Computational Statistics and Data Analysis 106 (2017) 153–164

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Estimating functional linear mixed-effects regression models

Baisen Liu [a], Liangliang Wang [b], Jiguo Cao [b,*]

[a] School of Statistics, Dongbei University of Finance and Economics, Dalian 116025, China

[b] Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC V5A1S6, Canada

ARTICLE INFO

Article history:
Received 5 January 2016
Received in revised form 26 May 2016
Accepted 19 September 2016
Available online 7 October 2016

Keywords:
EM algorithm
Functional linear regression
Smoothing spline
Random effects model

ABSTRACT

A new functional linear mixed model is proposed to investigate the impact of functional predictors on a scalar response when repeated measurements are available on multiple subjects. The advantage of the proposed model is that under the proposed model, each subject has both individual scalar covariate effects and individual functional effects over time, while it shares the common population scalar covariate effects and the common population slope functions. A smoothing spline method is proposed to estimate the population fixed and random slope functions, and a REML-based EM algorithm is developed to estimate fixed effects and variance parameters for random effects. Simulation studies illustrate that for finite samples the proposed estimation method can provide accurate estimates for the functional linear mixed-effects model. The proposed model is applied to investigate the effect of daily ozone concentration on annual nonaccidental mortality rates and also to study the effect of daily temperature on annual precipitation.

1. Introduction

When a random variable is observed at multiple time points or spatial locations, the data can be viewed as a function of time or spatial location. This type of data is generally called functional data (Ramsay and Silverman, 2005). Now in the big data era, functional data analysis (FDA) has become very attractive in statistical methodology and applied data analysis. Functional linear models (FLMs), introduced by Ramsay and Dalzell (1991), are some of the most popular models in FDA. FLMs can be used to model the relationship between functional variables and to predict a scalar response from functional covariates. Developments in modern technology have allowed FLMs to be applied to model functional data in many fields such as economics, medicine, environment, and climate [see for instance, Ramsay and Silverman, 2002, 2005, and Ferraty and Vieu, 2006, for several case studies].

The properties of FLMs have been thoroughly examined in the literature. For example, Yao et al. (2005) studied FLMs for sparse longitudinal data and suggested a nonparametric estimation method based on functional principal components analysis (FPCA). Their proposed functional regression approach is flexible enough to allow sparse measurements of functional predictors and response. Cai and Hall (2006) discussed the prediction problem in FLMs based on the FPCA technique. Crambes et al. (2009) proposed a smoothing spline estimator for the functional slope parameter, and extended this estimator to covariates with measurement-errors. Yuan and Cai (2010) suggested a smoothness regularization method for estimating FLMs based on the reproducing kernel Hilbert space (RKHS) approach. They provided a unified treatment for both the prediction and estimation problems by developing a tool that relys on simultaneous diagonalization of two positive-definite kernels. Wu et al. (2010) proposed a varying-coefficient FLM which allows for the slope to be modeled

* Correspondence to: Room SC K10545, 8888 University Drive, Burnaby, BC, Canada V5A1S6.
E-mail address: jiguo_cao@sfu.ca (J. Cao).

http://dx.doi.org/10.1016/j.csda.2016.09.009
0167-9473/

as a dependent function of additional scalar covariates. Gertheiss et al. (2013) extended the classical functional principal components regression (FPCR) by developing a longitudinal FPCR that allows for different effects of subject-specific trends in curves and for visit-specific deviations from that trend in longitudinal functional data. Scheipl et al. (2015) proposed a flexible functional additive mixed model that incorporates linear and nonlinear effects of functional and scalar covariates and allowed for flexible correlation structures of data. A systematic review on FLMs can be found in Morris (2015).

One conventional FLM involves linking a scalar response variable $Y_j, j = 1, \ldots, m$, to a functional predictor $X_j(t)$ through the following model:

$$Y_j = \alpha + \int_S \beta(t)X_j(t)\mathrm{d}t + \epsilon_j, \tag{1}$$

where α is the intercept, $\beta(t)$ is a smooth slope function, and the ϵ_j's are independent and identically distributed (i.i.d.) random variables with mean 0 and variance σ_ϵ^2. In this model, S is often assumed to be a compact subset of an Euclidean space such as $[0, 1]$. The slope function $\beta(t)$ represents the accumulative effect of the functional covariate $X_j(t)$ on the scalar response Y_j.

For purposes of illustration, we take the air pollution data as an example. This data is from the R package `NMMAPSdata` (Peng and Welty, 2004), which catalogues air pollution, weather, and mortality data for American cities from 1987 to 2000. Our aim is to investigate the impact of daily ozone concentration on nonaccidental mortality rates. The scalar response Y_{ij} is the logarithm of annual nonaccidental mortality rates in the jth year at the ith city, and the functional predictor $X_{ij}(t)$ is daily ozone concentration. In a preliminary analysis, we performed the classical FLM (1) on each individual city and found that there was dramatic variation in the estimate $\hat{\beta}(t)$ among different cities. This indicates that the effect of daily ozone concentration on the annual nonaccidental mortality rates is not the same in different cities. Therefore, it may not be appropriate to pool the data for all US cities and provide a single estimate for the average effect of daily ozone on the annual nonaccidental mortality rates. On the other hand, we may not fully exploit the information available in the data if we fit separate functional linear models for each city.

To address this dilemma, we generalize the FLM (1) to allow for the incorporation of random effects in the slope function. We call this new model the functional linear mixed-effects model (FLMM). Assume that we repeatedly observe a distinct functional predictor and scalar outcome for each subject over several visits. Then the observed data has the structure $\{Y_{ij}, \mathbf{X}_{ij}(t), \mathbf{W}_{ij}, \mathbf{Z}_{ij}\}, i = 1, \ldots, n, j = 1, \ldots, m_i$, where Y_{ij} is the jth repeated measurement of the scalar response for the ith subject, $\mathbf{W}_{ij}$ and $\mathbf{Z}_{ij}$ are vectors of scalar covariates, and $\mathbf{X}_{ij}(t) = (X_{ij1}(t), X_{ij2}(t), \ldots, X_{ijd}(t))$ are the corresponding functional predictors. The functional linear mixed-effects model can be expressed as follows:

$$Y_{ij} = \mathbf{W}_{ij}'\boldsymbol{\alpha} + \mathbf{Z}_{ij}'\boldsymbol{\gamma}_i + \sum_{\ell=1}^{d}\int_S [\beta_\ell(t) + b_{i\ell}(t)]X_{ij\ell}(t)\mathrm{d}t + \epsilon_{ij}, \tag{2}$$

where $\boldsymbol{\alpha}$ is a p-dimensional vector of the fixed effect, $\boldsymbol{\gamma}_i$ is the q-dimensional vector of random effect of scalar covariates, $\boldsymbol{\beta}(t) = (\beta_1(t), \ldots, \beta_d(t))'$ represents the population effect of $\mathbf{X}_{ij}(t)$ on Y_{ij}, $\mathbf{b}_i(t) = (b_{i1}(t), \ldots, b_{id}(t))'$ stands for the random effect of $\mathbf{X}_{ij}(t)$ on Y_{ij} for the ith subject, and ϵ_{ij} is the i.i.d. random variable with mean 0 and variance σ_ϵ^2. In this article, we assume that $\boldsymbol{\gamma}_i \sim N(0, \sigma_\epsilon^2\boldsymbol{\Psi})$, $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$, and $b_{i\ell}(t)$ follows a Gaussian stochastic process with mean 0 and covariance function $\gamma_\ell(s, t)$, that is, $b_{i\ell}(t) \sim GP(0, \gamma_\ell(s, t))$. We also assume that $\boldsymbol{\gamma}_i$, ϵ_{ij}, $\mathbf{b}_i(t)$, and $\mathbf{X}_{ij}(t)$ are mutually independent.

The functional linear mixed-effects model above is attractive, because it can estimate the population effect and random effect of the functional predictor $\mathbf{X}(t)$ (e.g., the daily temperature and ozone concentrations) on the scalar response Y (e.g., the annual nonaccidental deaths), as well as the population effect $\boldsymbol{\alpha}$ and random effect $\boldsymbol{\gamma}_i$. The application of the proposed functional linear mixed-effects model to the air pollution problem is not unique; many similar applications can be found in environmental and biological problems.

The proposed functional linear mixed-effect model (2) is different from the following functional mixed model (Goldsmith et al., 2011, 2012):

$$Y_{ij} = \mathbf{Z}_i\mathbf{b}_i + \int_S \beta(t)X_{ij}(t)\mathrm{d}t + \epsilon_{ij}, \tag{3}$$

where $\mathbf{b}_i \sim N(\mathbf{0}, \sigma_{\mathbf{b}}^2\mathbf{I})$ accounts for the correlation in the repeated outcomes for the ith subject. The primary distinction between models (2) and (3) is that the subject-specific random effect $\mathbf{b}_i$ in (3) remains constant across visits, while the random effect $\mathbf{b}_i(t)$ in (2) allows this effect to vary with time. By including the random effect $\mathbf{b}_i(t)$ in (2), this model can characterize trends in the effects of functional predictors on scalar outcomes for different subjects.

