

# Estimating Functional Linear Mixed-Effects Regression Models

Baisen Liu<sup>a</sup>, Jiguo Cao<sup>b,\*</sup>

<sup>a</sup>*School of Statistics, Dongbei University of Finance and Economics, Dalian 116025, China*

<sup>b</sup>*Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC V5A1S6, Canada*

---

## Abstract

The functional linear model is a popular tool to investigate the relationship between a scalar/functional response variable and a scalar/functional covariate. We generalize this model to a functional linear mixed-effects model when repeated measurements are available on multiple subjects. Each subject has an individual intercept and slope function, while shares common population intercept and slope function. This model is flexible in the sense of allowing the slope random effects to change with the time. We propose a penalized spline smoothing method to estimate the population and random slope functions. A REML-based EM algorithm is developed to estimate the variance parameters for the random effects and the data noise. Simulation studies show that our estimation method provides an accurate estimate for the functional linear mixed-effects model with the finite samples. The functional linear mixed-effects model is demonstrated by investigating the effect of the 24-hour nitrogen dioxide on the daily maximum ozone concentrations and also studying the effect of the daily temperature on the annual precipitation.

*Keywords:* EM algorithm, Functional Linear Regression, Penalized Splines, Random Effects Model

---

\*Corresponding email: jiguo\_cao@sfu.ca

## 1. Introduction

When a random variable is measured or observed at multiple time points or spatial locations, the data can be viewed as a function of time or spatial locations. This type of data is generally called as functional data (Ramsay and Silverman, 2005). In the current big data era, functional data analysis (FDA) has become very popular in statistical methodology and applied data analysis. Functional linear models (FLMs) is one of the most popular models in FDA. It models the relationship between functional variables and/or predicts the scalar response from the functional input. FLMs have been studied extensively since Ramsay and Dalzell (1991) introduced them. With the developments of modern technology, FLMs have been popularly applied to model functional data in many fields such as economics, medicine, environment, climate [see for instance, Ramsay and Silverman (2002), Ramsay and Silverman (2005), and Ferraty and Vieu (2006), for several case studies].

There is extensive literature studying estimations and properties of FLMs. For example, Chiou et al. (2003) applied a quasi-likelihood approach to study a FLM with a functional response and a finite-dimensional vector of scalar predictors. Yao et al. (2005) studied FLMs for sparse longitudinal data and suggested a nonparametric estimation method based on the functional principal components analysis (FPCA). Their proposed functional regression approach is flexible to allow for different measurement time points of functional predictors and the functional 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 it 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 on simultaneous diagonalization of two positive-definite kernels. Wu et al. (2010) proposed a varying-coefficient FLM which allows for the slope function depend-

ing on some additional scalar covariates. A systematic review on FLMs can be found in Morris (2015).

One popular FLM is to link a scalar response variable  $Y_j, j = 1, \dots, m$ , with a functional predictor  $X_j(t)$  through the following model

$$Y_j = \alpha + \int_S \beta(t)X_j(t)dt + \epsilon_j, \quad (1)$$

where  $\alpha$  is the intercept,  $\beta(t)$  is a smooth slope function,  $\epsilon_j$ 's are independent and identically distributed (i.i.d.) random variables with mean 0 and variance  $\sigma_\epsilon^2$ , and  $S$  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). The data have hourly measurements of ozone and nitrogen dioxide  $NO_2$  concentrations for some U.S. cities. Our aim is to study the relationship of the daily maximum ozone concentration and the functional predictors nitrogen dioxide  $NO_2(t)$  measured during 24 hours (from 0 am to 11 pm) of that day. For the  $i$ -th city, the scalar response  $Y_{ij}$  is the maximum ozone concentration during 24 hours (from 0 am to 11 pm) in the  $j$ -th day, and the functional covariate  $X_{ij}(t)$  is the hourly  $NO_2(t)$  concentration measured during 24 hours. In a preliminary analysis, we performed a functional linear regression model (Cardot et al., 2007) on each individual city and found that there was a dramatic variation of the estimated  $\hat{\beta}(t)$ . This indicates that each city has different effects of the hourly  $NO_2(t)$  concentration on the daily maximum ozone. Therefore, it may not be appropriate to pool all the data of U.S. cities together and provide only one average effect of hourly  $NO_2(t)$  on the daily maximum ozone. On the other hand, we may not use all of the data information available if we fit the functional linear model for each individual city separately.