Many nonparametric smoothers used for FLMs can be applied to fit model (2). In this article, we employ the smoothing spline method (Ramsay and Silverman, 2005) to estimate $\beta_\ell(t)$ and $b_{i\ell}(t)$ in (2). Next, we transform model (2) into a linear mixed-effects model (LMM). Finally, we propose an REML-based EM algorithm to fit the LMM, the efficiency of which is illustrated by examples.

The remainder of this article is organized as follows. Section 2 introduces a smoothing spline method to estimate the above functional linear mixed-effects model. Section 3 implements simulations to evaluate the finite sample performance of the smoothing spline method. The functional linear mixed-effects model is demonstrated through two real applications in Section 4. Conclusions are given in Section 5.

2. Method

Without making any parametric assumptions on the slope functions, for each $\ell = 1, \ldots, d$, we express $\beta_\ell(t)$ and $b_{i\ell}(t)$ as linear combinations of basis functions

$$\beta_\ell(t) = \sum_{j=1}^{J} c_{j\ell}\phi_j(t) = \boldsymbol{\phi}'(t)\mathbf{c}_\ell, \tag{4}$$

and

$$b_{i\ell}(t) = \sum_{k=1}^{K} b_{ik\ell}\psi_k(t) = \boldsymbol{\psi}'(t)\mathbf{b}_{i\ell}, \tag{5}$$

where $\boldsymbol{\phi}(t) = (\phi_1(t), \ldots, \phi_J(t))'$ and $\boldsymbol{\psi}(t) = (\psi_1(t), \ldots, \psi_K(t))'$ are two vectors of basis functions with dimensions J and K, respectively, and $\mathbf{c}_\ell = (c_{1\ell}, \ldots, c_{J\ell})'$ and $\mathbf{b}_{i\ell} = (b_{i1\ell}, \ldots, b_{iK\ell})'$ are the corresponding vectors of basis coefficients to estimate. Many popular basis functions such as B-splines and Fourier basis can be exploited depending on the situation. If the functional predictor $X_\ell(t)$ is a periodic function, Fourier basis functions are recommended, otherwise cubic B-splines are the most popular choice. Generally, we recommend using a large number of basis functions to ensure reasonably flexible estimates for $\beta_\ell(t)$ and $b_{i\ell}(t)$, and controlling the roughness of the estimated functions by adding a roughness penalty.

Let $\sigma_\epsilon^2\mathbf{D}_\ell = \mathrm{Cov}(\mathbf{b}_{i\ell}) = \mathrm{E}(\mathbf{b}_{i\ell}\mathbf{b}_{i\ell}')$ denote the variance–covariance matrix of the random-effects, so that $\mathbf{b}_{i\ell} \sim N(0, \sigma_\epsilon^2\mathbf{D}_\ell)$. Accordingly the covariance function $\gamma_\ell(s,t)$ for the random effect $b_{i\ell}(t)$ can be expressed as $\gamma_\ell(s,t) = \sigma_\epsilon^2\boldsymbol{\psi}'(s)\mathbf{D}_\ell\boldsymbol{\psi}(t)$. We first consider the scenario where the functional predictors $\mathbf{X}_{ij}(t)$ are observed without measurement errors. When $\mathbf{X}_{ij}(t)$ are observed with measurement errors, many nonparametric smoothing approaches can be applied to reconstruct the underlying functional predictors, $\mathbf{X}_{ij}(t)$, such as smoothing spline or FPCA (Yao et al., 2005), which will be discussed in Section 2.5.

2.1. Estimating fixed and random effects

For $\ell = 1, \ldots, d$, let $\mathbf{B}_{ij\ell} = \int_S X_{ij\ell}(t)\boldsymbol{\phi}(t)\mathrm{d}t$ and $\mathbf{C}_{ij\ell} = \int_S X_{ij\ell}(t)\boldsymbol{\psi}(t)\mathrm{d}t$. The FLMM model (2) can be expressed as

$$Y_{ij} = \mathbf{W}_{ij}'\boldsymbol{\alpha} + \mathbf{Z}_{ij}'\boldsymbol{\gamma}_i + \sum_{\ell=1}^{d}(\mathbf{B}_{ij\ell}'\mathbf{c}_\ell + \mathbf{C}_{ij\ell}'\mathbf{b}_{i\ell}) + \epsilon_{ij}, \quad 1 \le i \le n,\ 1 \le j \le m_i. \tag{6}$$

Let $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \mathbf{c}_1', \ldots, \mathbf{c}_d')'$ and $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_1', \ldots, \boldsymbol{\vartheta}_n')'$ with $\boldsymbol{\vartheta}_i = (\boldsymbol{\gamma}_i', \mathbf{b}_{i1}', \ldots, \mathbf{b}_{id}')'$, $i = 1, \ldots, n$. The fixed effects $\{\boldsymbol{\alpha}, \boldsymbol{\beta}(t)\}$, and random effects $\{\boldsymbol{\gamma}_i\mathbf{b}_i(t)\}$ are estimated by minimizing

$$\begin{aligned} H(\boldsymbol{\theta}, \boldsymbol{\vartheta}) = {} & \sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{1}{2\sigma_\epsilon^2}\left[Y_{ij} - \mathbf{W}_{ij}'\boldsymbol{\alpha} - \mathbf{Z}_{ij}'\boldsymbol{\gamma}_i - \sum_{\ell=1}^{d}(\mathbf{B}_{ij\ell}'\mathbf{c}_\ell + \mathbf{C}_{ij\ell}'\mathbf{b}_{i\ell})\right]^2 \\ & + \sum_{\ell=1}^{d}\left[\frac{\lambda_\ell^\beta}{2}\int_S\left\{\frac{\mathrm{d}^2\beta_\ell(t)}{\mathrm{d}t^2}\right\}^2\mathrm{d}t + \frac{\lambda_\ell^b}{2\sigma_\epsilon^2}\sum_{i=1}^{n}\int_S\left\{\frac{\mathrm{d}^2 b_{i\ell}(t)}{\mathrm{d}t^2}\right\}^2\mathrm{d}t\right] \\ & + \frac{1}{2\sigma_\epsilon^2}\sum_{\ell=1}^{d}\sum_{i=1}^{n}\mathbf{b}_{i\ell}'\mathbf{D}_\ell^{-1}\mathbf{b}_{i\ell} + \frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^{n}\boldsymbol{\gamma}_i'\boldsymbol{\Psi}^{-1}\boldsymbol{\gamma}_i. \end{aligned} \tag{7}$$

Define three vectors $\mathbf{Y}_i = [Y_{i1}, \ldots, Y_{im_i}]'$, $\widetilde{\mathbf{W}}_i = [\widetilde{\mathbf{W}}_{i1}, \ldots, \widetilde{\mathbf{W}}_{im_i}]'$, and $\widetilde{\mathbf{Z}}_i = [\widetilde{\mathbf{Z}}_{i1}, \ldots, \widetilde{\mathbf{Z}}_{im_i}]'$, where $\widetilde{\mathbf{W}}_{ij} = \left[\mathbf{W}_{ij}', \mathbf{B}_{ij,1}', \ldots, \mathbf{B}_{ij,d}'\right]'$ and $\widetilde{\mathbf{Z}}_{ij} = \left[\mathbf{Z}_{ij}', \mathbf{C}_{ij,1}', \ldots, \mathbf{C}_{ij,d}'\right]'$. Moreover, define a $J \times J$ matrix $\mathbf{G}_\beta = \int_S(\mathrm{d}^2\boldsymbol{\phi}(t)/\mathrm{d}t^2)(\mathrm{d}^2\boldsymbol{\phi}(t)/\mathrm{d}t^2)'\mathrm{d}t$ and a $K \times K$ matrix $\mathbf{G}_b = \int_S(\mathrm{d}^2\boldsymbol{\psi}(t)/\mathrm{d}t^2)(\mathrm{d}^2\boldsymbol{\psi}(t)/\mathrm{d}t^2)'\mathrm{d}t$. Then, $H(\boldsymbol{\theta}, \boldsymbol{\vartheta})$ can be expressed in matrix form as

$$\begin{aligned} H(\boldsymbol{\theta}, \boldsymbol{\vartheta}) = {} & \sum_{i=1}^{n}\frac{1}{2\sigma_\epsilon^2}\|\mathbf{Y}_i - \widetilde{\mathbf{W}}_i\boldsymbol{\theta} - \widetilde{\mathbf{Z}}_i\boldsymbol{\vartheta}_i\|^2 + \sum_{\ell=1}^{d}\left[\frac{\lambda_\ell^\beta}{2}\mathbf{c}_\ell'\mathbf{G}_\beta\mathbf{c}_\ell + \frac{\lambda_\ell^b}{2\sigma_\epsilon^2}\sum_{i=1}^{n}\mathbf{b}_{i\ell}'\mathbf{G}_b\mathbf{b}_{i\ell}\right] \\ & + \frac{1}{2\sigma_\epsilon^2}\sum_{\ell=1}^{d}\sum_{i=1}^{n}\mathbf{b}_{i\ell}'\mathbf{D}_\ell^{-1}\mathbf{b}_{i\ell} + \frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^{n}\boldsymbol{\gamma}_i'\boldsymbol{\Psi}^{-1}\boldsymbol{\gamma}_i. \end{aligned}$$

The estimators $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\vartheta}}$, obtained by minimizing $H(\boldsymbol{\theta}, \boldsymbol{\vartheta})$, are given by

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \left(\widetilde{\mathbf{W}}'\widetilde{\mathbf{V}}^{-1}\widetilde{\mathbf{W}} + \widetilde{\mathbf{G}}_\theta\right)^{-1}\widetilde{\mathbf{W}}'\widetilde{\mathbf{V}}^{-1}\mathbf{Y}, \\ \hat{\boldsymbol{\vartheta}}_i &= \widetilde{\mathbf{D}}_\vartheta\widetilde{\mathbf{Z}}_i'\widetilde{\mathbf{V}}_i^{-1}(\mathbf{Y}_i - \widetilde{\mathbf{W}}_i\hat{\boldsymbol{\theta}}). \end{aligned} \tag{8}$$

where $\mathbf{Y} = (\mathbf{Y}_1', \ldots, \mathbf{Y}_n')'$, $\widetilde{\mathbf{W}} = [\widetilde{\mathbf{W}}_1', \ldots, \widetilde{\mathbf{W}}_n']'$, $\widetilde{\mathbf{V}} = \mathrm{diag}(\widetilde{\mathbf{V}}_1, \ldots, \widetilde{\mathbf{V}}_n)$ with $\widetilde{\mathbf{V}}_i = \widetilde{\mathbf{Z}}_i\widetilde{\mathbf{D}}_\vartheta\widetilde{\mathbf{Z}}_i' + \mathbf{I}_{m_i}$ $(1 \le i \le n)$, $\widetilde{\mathbf{D}}_\vartheta = \mathrm{diag}(\boldsymbol{\Psi}, \widetilde{\mathbf{D}}_1, \ldots, \widetilde{\mathbf{D}}_d)$ with $\widetilde{\mathbf{D}}_\ell = \left(\mathbf{D}_\ell^{-1} + \lambda_\ell^b\mathbf{G}_b\right)^{-1}$ $(1 \le \ell \le d)$, $\widetilde{\mathbf{G}}_\theta = \mathrm{diag}(\mathbf{0}_{p\times p}, \mathrm{diag}(\lambda_1^\beta, \ldots, \lambda_d^\beta) \bigotimes \mathbf{G}_\beta)$, $\widetilde{\mathbf{Z}} = \mathrm{diag}(\widetilde{\mathbf{Z}}_1', \ldots, \widetilde{\mathbf{Z}}_n')$, and $\bigotimes$ denotes the kronecker product.

Once the estimates $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\alpha}}, \widehat{\mathbf{c}}_1', \ldots, \widehat{\mathbf{c}}_d')'$ and $\hat{\boldsymbol{\vartheta}}_i = (\hat{\boldsymbol{\gamma}}_i, \widehat{\mathbf{b}}_{i1}', \ldots, \widehat{\mathbf{b}}_{id}')'$ have been obtained, the estimates of $\beta_\ell(t)$ and $b_{i\ell}(t)$ are given by

$$\hat{\beta}_\ell(t) = \boldsymbol{\phi}'(t)\widehat{\mathbf{c}}_\ell, \qquad \hat{b}_{i\ell}(t) = \boldsymbol{\psi}'(t)\widehat{\mathbf{b}}_{i\ell}, \quad i = 1, \ldots, n, \ \ell = 1, \ldots, d. \tag{9}$$

2.2. The REML-based EM algorithm

To simplify notation, we set $\mathbf{D}_0 = \boldsymbol{\Psi}$ and $\lambda_0^b = 0$. To estimate the fixed-effects, $\boldsymbol{\theta}$, the random-effects, $\boldsymbol{\vartheta}_i$, and the variance parameters, σ_ϵ^2 and $\mathbf{D} = \mathrm{diag}(\mathbf{D}_0, \mathbf{D}_1, \ldots, \mathbf{D}_d)$, we recommend an EM algorithm procedure called the *REML-based EM-algorithm*. This algorithm was proposed by Wu and Zhang (2006) to estimate nonparametric mixed-effects regression models with longitudinal data. The REML-based EM-algorithm has three steps, which are outlined as follows.

Initializing. Initializing the starting values for σ_ϵ^2 and $\mathbf{D}$, denoted by $\sigma_\epsilon^{2(0)}$ and $\mathbf{D}^{(0)}$, respectively. For example, we could set $\sigma_\epsilon^{2(0)} = 1$ and $\mathbf{D}^{(0)}$ to be the identity matrix.

Step 1. Set $r = r + 1$. Compute

$$\begin{aligned}
\widetilde{\mathbf{D}}_\ell^{(r-1)} &= \{\mathbf{D}_\ell^{(r-1)}\}^{-1} + \lambda_\ell^b\mathbf{G}_b, \quad \ell = 0, 1, \ldots, d,\\
\widetilde{\mathbf{D}}_\vartheta^{(r-1)} &= \left\{\mathrm{diag}(\widetilde{\mathbf{D}}_0^{(r-1)}, \widetilde{\mathbf{D}}_1^{(r-1)}, \ldots, \widetilde{\mathbf{D}}_d^{(r-1)})\right\}^{-1},\\
\widetilde{\mathbf{V}}_i^{(r-1)} &= \widetilde{\mathbf{Z}}_i\widetilde{\mathbf{D}}_\vartheta^{(r-1)}\widetilde{\mathbf{Z}}_i' + \mathbf{I}_{m_i}, \quad i = 1, 2, \ldots, n.
\end{aligned}$$

Denote $\widetilde{\mathbf{V}}^{(r-1)} = \mathrm{diag}(\widetilde{\mathbf{V}}_1^{(r-1)}, \ldots, \widetilde{\mathbf{V}}_n^{(r-1)})$. Then estimate $\hat{\boldsymbol{\theta}}^{(r)}$ and $\hat{\boldsymbol{\vartheta}}_i^{(r)}$ by

$$\begin{aligned}
\hat{\boldsymbol{\theta}}^{(r)} &= [\widetilde{\mathbf{W}}'\{\widetilde{\mathbf{V}}^{(r-1)}\}^{-1}\widetilde{\mathbf{W}} + \widetilde{\mathbf{G}}_\theta]^{-1}\widetilde{\mathbf{W}}'\{\widetilde{\mathbf{V}}^{(r-1)}\}^{-1}\mathbf{Y},\\
\hat{\boldsymbol{\vartheta}}_i^{(r)} &= \widetilde{\mathbf{D}}_\vartheta^{(r-1)}\widetilde{\mathbf{Z}}_i'\{\widetilde{\mathbf{V}}_i^{(r-1)}\}^{-1}(\mathbf{Y}_i - \widetilde{\mathbf{W}}_i\hat{\boldsymbol{\theta}}^{(r)}), \quad i = 1, 2, \ldots, n.
\end{aligned}$$

Step 2. Compute the residuals $\hat{\boldsymbol{\epsilon}}_i^{(r)} = \mathbf{Y}_i - \widetilde{\mathbf{W}}_i\hat{\boldsymbol{\theta}}^{(r)} - \widetilde{\mathbf{Z}}_i\hat{\boldsymbol{\vartheta}}_i^{(r)}$ and the matrix $\mathbf{H}_i^{(r-1)} = \{\widetilde{\mathbf{V}}_i^{(r-1)}\}^{-1} - \{\widetilde{\mathbf{V}}_i^{(r-1)}\}^{-1}\widetilde{\mathbf{W}}_i[\widetilde{\mathbf{W}}'\{\widetilde{\mathbf{V}}^{(r-1)}\}^{-1}\widetilde{\mathbf{W}} + \widetilde{\mathbf{G}}_\theta]^{-1}\widetilde{\mathbf{W}}_i'\{\widetilde{\mathbf{V}}_i^{(r-1)}\}^{-1}$. Then, the updates of $\sigma_\epsilon^{2(r)}$ and $\mathbf{D}^{(r)} = \mathrm{diag}(\mathbf{D}_0^{(r)}, \mathbf{D}_1^{(r)}, \ldots, \mathbf{D}_d^{(r)})$ are given by

$$\begin{aligned}
\sigma_\epsilon^{2(r)} &= N^{-1}\sum_{i=1}^n\left\{\{\hat{\boldsymbol{\epsilon}}_i^{(r)}\}'\hat{\boldsymbol{\epsilon}}_i^{(r)} + \sigma_\epsilon^{2(r-1)}[m_i - \mathrm{trace}(\mathbf{H}_i^{(r-1)})]\right\},\\
\mathbf{D}^{(r)} &= n^{-1}\sum_{i=1}^n\left\{\frac{1}{\sigma_\epsilon^2}\hat{\boldsymbol{\vartheta}}_i^{(r)}\{\hat{\boldsymbol{\vartheta}}_i^{(r)}\}' + [\widetilde{\mathbf{D}}_\vartheta^{(r-1)} - \widetilde{\mathbf{D}}_\vartheta^{(r-1)}\widetilde{\mathbf{Z}}_i'\mathbf{H}_i^{(r-1)}\widetilde{\mathbf{Z}}_i\widetilde{\mathbf{D}}_\vartheta^{(r-1)}]\right\}.
\end{aligned}$$

Step 3. Set a tolerance rate $\delta > 0$ (e.g., $\delta = 10^{-6}$). Repeat Steps 1 and 2, until the following convergence conditions

$$\max\left\{\|\boldsymbol{\theta}^{(r)} - \boldsymbol{\theta}^{(r-1)}\|, \|\hat{\boldsymbol{\vartheta}}^{(r)} - \hat{\boldsymbol{\vartheta}}^{(r-1)}\|, \|\mathbf{D}^{(r)} - \mathbf{D}^{(r-1)}\|, |\sigma_\epsilon^{2(r)} - \sigma_\epsilon^{2(r-1)}|\right\} < \delta,$$

are satisfied, where $\|\mathbf{A}\|$ denotes the Frobenius norm of the matrix $\mathbf{A}$.