To address this dilemma, we generalize the FLM (1) to incorporate random effects into the slope function, and call it the functional linear mixed-effects model (FLMM). Assume that we repeatedly observed a distinct functional pre-

dictor and scalar outcome for each subject over several visits. Then the observed data has the structure  $\{Y_{ij}, X_{ij}(t)\}$  for  $1 \leq i \leq n$  and  $1 \leq j \leq m_i$ , where  $Y_{ij}$  is the  $j$ -th repeated measurement of the scalar response for the  $i$ -th subject, and  $X_{ij}(t)$  is the corresponding functional predictor. The functional linear mixed-effects model can be expressed as following:

$$Y_{ij} = \alpha_0 + a_i + \int_S [\beta(t) + b_i(t)] X_{ij}(t) dt + \epsilon_{ij}, \quad i = 1, 2, \dots, n, j = 1, 2, \dots, m_i, \quad (2)$$

where  $\alpha_0$  is the population intercept,  $a_i$  is the intercept random effect,  $\beta(t)$  represents the population effect of  $X_{ij}(t)$  on  $Y_{ij}$ ,  $b_i(t)$  stands for the random effect of  $X_{ij}(t)$  on  $Y_{ij}$  for the  $i$ -th subject, and  $\epsilon_{ij}$  is the i.i.d. random variable with mean 0 and variance  $\sigma_\epsilon^2$ . In this article, we assume that  $a_i \sim N(0, \sigma_a^2)$ ,  $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$ , and  $b_i(t)$  follows a Gaussian stochastic process with mean 0 and covariance function  $\gamma(s, t)$ , that is,  $b_i(t) \sim GP(0, \gamma(s, t))$ . We also assume that  $a_i$ ,  $\epsilon_{ij}$ ,  $b_i(t)$ , and  $X_{ij}(t)$  are mutually independent. The above functional linear mixed-effects model is very attractive, because it can estimate the population effect and random effect of the functional predictor  $X(t)$  (e.g. the hourly  $NO_2(t)$  in the air pollution study) on the scalar response  $Y$  (e.g. daily maximum ozone concentrations) as well as the population intercept and intercept random effect simultaneously. The application of the proposed functional linear mixed-effects model on the air pollution problem is not unique, and many similar applications can be found in environmental or biological problems.

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

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

where  $\mathbf{b}_i \sim N(\mathbf{0}, \sigma_b^2 \mathbf{I})$  accounts for correlations in the repeated outcomes for the  $i$ -th subject. The highlight of the distinction between (2) and (3) is: the subject-specific random effect  $\mathbf{b}_i$  of (3) remains the same across visits, while the random effect  $b_i(t)$  of (2) allows for varying with time. The including of the random effect  $b_i(t)$  in (2) can characterize the different trend effect of functional predictor on scalar outcomes for different subjects.

Many nonparametric smoothers used for the FLMs can be applied to fit the model (2). In this article, we use the idea of penalized splines smoothers of Ramsay and Silverman (2005) to estimate  $\beta(t)$  and  $b_i(t)$  in (2). Then, the model (2) is transformed by a linear mixed-effects model (LMM). Then a REML-based EM algorithm is proposed to fit the LMMs, and its efficiency 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 implemented some simulations to evaluate the finite sample performance of the smoothing spline method. Then the functional linear mixed-effects model is demonstrated by two real applications in Section 4. Conclusions are given in Section 5.