2.3. Smoothing parameter selection

The smoothness of $\beta_\ell(t)$ and $b_{i\ell}(t)$ $(1 \le \ell \le d)$ are controlled by the smoothing parameter λ_ℓ^β and λ_ℓ^b, respectively. The optimal value for $\boldsymbol{\lambda}^\beta = (\lambda_1^\beta, \ldots, \lambda_d^\beta)'$ is chosen by minimizing the generalized cross-validation (GCV) criterion. The optimal value of $\boldsymbol{\lambda}^b = (\lambda_1^b, \ldots, \lambda_d^b)'$ is typically chosen using the REML method since $\boldsymbol{\lambda}^b$ is a parameter in the variance of the random effects.

(a) *Selecting the optimal* $\boldsymbol{\lambda}^\beta$

Define the following GCV criterion

$$\mathrm{GCV}(\boldsymbol{\lambda}^\beta) = \frac{\mathrm{SSE}(\boldsymbol{\lambda}^\beta)}{\mathrm{trace}\{\mathbf{I}_N - \mathbf{S}(\boldsymbol{\lambda}^\beta)\}^2}, \tag{10}$$

where $\mathrm{SSE}(\boldsymbol{\lambda}^\beta) = \sum_{i=1}^n \|\mathbf{Y}_i - \widetilde{\mathbf{W}}_i\widehat{\boldsymbol{\theta}} - \widetilde{\mathbf{Z}}_i\widehat{\boldsymbol{\vartheta}}_i\|^2$, $N = \sum_{i=1}^n m_i$, and $\mathbf{S}(\boldsymbol{\lambda}^\beta)$ is given by

$$\mathbf{S}(\boldsymbol{\lambda}^\beta) = \widetilde{\mathbf{W}}\left(\widetilde{\mathbf{W}}'\widetilde{\mathbf{W}} + \widetilde{\mathbf{G}}_\theta\right)^{-1}\widetilde{\mathbf{W}}'.$$

Then, the optimal $\boldsymbol{\lambda}^\beta$ is chosen by minimizing (10).

(b) *Selecting the optimal* $\boldsymbol{\lambda}^b$

Consider the linear mixed model:

$$\mathbf{Y} = \widetilde{\mathbf{W}}\boldsymbol{\theta} + \widetilde{\mathbf{Z}}\boldsymbol{\vartheta} + \mathbf{e},$$

where $\boldsymbol{\vartheta} \sim N(\mathbf{0}, \sigma_\epsilon^2 \widetilde{\mathbf{D}}_\vartheta)$, $\widetilde{\mathbf{D}}_\vartheta = \mathrm{diag}(\boldsymbol{\Psi}, \widetilde{\mathbf{D}}_1, \ldots, \widetilde{\mathbf{D}}_d)$ with $\widetilde{\mathbf{D}}_\ell = (\mathbf{D}_\ell^{-1} + \lambda_\ell^b \mathbf{G}_b)^{-1}$ $(1 \le \ell \le d)$, and $\mathbf{e} \sim N(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}_N)$. We can apply the REML method to estimate $\boldsymbol{\lambda}^b$ by maximizing

$$l_R(\boldsymbol{\lambda}^b) \propto -\frac{1}{2}\log|\widetilde{\mathbf{V}}| - \frac{1}{2}\log|\widetilde{\mathbf{W}}'\widetilde{\mathbf{V}}^{-1}\widetilde{\mathbf{W}}| - \frac{1}{2}(\mathbf{Y} - \widetilde{\mathbf{W}}\hat{\boldsymbol{\theta}})'\widetilde{\mathbf{V}}^{-1}(\mathbf{Y} - \widetilde{\mathbf{W}}\hat{\boldsymbol{\theta}}),$$

using a numerical approach such as the Newton–Raphson method.

2.4. Constructing confidence intervals

To construct confidence intervals for $\boldsymbol{\alpha}$ and point-wise confidence intervals for $\boldsymbol{\beta}(t)$, we need to calculate the covariance matrix of $\hat{\boldsymbol{\theta}}$:

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}}) = \mathrm{Cov}\left\{\left(\widetilde{\mathbf{W}}'\widetilde{\mathbf{V}}^{-1}\widetilde{\mathbf{W}} + \widetilde{\mathbf{G}}_\theta\right)^{-1}\widetilde{\mathbf{W}}'\widetilde{\mathbf{V}}^{-1}\mathbf{Y}_i\right\}.$$

However, $\mathrm{Cov}(\mathbf{Y}_i)$ can be replaced by $\widetilde{\mathbf{V}}_i$ to account for our roughness penalty on $\mathbf{b}_i(t)$, $i = 1, \ldots, n$. For simplicity, we set

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}}) = \sigma_\epsilon^2\left(\widetilde{\mathbf{W}}'\widetilde{\mathbf{V}}^{-1}\widetilde{\mathbf{W}} + \widetilde{\mathbf{G}}_\theta\right)^{-1}. \tag{11}$$

Let $\widehat{\mathrm{Cov}}(\hat{\boldsymbol{\theta}})$ be the estimator of the covariance matrix (11) and partition it as

$$\widehat{\mathrm{Cov}}(\hat{\boldsymbol{\theta}}) = \hat{\sigma}_\epsilon^2 \begin{pmatrix} \hat{\boldsymbol{\Sigma}}_{11} & * & \cdots & * \\ * & \hat{\boldsymbol{\Sigma}}_{22,1} & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & \hat{\boldsymbol{\Sigma}}_{22,d} \end{pmatrix}.$$

The 95% confidence intervals for α_i may be approximated by

$$\hat{\boldsymbol{\alpha}}_i \pm 1.96\hat{\sigma}_\epsilon\sqrt{\hat{\boldsymbol{\Sigma}}_{11}(i)}, \quad i = 1, \ldots, p,$$

where $\hat{\boldsymbol{\Sigma}}_{11}(i)$ denotes the ith diagonal element of $\hat{\boldsymbol{\Sigma}}_{11}$. The 95% pointwise confidence intervals for $\beta_\ell(t)$ may be approximated by

$$\hat{\beta}_\ell(t) \pm 1.96\sqrt{\widehat{\mathrm{Var}}[\hat{\beta}_\ell(t)]}, \quad \ell = 1, \ldots, d, \text{ for all } t \in S,$$

where $\widehat{\mathrm{Var}}[\hat{\beta}_\ell(t)] = \hat{\sigma}_\epsilon^2\boldsymbol{\phi}'(t)\hat{\boldsymbol{\Sigma}}_{22\ell}\boldsymbol{\phi}(t)$. Moreover, an estimate of $\gamma_\ell(s,t)$ is given by

$$\hat{\gamma}_\ell(s,t) = \hat{\sigma}_\epsilon^2\boldsymbol{\psi}'(s)\left(\widehat{\mathbf{D}}_\ell^{-1} + \lambda_\ell^b\mathbf{G}_\ell\right)^{-1}\boldsymbol{\psi}(t), \quad \ell = 1, \ldots, d.$$

2.5. Reconstructing the predictors $X_{ij\ell}(t)$

If the functional predictors $\{X_{ij\ell}(t)\}_{\ell=1}^d$ in the functional linear mixed-effects model (2) are subject to measurement errors, we must account for this error to ensure that our estimates and inferences are not biased. In this article, we suggest first reconstructing the functional predictors $\{X_{ij\ell}(t)\}_{\ell=1}^d$ using a large number of functional principal components obtained from a smooth estimator of the covariance matrix (Goldsmith et al., 2012). Then, we treat the estimates $\{\hat{X}_{ij\ell}(t)\}_{\ell=1}^d$ as the true predictors and apply the REML-based EM algorithm.