## 2. Method

Without giving any parametric assumption on the slope functions,  $\beta(t)$  and  $b_i(t)$ , we estimate them as linear combinations of splines basis functions

$$\beta(t) = \sum_{j=1}^J c_j \phi_j(t) = \boldsymbol{\phi}'(t) \mathbf{c}, \quad b_i(t) = \sum_{k=1}^K b_{ik} \psi_k(t) = \boldsymbol{\psi}'(t) \mathbf{b}_i.$$

where  $\boldsymbol{\phi}(t) = (\phi_1(t), \dots, \phi_J(t))'$  and  $\boldsymbol{\psi}(t) = (\psi_1(t), \dots, \psi_K(t))'$  are two vectors of basis functions with dimensions  $J$  and  $K$ , respectively, and  $\mathbf{c} = (c_1, \dots, c_J)'$  and  $\mathbf{b}_i = (b_{i1}, \dots, b_{iK})'$  are the corresponding vectors of basis coefficients to estimate. Let  $\mathbf{D} = \text{Cov}(\mathbf{b}_i) = \text{E}(\mathbf{b}_i \mathbf{b}_i')$  denote the variance-covariance matrix of random-effects, then  $\mathbf{b}_i \sim N(0, \mathbf{D})$ , and the covariance function  $\gamma(s, t)$  for the random effect  $b_i(t)$  can be expressed as  $\gamma(s, t) = \boldsymbol{\psi}'(s) \mathbf{D} \boldsymbol{\psi}(t)$ . We first consider the scenario of the functional predictors  $X_{ij}(t)$  observed without measurement errors. When  $X_{ij}(t)$  is observed with measurement errors, many nonparametric smoothing approaches can be applied to reconstruct the underlying functional predictors  $X_{ij}(t)$ , such as the functional principal component analysis method (Yao et al., 2005), which is introduced in Subsection 2.5.

### 2.1. Estimating Fixed and Random Effects

Let  $\boldsymbol{\theta} = (\alpha_0, \mathbf{c}')'$  and  $\boldsymbol{\xi} = (\boldsymbol{\xi}'_1, \dots, \boldsymbol{\xi}'_n)'$  with  $\boldsymbol{\xi}_i = (a_i, \mathbf{b}'_i)'$ ,  $i = 1, \dots, n$ . The fixed effects  $\{\alpha_0, \beta(t)\}$ , and random effects  $\{a_i, b_i(t)\}$  are estimated by minimizing

$$\begin{aligned} H(\boldsymbol{\theta}, \boldsymbol{\xi}) &= \sum_{i=1}^n \sum_{j=1}^{m_i} \frac{1}{2\sigma_\epsilon^2} \left( Y_{ij} - \alpha_0 - a_i - \int_S [\beta(t) + b_i(t)] X_{ij}(t) dt \right)^2 + \frac{1}{2} \sum_{i=1}^n \mathbf{b}'_i \mathbf{D}^{-1} \mathbf{b}_i \\ &\quad + \left[ \frac{\lambda_\beta}{2} \int_S \left\{ \frac{d^2 \beta(t)}{dt^2} \right\}^2 dt + \frac{\lambda_b}{2} \sum_{i=1}^n \int_S \left\{ \frac{d^2 b_i(t)}{dt^2} \right\}^2 dt \right] + \frac{1}{2\sigma_a^2} \sum_{i=1}^n a_i^2. \end{aligned} \quad (4)$$

Define three vectors  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{im_i})'$ ,  $\mathbf{W}_i = [\mathbf{w}_{i1}, \dots, \mathbf{w}_{im_i}]'$  and  $\mathbf{Z}_i = [\mathbf{z}_{i1}, \dots, \mathbf{z}_{im_i}]'$  where  $\mathbf{w}_{ij} = (1, \int_S \phi'(t) X_{ij}(t) dt)'$  and  $\mathbf{z}_{ij} = (1, \int_S \psi'(t) X_{ij}(t) dt)'$ . Define a  $J \times J$  matrix  $\mathbf{G} = \int_S (d^2 \phi(t)/dt^2)(d^2 \phi(t)/dt^2)' dt$  and a  $K \times K$  matrix  $\mathbf{G}_b = \int_S (d^2 \psi(t)/dt^2)(d^2 \psi(t)/dt^2)' dt$ . Then  $H(\boldsymbol{\theta}, \boldsymbol{\xi})$  can be expressed in a matrix form

$$\begin{aligned} H(\boldsymbol{\theta}, \boldsymbol{\xi}) &= \sum_{i=1}^n \frac{1}{2\sigma_\epsilon^2} \|\mathbf{Y}_i - \mathbf{W}_i \boldsymbol{\theta} - \mathbf{Z}_i \boldsymbol{\xi}_i\|^2 + \frac{1}{2} \sum_{i=1}^n \mathbf{b}'_i \mathbf{D}^{-1} \mathbf{b}_i \\ &\quad + \left( \frac{\lambda_\beta}{2} \mathbf{c}' \mathbf{G} \mathbf{c} + \frac{\lambda_b}{2} \sum_{i=1}^n \mathbf{b}'_i \mathbf{G}_b \mathbf{b}_i \right) + \frac{1}{2\sigma_a^2} \sum_{i=1}^n a_i^2. \end{aligned}$$

Then the estimates for  $\boldsymbol{\theta}$  and  $\boldsymbol{\xi}$  are obtained by minimizing  $H(\boldsymbol{\theta}, \boldsymbol{\xi})$ :

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \left( \mathbf{W}' \tilde{\mathbf{V}}^{-1} \mathbf{W} + \lambda \tilde{\mathbf{G}} \right)^{-1} \mathbf{W}' \tilde{\mathbf{V}}^{-1} \mathbf{Y}, \\ \hat{\boldsymbol{\xi}} &= (\mathbf{I}_n \otimes \tilde{\mathbf{D}}_\xi) \mathbf{Z}' \tilde{\mathbf{V}}^{-1} (\mathbf{Y} - \mathbf{W} \hat{\boldsymbol{\theta}}). \end{aligned} \quad (5)$$

where  $\mathbf{Y} = (\mathbf{Y}'_1, \dots, \mathbf{Y}'_n)'$ ,  $\mathbf{W} = (\mathbf{W}'_1, \dots, \mathbf{W}'_n)'$ ,  $\tilde{\mathbf{V}} = \text{diag}(\tilde{\mathbf{V}}_1, \dots, \tilde{\mathbf{V}}_n)$  with  $\tilde{\mathbf{V}}_i = \mathbf{Z}_i \tilde{\mathbf{D}}_\xi \mathbf{Z}'_i + \sigma_\epsilon^2 \mathbf{I}_{m_i}$ ,  $i = 1, \dots, n$ ,  $\tilde{\mathbf{G}} = \text{diag}(0, \mathbf{G})$ ,  $\tilde{\mathbf{D}}_\xi = (\mathbf{D}_\xi^{-1} + \lambda_b \mathbf{G}_\xi)^{-1}$ ,  $\mathbf{D}_\xi = \text{diag}(\sigma_a^2, \mathbf{D})$ ,  $\mathbf{G}_\xi = \text{diag}(0, \mathbf{D}_b)$ ,  $\mathbf{Z} = \text{diag}(\mathbf{Z}'_1, \dots, \mathbf{Z}'_n)$ , and  $\otimes$  denotes the kronecker product.

Once obtaining the estimates  $\hat{\boldsymbol{\theta}} = (\hat{\alpha}_0, \hat{\mathbf{c}})'$  and  $\hat{\boldsymbol{\xi}}_i = (\hat{a}_i, \hat{\mathbf{b}}'_i)'$ , the estimates of  $\beta(t)$  and  $b_i(t)$ ,  $i = 1, \dots, n$ , can be given by

$$\hat{\beta}(t) = \phi'(t) \hat{\mathbf{c}}, \quad \hat{b}_i(t) = \psi'(t) \hat{\mathbf{b}}_i, \quad i = 1, \dots, n. \quad (6)$$

## 2.2. The REML-based EM algorithm

To estimate the fixed-effects,  $\boldsymbol{\theta}$ , the random-effects,  $\boldsymbol{\xi}$ , and the variance parameters,  $\sigma_a^2$ ,  $\sigma_\epsilon^2$  and  $\mathbf{D}$ , we recommend an EM algorithm procedure called the *REML-based EM-algorithm*. It was proposed by Wu and Zhang (2006) for estimating nonparametric mixed-effects regression models with longitudinal data. The REML-based EM-algorithm has three steps, which are outlined as follow.