Define the covariance function of $X(t)$ to be

$$C(s,t) = \mathrm{Cov}(X(t), X(s)).$$

Mercer's theorem (Ash and Gardner, 1975) states that $C(s,t)$ has eigen-decomposition

$$C(s,t) = \sum_{k=1}^{\infty}\lambda_k\varphi_k(s)\varphi_k(t),$$

where $\lambda_1 \ge \lambda_2 \ge \cdots \ge 0$ satisfy $\sum_{k=1}^{\infty}\lambda_k < \infty$, and the $\varphi_k(t)$'s form a complete orthonormal basis in $S \times S$. Then, $X(t)$ can be reformulated according to the Karhunen–Loeve decomposition (Rice and Silverman, 1991)

$$X(t) = \mu(t) + \sum_{k=1}^{\infty}\xi_k\varphi_k(t)$$

where $\varphi_k(\cdot)$ is the orthonormal eigenfunction, also called the functional principal component (FPC). The coefficients ξ_k are called the FPC scores of $X(t)$, and satisfy $E(\xi_k) = 0$, $E(\xi_k^2) = \lambda_k$, and $E(\xi_k \xi_l) = 0$ for $k \neq l$.

Suppose we have the additive measurement error model,

$$W_{ij\ell}(t) = X_{ij\ell}(t) + e_{ij\ell}(t), \quad \ell = 1, \ldots, d,$$

where $W_{ij\ell}(t)$ is the observed value at time t, $X_{ij\ell}(t)$ is the underlying true value for the ith subject at the time t, and $e_{ij\ell}(t)$ represents the measurement error at the time t. We assume that $e_{ij\ell}(t)$ is a mean zero process, and that $\{X_{ij\ell}(t), e_{ij\ell}(t)\}$ are mutually independent. We estimate $C(s, t)$ using a method-of-moments approach, and then smooth the off-diagonal elements of the observed covariance matrix to remove the nugget effect that is caused by measurement error (Staniswalis and Lee, 1998; Yao et al., 2005; Goldsmith et al., 2012).

We use the principal analysis by conditional estimation (PACE) algorithm proposed by Yao et al. (2005) to estimate the mean curve, $\mu_{i\ell}(t)$, the FPCs, $\varphi_{ik\ell}(t)$ and the FPC scores, $\xi_{ijk\ell}$, from the observations, $W_{ij\ell}(t_{ijk})$. Let $\hat{\mu}_{i\ell}(t)$, $\hat{\varphi}_{ik\ell}(t)$, and $\hat{\xi}_{ijk\ell}$ be the respective estimators of $\mu_{i\ell}(t)$, $\varphi_{ik\ell}(t)$ and $\xi_{ijk\ell}$. Then, $X_{ij\ell}(t)$ can be estimated by

$$\hat{X}_{ij\ell}(t) = \hat{\mu}_{i\ell}(t) + \sum_{k=1}^{M} \hat{\xi}_{ijk\ell} \hat{\varphi}_{ik\ell}(t) \tag{12}$$

where the number of FPCs, M, can be chosen by AIC, BIC, a cross-validation method, or empirical experience based on the percentage of explained variance (such as 90% or 95%).

When the functional predictor $X_{ij\ell}(t)$ is observed at dense time points, another method to re-construct $X_{ij\ell}(t)$ is the smoothing spline approach, the details of which may be found in Ramsay and Silverman (2005).

3. Simulation studies

In this section, we perform numerical experiments to assess the efficiency of our estimation procedure for the functional linear mixed-effects model (2). The performance of our estimation method is evaluated by the root mean square error (RMSE) for the scalar fixed effect $\hat{\alpha}_k$, and by the root mean integrated square error (RMISE) for the estimated population slope function $\hat{\beta}_\ell(t)$ and the individual slope function $\hat{\beta}_{i\ell}(t) = \hat{\beta}_\ell(t) + \hat{b}_{i\ell}(t)$. The RMSE for $\hat{\alpha}_k$ is given by

$$\text{RMSE}(\hat{\alpha}_k) = \left[\frac{1}{M} \sum_{m=1}^{M} (\hat{\alpha}_k^{(m)} - \alpha_k)^2 \right]^{\frac{1}{2}},$$

where $\hat{\alpha}_k^{(m)}$, $m = 1, \ldots, M$, are the estimates of α_k in the mth simulation replicate. The respective RMISE for the estimated population slope function $\hat{\beta}_\ell(t)$ and the individual slope function $\hat{\beta}_{1\ell}(t), \ldots, \hat{\beta}_{n\ell}(t)$ are given by

$$\text{RMISE}(\hat{\beta}_\ell(t)) = \left[\frac{1}{M} \sum_{m=1}^{M} \frac{\int_S (\hat{\beta}_\ell^{(m)}(t) - \beta_\ell(t))^2 \mathrm{d}t}{\int_S \beta_\ell^2(t) \mathrm{d}t} \right]^{\frac{1}{2}},$$

and

$$\text{RMISE}(\hat{\beta}_{1\ell}(t), \ldots, \hat{\beta}_{n\ell}(t)) = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{1}{M} \sum_{m=1}^{M} \frac{\int_S (\hat{\beta}_{i\ell}^{(m)}(t) - \beta_{i\ell}^{(m)}(t))^2 \mathrm{d}t}{\int_S \{\beta_{i\ell}^{(m)}(t)\}^2 \mathrm{d}t} \right]^{\frac{1}{2}},$$

where $\hat{\beta}_\ell^{(m)}(t)$ and $\hat{b}_{i\ell}^{(m)}(t)$, $m = 1, \ldots, M$ are the estimates of $\beta_\ell(t)$ and $b_{i\ell}^{(m)}(t)$ in the mth simulation replicate, $\beta_{i\ell}^{(m)}(t) = \beta_\ell(t) + b_{i\ell}^{(m)}(t)$, and $\hat{\beta}_{i\ell}^{(m)}(t) = \hat{\beta}_\ell^{(m)}(t) + \hat{b}_{i\ell}^{(m)}(t)$, $1 \leq i \leq n$, $1 \leq \ell \leq d$.

We assume that the scalar response Y_{ij} is generated from the following model:

$$Y_{ij} = \mathbf{W}_{ij}' \boldsymbol{\alpha} + \mathbf{Z}_{ij}' \boldsymbol{\gamma}_i + \sum_{\ell=1}^{2} \int_0^1 [\beta_\ell(t) + b_{i\ell}(t)] X_{ij\ell}(t) \mathrm{d}t + \epsilon_{ij}, \tag{13}$$

where $\mathbf{W}_{ij} = (1, W_{ij1}, W_{ij2})'$, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m_i$, with W_{ij1} generated from the *Bernoulli*(1, 0.5) distribution and W_{ij2} generated from the *Uniform*(0, 1) distribution. We assume that $\boldsymbol{\alpha}$ is a three-dimensional vector of the fixed effects with the true value $(3.0, 1.0, 0.5)'$, $\mathbf{Z}_{ij} = \mathbf{W}_{ij}$, and $\boldsymbol{\gamma}_i$ is a three-dimensional vector of the random effects with the true distribution as a multivariate normal distribution *Normal*$((0, 0, 0)', \text{diag}(0.5, 0.5, 0.2))$. For the population effect of $X_{ij\ell}(t)$ on Y_{ij}, $\beta_\ell(t)$, we set the true functions $\beta_1(t) = 1 + 2t^2 + e^{-3t}$ and $\beta_2(t) = 1 + 2\sin(2\pi t) + \cos(2\pi t)$, respectively. Next, we assume that the random effect, $b_{i\ell}(t)$, is generated as $b_{i1}(t) = \eta_{0i} + \eta_{1i} t^2 + \eta_{2i} e^{-3t}$ and $b_{i2}(t) = \zeta_{0i} + \zeta_{1i} \sin(2\pi t) + \zeta_{2i} \cos(2\pi t)$, where the random coefficients $(\eta_{0i}, \eta_{1i}, \eta_{2i})'$ and $(\zeta_{0i}, \zeta_{1i}, \zeta_{2i})'$ are independently distributed according to a multivariate normal distribution $N((0, 0, 0)', \text{diag}(0.2^2, 0.4^2, 0.2^2))$. Finally, we simulate $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$, where true value of σ_ϵ is assumed to be 1.0.

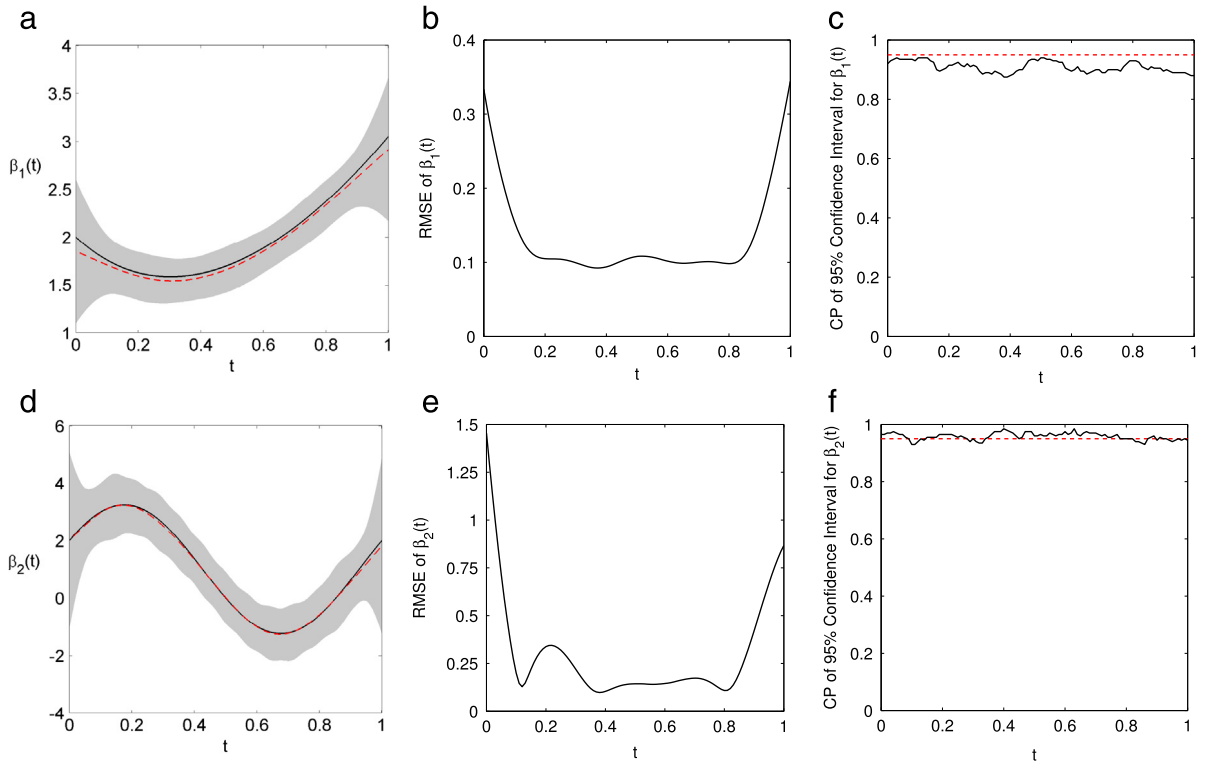


Fig. 1. The pointwise mean, pointwise root mean squared error (RMSE) of the estimated population slope function $\hat{\beta}_1(t)$ (top panels) and $\hat{\beta}_2(t)$ (bottom panels), and the pointwise coverage probability (CP) of their estimated 95% confidence intervals when $n = 100$, $m_i = 10$, $\sigma_X^2 = 1.0$ in 200 simulation replicates. In panel (a), the true population slope function is $\beta_1(t) = 1 + 2t^2 + \exp(-3t)$ (solid line); the dashed line and the shaded area are the average of the estimate of $\beta_1(t)$ and the average of the pointwise 95% confidence intervals, respectively. In panel (d), the true population slope function is $\beta_2(t) = 1 + 2\sin(2\pi t) + \cos(2\pi t)$ (solid line); the dashed line and the shaded area are the average of the estimate of $\beta_2(t)$ and the average of the estimated pointwise 95% confidence intervals, respectively. The horizontal dashed lines in panels (c) and (f) indicate the value 95%.

We assume that the two functional predictors $X_{ij1}(t)$ and $X_{ij2}(t)$ are observed at 101 equally-spaced time points in [0, 1] with the additive normal measurement errors

$$W_{ij\ell}(t_k) = X_{ij\ell}(t_k) + e_{ijk\ell}, \quad e_{ijk\ell} \sim N(0, \sigma_X^2), \ \ell = 1, 2,$$

where $\sigma_X = 0.0$ or 1.0. The true underlying predictors $X_{ij\ell}(t)$ are given by

$$X_{ij\ell}(t) = \mu_{i\ell}(t) + \sqrt{2}\sum_{k=1}^{4} \xi_{ijk\ell}\varphi_k(t), \quad t \in [0, 1], \ \ell = 1, 2,$$

where $\mu_{i\ell}(t) = \delta_{i0\ell} + \delta_{i1\ell}\sin(\pi t)$ with mutually independent random variables $\delta_{i0\ell} \sim \mathit{Uniform}[-2, 2]$, $\delta_{i1\ell} \sim \mathit{Normal}(0, 4)$, $\xi_{ijk\ell} \sim \mathit{Normal}(0, 2/2^k)$, $k = 1, 2, 3, 4$, and $\varphi_1(t) = \sin(2\pi t)$, $\varphi_2(t) = \cos(2\pi t)$, $\varphi_3(t) = \sin(4\pi t)$ and $\varphi_4(t) = \cos(4\pi t)$. We allow the number of repeated measurements for each individual, m_i, to take on values 5 and 10, and we let the total number of subjects take on values 50 and 100.

The functional linear mixed-effects model (13) is estimated using the method introduced in Section 2. Both the population slope function $\beta_\ell(t)$ and the random slope function $b_{i\ell}(t)$ are estimated using 17 cubic B-spline basis function with 15 equally-spaced knots in [0, 1]. Fig. 1 displays the pointwise mean and pointwise root mean squared error of the estimated population slopes $\hat{\beta}_\ell(t)$, $\ell = 1, 2$ in 200 simulation replicates when $n = 100$, $m_i = 10$, and $\sigma_X = 1.0$. Fig. 1 shows that the pointwise mean of $\hat{\beta}_\ell(t)$ is very close to the true $\beta_\ell(t)$. Fig. 1 also shows the pointwise coverage probability of the estimated 95% confidence intervals for $\hat{\beta}_\ell(t)$, which are close to 95%.

The estimation results for the scalar fixed effects are summarized in Table 1. The RMSEs for all three scalar fixed effects decrease when the number of observation per subject, m_i, increases from 5 to 10 or when the number of subjects increases from 50 to 100. Comparing the results when the functional predictor $X_{ij\ell}(t)$ is contaminated with random noises ($\sigma_X = 1.0$) to the results when $X_{ij\ell}(t)$ is observed directly ($\sigma_X = 0$), we see that the RMSEs for all three scalar fixed effects display little change and the lengths of 95% confidence intervals decrease substantially. The coverage probabilities of the 95% confidence intervals for all three scalar fixed effects are close to 95%.

Table 1
The mean, standard deviation (SD), root mean square error (RMSE), average interval lengths of 95% confidence intervals (AIL), and coverage probabilities (CP) of 95% confidence intervals for $(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2)'$ obtained by applying the REML-based EM algorithm on the simulated data with 200 simulation replicates. The true values of $(\alpha_0, \alpha_1, \alpha_2)'$ is $(3.0, 1.0, 0.5)'$.

σ_X	n	m_i	Fixed effects				
				Mean	SD	RMSE	AIL	CP (%)
0.0	50	5	α_0	3.015	0.137	0.158	0.535	93
			α_1	0.987	0.166	0.170	0.651	96
			α_2	0.500	0.083	0.089	0.324	94
		10	α_0	2.995	0.112	0.132	0.440	91
			α_1	0.997	0.124	0.125	0.486	94
			α_2	0.503	0.059	0.066	0.230	92
0.0	100	5	α_0	3.003	0.096	0.108	0.377	91
			α_1	1.004	0.115	0.116	0.449	93
			α_2	0.495	0.057	0.061	0.222	94
		10	α_0	2.998	0.080	0.084	0.313	94
			α_1	0.990	0.085	0.088	0.335	92
			α_2	0.499	0.041	0.044	0.160	93
1.0	50	5	α_0	3.021	0.138	0.147	0.541	92
			α_1	0.991	0.168	0.177	0.657	94
			α_2	0.513	0.084	0.085	0.329	92
		10	α_0	3.005	0.113	0.129	0.443	92
			α_1	1.000	0.124	0.124	0.486	95
			α_2	0.501	0.061	0.063	0.238	96
1.0	100	5	α_0	3.001	0.098	0.109	0.386	94
			α_1	1.000	0.118	0.115	0.462	96
			α_2	0.500	0.059	0.059	0.231	97
		10	α_0	2.997	0.083	0.078	0.324	97
			α_1	1.011	0.088	0.082	0.345	96
			α_2	0.503	0.042	0.041	0.165	96

Table 2
The RMISE of the estimated population slope $\hat{\beta}_\ell(t)$ and the estimated individual slope $\hat{\beta}_{i\ell}(t)$ obtained by applying the REML-based EM algorithm on the simulated data with 200 simulation replicates. The true population slope functions are $\beta_1(t) = 1 + 2t^2 + e^{-3t}$ and $\beta_2(t) = 1 + 2\sin(2\pi t) + \cos(2\pi t)$ for $t \in (0, 1)$.

σ_X	n	m_i	Population slope		Individual slope	
			$\beta_1(t)$	$\beta_2(t)$	$\beta_{i1}(t)$	$\beta_{i2}(t)$
0.0	50	5	0.040	0.157	0.635	0.727
		10	0.027	0.090	0.626	0.709
0.0	100	5	0.025	0.090	0.616	0.701
		10	0.017	0.058	0.614	0.691
1.0	50	5	0.041	0.228	0.636	0.737
		10	0.030	0.182	0.624	0.721
1.0	100	5	0.027	0.200	0.623	0.720
		10	0.020	0.144	0.614	0.707

The estimation results for the population and individual functional effects are summarized in Table 2. As expected, there is a substantial decrease in RMISEs of the estimated population slopes $\hat{\beta}_\ell(t)$ with more replicates per subject. For instance, RMISEs of $\hat{\beta}_1(t)$ and $\hat{\beta}_2(t)$ decrease 32.5% and 42.7%, respectively, when the number of replicates per subject m_i increases from 5 to 10 in the scenario $\sigma_X = 0.0$ and $n = 50$. On the other hand, RMISEs of the estimated individual slopes $\hat{\beta}_{i\ell}(t)$ have little decrease with more replicates per subject.