**Initializing.** Initializing the starting values for  $\sigma_a^2$ ,  $\sigma_\epsilon^2$  and  $\mathbf{D}$ , denoted by  $\sigma_a^{2(0)}$ ,  $\sigma_\epsilon^{2(0)}$  and  $\mathbf{D}^{(0)}$ , respectively. For example, we can choose  $\sigma_a^{2(0)} = \sigma_\epsilon^{2(0)} = 1$  and  $\mathbf{D}^{(0)}$  as an identity matrix.

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

$$\begin{aligned}\tilde{\mathbf{D}}_\xi^{(r-1)} &= [\{\mathbf{D}_\xi^{(r-1)}\}^{-1} + \lambda_b \mathbf{G}_\xi]^{-1}, \\ \tilde{\mathbf{V}}_i^{(r-1)} &= \mathbf{Z}_i' \tilde{\mathbf{D}}_\xi^{(r-1)} \mathbf{Z}_i + \sigma_\epsilon^{2(r-1)} \mathbf{I}_{m_i}, i = 1, 2, \dots, n.\end{aligned}$$

Denote  $\tilde{\mathbf{V}}^{(r-1)} = \text{diag}(\tilde{\mathbf{V}}_1^{(r-1)}, \dots, \tilde{\mathbf{V}}_n^{(r-1)})$ . Then estimate  $\hat{\boldsymbol{\theta}}^{(r)}$  and  $\hat{\boldsymbol{\xi}}_i^{(r)}$  by

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

**Step 2.** Compute the residuals  $\hat{\boldsymbol{\epsilon}}_i^{(r)} = \mathbf{Y}_i - \mathbf{W}_i \hat{\boldsymbol{\theta}}^{(r)} - \mathbf{Z}_i \hat{\boldsymbol{\xi}}_i^{(r)}$  and the matrix  $\mathbf{H}_i^{(r-1)} = \{\tilde{\mathbf{V}}_i^{(r-1)}\}^{-1} - \{\tilde{\mathbf{V}}_i^{(r-1)}\}^{-1} \mathbf{W}_i [\mathbf{W}'\{\tilde{\mathbf{V}}^{(r-1)}\}^{-1} \mathbf{W} + \lambda_\beta \tilde{\mathbf{G}}]^{-1} \mathbf{W}_i' \{\tilde{\mathbf{V}}_i^{(r-1)}\}^{-1}$ . Then the updates of  $\sigma_\epsilon^{2(r)}$  and  $\mathbf{D}_\xi^{(r)} = \text{diag}(\sigma_a^{2(r)}, \mathbf{D}^{(r)})$  are given by

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

**Step 3.** Repeat Steps 2 and 3, until some convergence conditions are satisfied.

## 2.3. Smoothing Parameter Selection

The smoothness of  $\beta(t)$  and  $b_i(t)$ ,  $i = 1, \dots, n$ , are controlled by the smoothing parameter  $\lambda_\beta$  and  $\lambda_b$ , respectively. Define  $\boldsymbol{\lambda} = (\lambda_\beta, \lambda_b)'$ , then the optimal

value for  $\boldsymbol{\lambda}$  is chosen by minimizing the generalized cross-validation (GCV) criterion defined as follows

$$\text{GCV}(\boldsymbol{\lambda}) = \frac{\text{SSE}(\boldsymbol{\lambda})}{(N - df(\boldsymbol{\lambda}))^2},$$