4. Applications

In this section, we perform the afore-proposed FLMMs via the EM algorithm to analyze two real datasets.

4.1. The NMMAPS data analysis

First, we re-visit the air pollution study introduced in Section 1. The functional linear mixed-effects model (2) is used to study the effect of daily ozone concentration on annual nonaccidental mortality rates for population between 65 and

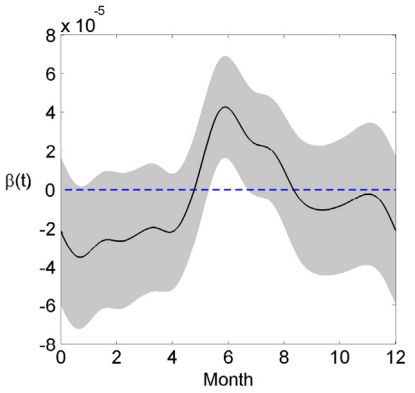


Fig. 2. The estimated population slope functions $\hat{\beta}(t)$, which represent the effect of the daily ozone concentration on the annual nonaccidental mortality rate. The shaded area indicates the pointwise 95% confidence intervals for $\hat{\beta}(t)$.

74 years old in US cities. We fit the following functional linear mixed-effect model

$$Y_{ij} = \alpha + \gamma_i + \int_0^{365} [\beta(t) + b_i(t)]X_{ij}(t)\mathrm{d}t + \epsilon_{ij}, \tag{14}$$

where Y_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, m_i$ is the logarithm of the yearly nonaccidental mortality rates in the ith city and the jth year, and $X_{ij}(t)$ is the respective daily ozone concentration. The scalar random effect γ_i is assumed to follow a $Normal(0, \sigma_\gamma^2)$ distribution, and the measurement error ϵ_{ij} is assumed to follow $Normal(0, \sigma_\epsilon^2)$ distribution. The data are collected for $n = 63$ cities in the USA, from 1987 to 2000.

We use the Fourier basis functions for the population slope function $\beta(t)$ and for the random slope function $b_i(t)$ to account for their annual periodic effect. We follow the suggestion made by Ramsay and Silverman (2005) and use the harmonic acceleration operator to define the roughness penalty for the population slope function $\beta(t)$. The harmonic acceleration operator is defined by $L\beta(t) = \mathrm{d}^3\beta(t)/\mathrm{d}t^3 + \omega^2\mathrm{d}\beta(t)/\mathrm{d}t$, where $\omega = 2\pi/365$ is the period of the nonparametric function. Therefore, zero roughness implies that $\beta(t)$ is of the form $\beta(t) = a_1 + a_2\sin(\omega t) + a_3\cos(\omega t)$. The harmonic acceleration operator is also used to define the roughness penalty of the individual random effect $b_i(t)$. We implement the REML-based EM algorithm proposed in Section 2.2. The estimate for the intercept α is $\hat{\alpha} = -6.41$ with the estimated standard error 0.06, and the 95% confidence interval of α is $[-6.54, -6.29]$.

Fig. 2 displays the estimate population slope function $\hat{\beta}(t)$ and the corresponding approximate 95% pointwise confidence interval. We can see that the ozone concentration has a significantly positive effect on the annual nonaccidental mortality rate only in summer. This implies that a high ozone concentration in summer increases the mortality rate.

Fig. 3 displays the estimated individual slope function $\hat{\beta}_i(t) = \hat{\beta}(t) + \hat{b}_i(t)$ for four cities: Los Angeles, Boston, Honolulu and Kansas City. It is interesting to see that the individual slope functions $\hat{\beta}_i(t)$ for the four cities have a large variation. For instance, $\hat{\beta}_i(t)$ for Los Angeles is higher than the population slope function in the entire year, because Los Angeles have a mild weather, and the mortality rate in Los Angeles increases when the ozone concentration is higher. On the contrary, the individual slope function $\hat{\beta}_i(t)$ for Boston is lower than the population slope function $\hat{\beta}(t)$ in the entire year, because Boston usually has cold winters. When a winter in Boston has a high ozone concentration, it means that Boston has a large amount of sunshine and a warm winter, then the mortality rate in Boston is smaller.

In the functional mixed-effect model (14), we only consider the impact of ozone concentration on nonaccidental mortality, however, it is well known that other air pollutants also have impacts on health. Furthermore, other scalar factors such as medical budget, hospital environment and incomes may also have an effect on mortality. It is an open problem to add more factors in the functional mixed-effect model (14).

4.2. Weather data analysis

In this study, for each year, we explore the effect of daily temperature on annual precipitation. We use a dataset which catalogs annual precipitation, as well as daily temperature measurements for 38 Canadian weather stations from 1961 to 1991. Since most of the data from 1979 are missing, we do not use any data from the year 1979. The functional linear mixed-effects model of interest is

$$Y_{ij} = \alpha + \gamma_i + \int_0^{365} [\beta(t) + b_i(t)]X_{ij}(t)\mathrm{d}t + \epsilon_{ij}, \quad \gamma_i \sim N(0, \sigma_\gamma^2),\ \epsilon_{ij} \sim N(0, \sigma_\epsilon^2),$$

where Y_{ij} is the logarithm of annual precipitation at the ith weather station in the jth year, and $X_{ij}(t)$ is the daily temperature profile for $i = 1, \ldots, 38, j = 1, \ldots, 30$.

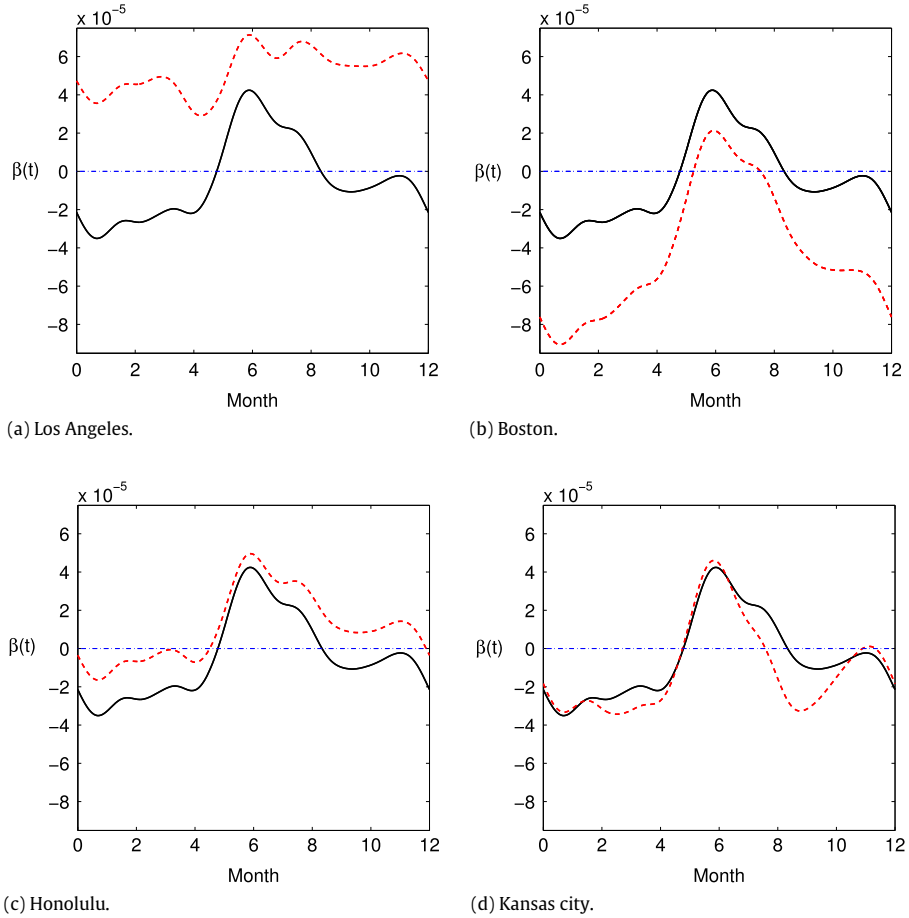


Fig. 3. The estimated individual slope functions $\hat{\beta}_i(t) = \hat{\beta}(t) + \hat{b}_i(t)$ for four cities: Los Angeles, Boston, Honolulu, and Kansas City, which represent the effect of the daily ozone concentration on the annual nonaccidental mortality rates. The solid line is the estimated population slope function $\hat{\beta}(t)$, and the dashed line is the individual slope function $\hat{\beta}_i(t)$.

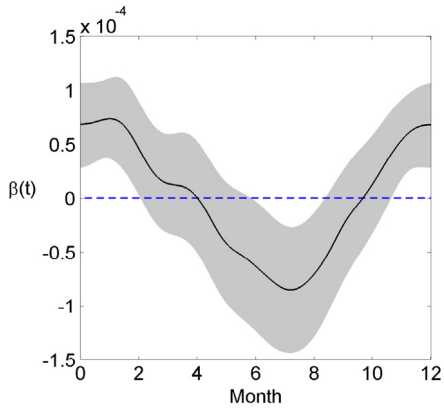
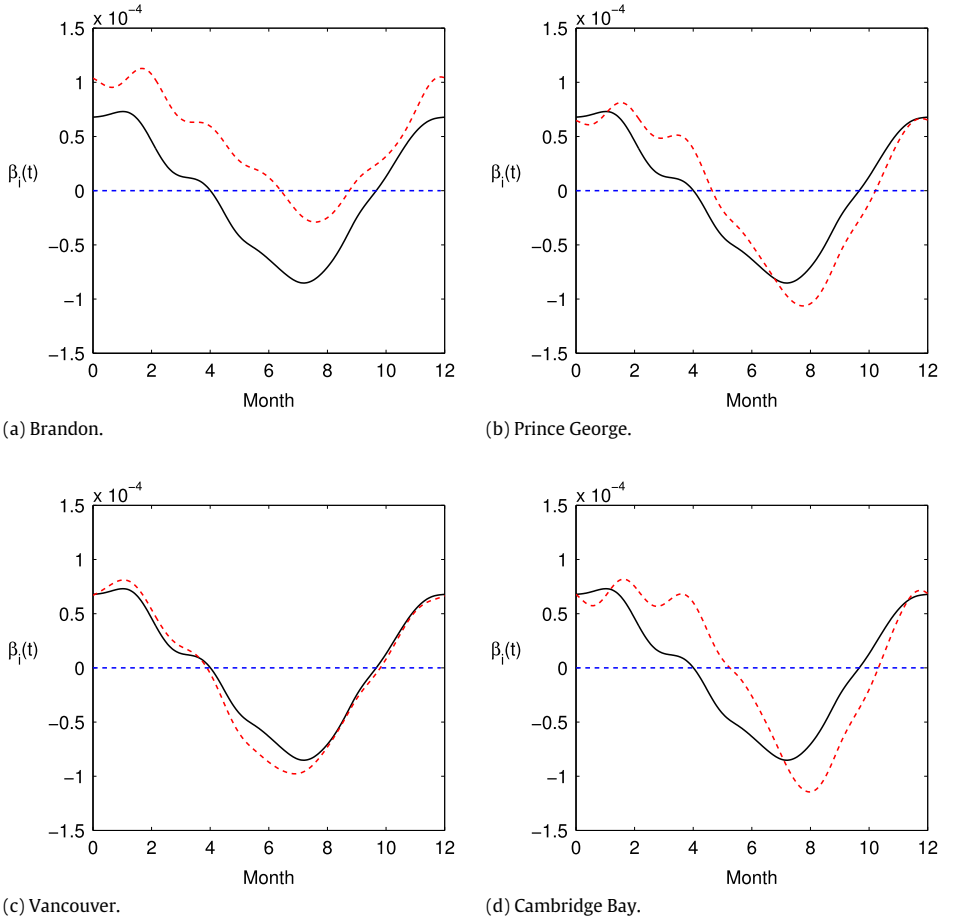


Fig. 4. The estimated population slope function $\hat{\beta}(t)$ for predicting the log total annual precipitation from the daily temperature in Canada. The shaded area indicates the pointwise 95% confidence intervals for $\hat{\beta}(t)$.

Due to the periodicity of weather data, we represent the population slope function $\beta(t)$ using 35 Fourier basis functions and represent the individual slope function $b_i(t)$ using 25 Fourier basis functions. We also use the harmonic acceleration operator to define the roughness penalty for the population slope function $\beta(t)$, and the individual random effect $b_i(t)$. We implement the REML-based EM algorithm proposed in Section 2.2. The estimate for the intercept α is $\hat{\alpha} = 2.98$, with standard error 0.05, and a 95% confidence interval for α is [2.88, 3.09].



(a) Brandon. (b) Prince George.

(c) Vancouver. (d) Cambridge Bay.

Fig. 5. The estimated individual slope function $\hat{\beta}_i(t) = \hat{\beta}(t) + \hat{b}_i(t)$ for predicting the log total annual precipitation from the daily temperature for four cities: Brandon, Prince George, Vancouver, Cambridge Bay in Canada. The solid line is the estimated population slope function $\hat{\beta}(t)$, and the dashed line is the individual slope function $\hat{\beta}_i(t)$.

Fig. 4 displays the estimated population slope function $\beta(t)$ and the 95% pointwise confidence intervals. It indicates that the temperature in the winter has a significant and positive effect on the annual precipitation. The temperature in the summer has a negative effect on the annual precipitation, but this effect is only marginally significant.

We also plot the individual slope function $\hat{\beta}_i(t) = \hat{\beta}(t) + \hat{b}_i(t)$ for four stations in Fig. 5. There are some readily apparent individual variations from the population slope function for each station. For example, we see significant variation in Brandon, which is located in western Manitoba on the banks of the Assiniboine River. Fig. 5 shows that the individual slope function of Brandon is higher than the population slope function for the entire year. This is because compared to other Canadian cities, Brandon is further south and enjoys a large amount of precipitation for most of the year. We note that this phenomenon could not be inferred using the regular functional linear regression model.

5. Conclusions

The functional linear regression model (1) is a popular tool for analyzing the relationship between a scalar response and a functional covariate. However, when there are repeated measurements on multiple subjects, using a single slope function can be too strict of an assumption. In this article, we propose a new version of functional linear mixed-effect model (2), which is more flexible in the sense that each subject has their own slope function while all subjects share a population slope function.

The population and random slope functions are estimated using a penalized spline smoothing method, in which the roughness of the slope functions are controlled by a penalty function. The variance parameters for the random slope function and the data noise are estimated by a REML-based EM algorithm. Our simulation studies illustrate that our estimation method can provide accurate estimates for the functional linear mixed-effect model.

The functional linear mixed-effect model is demonstrated using two real applications. The first application uses the functional linear mixed-effects model (2) to study the effect of daily ozone concentration on annual nonaccidental mortality

rate, and reveals some interesting results. For example, the ozone concentration only in summer has a significantly positive effect on the annual nonaccidental mortality rate. The individual slope functions $\hat{\beta}_i(t)$ for different cities have a large variation. A similar phenomenon is also observed in our second application, which investigates the effect of daily temperature on annual precipitation.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.csda.2016.09.009.

References

Ash, R.B., Gardner, M.F., 1975. Topics in Stochastic Processes. In: Probability and Mathematical Statistics, vol. 27. Academic Press, New York.
Cai, T.T., Hall, P., 2006. Prediction in functional linear regression. Ann. Statist. 34, 2159–2179.
Crambes, C., Kneip, A., Sarda, P., 2009. Smoothing splines estimators for functional linear regression. Ann. Statist. 37, 35–72.
Ferraty, F., Vieu, P., 2006. Nonparametric Functional Data Analysis: Methods, Theory, Applications and Implementations. Springer-Verlag, London.
Gertheiss, J., Goldsmith, J., Crainiceanu, C., Greven, S., 2013. Longitudinal scalar-on functions regression with application to tractography data. Biostatistics 14, 447–461.
Goldsmith, J., Crainiceanu, C.M., Caffo, B., Reich, D., 2012. Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements. J. R. Stat. Soc. Ser. C. Appl. Stat. 61, 453–469.
Goldsmith, J., Wand, M.P., Crainiceanu, C., 2011. Functional regression via variational bayes. Electron. J. Stat. 5, 572–602.
Morris, J.S., 2015. Functional regression. Stat. Appl. 2, 321–359.
Peng, R., Welty, L., 2004. The nmmapsdata package. R News 4, 10–14.
Ramsay, J.O., Dalzell, C.J., 1991. Some tools for functional data analysis. J. R. Stat. Soc. Ser. B Stat. Methodol. 53, 539–572.
Ramsay, J.O., Silverman, B.W., 2002. Applied Functional Data Analysis. Springer, New York.
Ramsay, J.O., Silverman, B.W., 2005. Functional Data Analysis, second ed. Springer, New York.
Rice, J.A., Silverman, B.W., 1991. Estimating the mean and covariance structure nonparametrically when the data are curves. J. R. Stat. Soc. Ser. B Stat. Methodol. 53, 233–243.
Scheipl, F., Staicu, A.M., Greven, S., 2015. Functional additive mixed models. J. Comput. Graph. Statist. 24, 477–501.
Staniswalis, J.G., Lee, J.J., 1998. Nonparametric regression analysis of longitudinal data. J. Amer. Statist. Assoc. 93, 1403–1418.
Wu, Y.C., Fan, J.Q., Müller, H.G., 2010. Varying-coefficient functional linear regression. Bernoulli 16, 730–758.
Wu, H.L., Zhang, J.T., 2006. Nonparametric Regression Methods for Longitudinal Data Analysis: Mixed-Effects Modeling Approaches. Wiley, New York.
Yao, F., Müller, H., Wang, J., 2005. Fucntional linear regression analysis for longitudinal data. Ann. Statist. 33, 2873–2903.
Yuan, M., Cai, T.T., 2010. A reproducing kernel Hilbert space approach to functional linear regression. Ann. Statist. 38, 3412–3444.