where  $\text{SSE}(\boldsymbol{\lambda}) = \sum_{i=1}^n \|\mathbf{Y}_i - \mathbf{W}_i \hat{\boldsymbol{\theta}} - \mathbf{Z}_i \hat{\boldsymbol{\xi}}_i\|^2$ ,  $N = \sum_{i=1}^n m_i$ , and  $df(\boldsymbol{\lambda})$  is the effective degrees of freedom, which is calculated as  $df(\boldsymbol{\lambda}) = \text{trace}(\mathbf{Q})$ , where  $\mathbf{Q}$  is given by

$$\mathbf{Q} = (\mathbf{W}, \mathbf{Z}) \left[ \frac{1}{\sigma_\epsilon^2} \begin{pmatrix} \mathbf{W}'\mathbf{W} & \mathbf{W}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{W} & \mathbf{Z}'\mathbf{Z} \end{pmatrix} + \begin{pmatrix} \lambda_\beta \tilde{\mathbf{G}} & 0 \\ 0 & \mathbf{I}_n \otimes \tilde{\mathbf{D}}_\xi \end{pmatrix} \right]^{-1} \begin{pmatrix} \mathbf{W}' \\ \mathbf{Z}' \end{pmatrix}.$$

Define the matrix  $\mathbf{S}_i = \mathbf{W}_i [\mathbf{W}' \tilde{\mathbf{V}}^{-1} \mathbf{W} + \lambda_\beta \tilde{\mathbf{G}}]^{-1} \mathbf{W}' \tilde{\mathbf{V}}^{-1}$ . The predictor  $\hat{\mathbf{Y}}_i = \mathbf{W}_i \hat{\boldsymbol{\theta}} + \mathbf{Z}_i \hat{\boldsymbol{\xi}}_i$  can then be expressed as  $\hat{\mathbf{Y}}_i = \mathbf{Q}_i \mathbf{Y}$  with

$$\mathbf{Q}_i = \mathbf{S}_i + \mathbf{Z}_i \tilde{\mathbf{D}}_\xi \mathbf{Z}_i' \tilde{\mathbf{V}}^{-1} (\mathbf{I}_{m_i} - \mathbf{S}_i).$$

Note that the smooth matrix  $\mathbf{Q} = (\mathbf{Q}'_1, \dots, \mathbf{Q}'_n)'$ .

#### 2.4. Constructing the Confidence Intervals

To construct the confidence intervals of  $\alpha_0$  and the point-wise confidence bands of  $\beta(t)$ , we need to calculate the covariance matrix of  $\hat{\boldsymbol{\theta}}$ :

$$\begin{aligned} \text{Cov}(\hat{\boldsymbol{\theta}}) &= \left( \sum_{i=1}^n \mathbf{W}'_i \tilde{\mathbf{V}}_i^{-1} \mathbf{W}_i + \lambda_\beta \tilde{\mathbf{G}} \right)^{-1} \left( \sum_{i=1}^n \mathbf{W}'_i \tilde{\mathbf{V}}_i^{-1} \text{Cov}(\mathbf{Y}_i) \tilde{\mathbf{V}}_i^{-1} \mathbf{W}_i \right) \\ &\quad \left( \sum_{i=1}^n \mathbf{W}'_i \tilde{\mathbf{V}}_i^{-1} \mathbf{W}_i + \lambda_\beta \tilde{\mathbf{G}} \right)^{-1}. \end{aligned} \quad (7)$$

In (7),  $\text{Cov}(\mathbf{Y}_i)$  can be replaced by  $\tilde{\mathbf{V}}_i$  to account for our roughness penalty on  $b_i(t)$ ,  $i = 1, \dots, n$ . For simplicity, instead of using  $\text{Cov}(\hat{\boldsymbol{\theta}})$ , we use

$$\text{Cov}(\hat{\boldsymbol{\theta}}) = \left( \sum_{i=1}^n \mathbf{W}'_i \tilde{\mathbf{V}}_i^{-1} \mathbf{W}_i + \lambda_\beta \tilde{\mathbf{G}} \right)^{-1}. \quad (8)$$

Let  $\widehat{\text{Cov}}(\hat{\boldsymbol{\theta}})$  be the estimator of the covariance matrix (8) and partition it as  $\widehat{\text{Cov}}(\hat{\boldsymbol{\theta}}) = \begin{pmatrix} \hat{\sigma}_{11}^2 & \hat{\boldsymbol{\Sigma}}_{12} \\ \hat{\boldsymbol{\Sigma}}'_{12} & \hat{\boldsymbol{\Sigma}}_{22} \end{pmatrix}$ . Then the 95% confidence intervals of  $\alpha_0$  is approximately as

$$(\hat{\alpha}_0 - 1.96\hat{\sigma}_{11}, \hat{\alpha}_0 + 1.96\hat{\sigma}_{11}),$$



and the 95% pointwise bands of  $\beta(t)$  can be approximately given by

$$\left( \hat{\beta}(t) - 1.96\sqrt{\widehat{\text{Var}}[\hat{\beta}(t)]}, \hat{\beta}(t) + 1.96\sqrt{\widehat{\text{Var}}[\hat{\beta}(t)]} \right), \text{ for all } t \in S,$$

where  $\widehat{\text{Var}}[\hat{\beta}(t)] = \Phi'(t)\hat{\Sigma}_{22}\Phi(t)$ . Moreover, the estimate of  $\gamma(s, t)$  can be given as

$$\hat{\gamma}(s, t) = \Psi'(s)\hat{\mathbf{D}}_b\Psi(t),$$

where we use  $\hat{\mathbf{D}}_b = (\mathbf{0}, \mathbf{I}_K) \left( \hat{\mathbf{D}}_\xi^{-1} + \lambda_b \mathbf{G}_\xi \right)^{-1} (\mathbf{0}, \mathbf{I}_K)'$  instead of  $\hat{\mathbf{D}}$  in order to account for our roughness penalty on  $b_i(t), i = 1, \dots, n$  in our method.

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

When covariates  $X_{ij}(t)$  in the functional linear mixed-effects model (2) are not be exactly observable but measured with errors, the estimators and inference may be biased if one ignores these measurement errors. Hence, we need to adjust the resulting bias. In this article, we suggest to reconstruct the functional predictors  $X_{ij}(t)$  by using a large number of functional principal components obtained from a smooth estimator of the covariance matrix estimator (Goldsmith et al., 2012) firstly. Then, we treat the estimated  $\hat{X}_{ij}(t)$  as the true predictors and applying the REML-based EM algorithm.

Define the covariance function of  $X(t)$  as

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

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

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

where  $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$  satisfying  $\sum_{k=1}^{\infty} \lambda_k < \infty$ , and  $\varphi_k(t)$ 's form a complete orthonormal basis in  $S \times S$ . Then,  $X(t)$  allows 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, which is also called the functional principal component (FPC). The coefficients  $\xi_k$  is called the FPC score of  $X(t)$ , which satisfies  $E(\xi_k) = 0$ ,  $E(\xi_k^2) = \lambda_k$ , and  $E(\xi_k \xi_l) = 0$  for  $k \neq l$ .

Suppose we have the following additive measurement error model,

$$W_{ij}(t) = X_{ij}(t) + e_{ij}(t),$$

where  $W_{ij}(t)$  is the observed value,  $X_{ij}(t)$  is the underlying true value for the  $i$ th subject at the time point  $t$ , and  $e_{ij}(t)$  represents the measurement error at the time point  $t$ . We assume that  $e_{ij}(t)$  is a mean zero process, and  $\{X_{ij}(t), e_{ij}(t)\}$  are mutually independent. We estimate  $C(s, t)$  by using a method-of-moments approach, and then smooth the off-diagonal elements of this 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  $\mu_i(t)$ , the FPCs  $\varphi_{ik}(t)$  and the FPC scores  $\xi_{ijk}$  from the observations  $W_{ij}(t_{ijk})$ . Let  $\hat{\mu}_i(t)$ ,  $\hat{\varphi}_{ik}(t)$ , and  $\hat{\xi}_{ijk}$  be the corresponding estimators of  $\mu_i(t)$ ,  $\varphi_{ik}(t)$  and  $\xi_{ijk}$ , respectively. Then an estimate of  $X_{ij}(t)$  is obtained as

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

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

### 3. Simulation studies

In this section, we perform some numerical experiments to assess the efficiency of our proposed estimating procedure for the functional linear mixed-effects model (2). The performance of our estimation method is evaluated by the following relative mean integrated square error (RMISE) for the estimated

population slope function  $\hat{\beta}(t)$  and the individual slope function  $\hat{\beta}_1(t), \dots, \hat{\beta}_n(t)$ ,

$$\text{RMISE}(\hat{\beta}(t)) = \frac{\int_S (\hat{\beta}(t) - \beta(t))^2 dt}{\int_S \beta^2(t) dt},$$

and

$$\text{RMISE}(\hat{\beta}_1(t), \dots, \hat{\beta}_n(t)) = \frac{\sum_{i=1}^n \int_S (\hat{\beta}_i(t) - \beta_i(t))^2 dt}{\sum_{i=1}^n \int_S \beta_i^2(t) dt},$$

where  $\beta_i(t) = \beta(t) + b_i(t)$ , and  $\hat{\beta}_i(t) = \hat{\beta}(t) + \hat{b}_i(t)$ ,  $i = 1, \dots, n$ .

We assume that the functional predictor  $X_{ij}(t)$  are observed at  $n$  equally-spaced time points  $\{t_k, k = 1, \dots, n\}$  of  $[0, 1]$  with the additive normal measurement errors:

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

where  $\sigma_e = 0.0$  or  $0.5$ , and the true underlying predictors  $X_{ij}(t)$  are given by

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

where  $\mu_i(t) = \delta_{i0} + \delta_{i1} \sin(\pi t)$  with independent random variables  $\delta_{i0} \sim U[-2, 2]$ ,  $\delta_{i1} \sim N(0, 4)$ ,  $\xi_{ijk} \sim N(0, 2/2^k)$ ,  $k = 1, 2, 3, 4$ , and  $\psi_1(t) = \sin(2\pi t)$ ,  $\psi_2(t) = \cos(2\pi t)$ ,  $\psi_3(t) = \sin(4\pi t)$  and  $\psi_4(t) = \cos(4\pi t)$ . We choose two types of functions for the individual slope functions  $\beta_i(t)$ : (1)  $\beta_i(t) = \eta_{0i} + \eta_{1i} t^2 + \eta_{2i} \exp(-3t)$  with the population slope function given by  $\beta(t) = 1 + 2t^2 + \exp(-3t)$ ; (2)  $\beta_i(t) = \eta_{0i} + \eta_{1i} \sin(2\pi t) + \eta_{2i} \cos(2\pi t)$  with the population slope function given by  $\beta(t) = 1 + 2 \sin(2\pi t) + \cos(2\pi t)$ . The random coefficients are generated as  $(\eta_{0i}, \eta_{1i}, \eta_{2i})' \sim N((1.0, 2.0, 1.0)', \text{diag}(0.2^2, 0.4^2, 0.2^2))$  for both of cases. The scalar response is generated from the following model:

$$Y_{ij} = \alpha_i + \int_0^1 \beta_i(t) X_{ij}(t) dt + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma_e^2), \quad i = 1, 2, \dots, n, j = 1, 2, \dots, m_i,$$

where  $\alpha_i$  is generated from  $N(3.0, 0.25)$ . The number of repeated measurements for each individual is varied as  $m_i = 5, 10, 20$ . The number of individuals is set as  $n = 50$  and  $100$ , and the standard deviation of the data noises is varied as  $\sigma_e = 0.5$  and  $1.0$ .

The functional linear mixed-effects model (2) is estimated using the method introduced in Section 2. In case 1, we choose 35 cubic B-splines basis functions

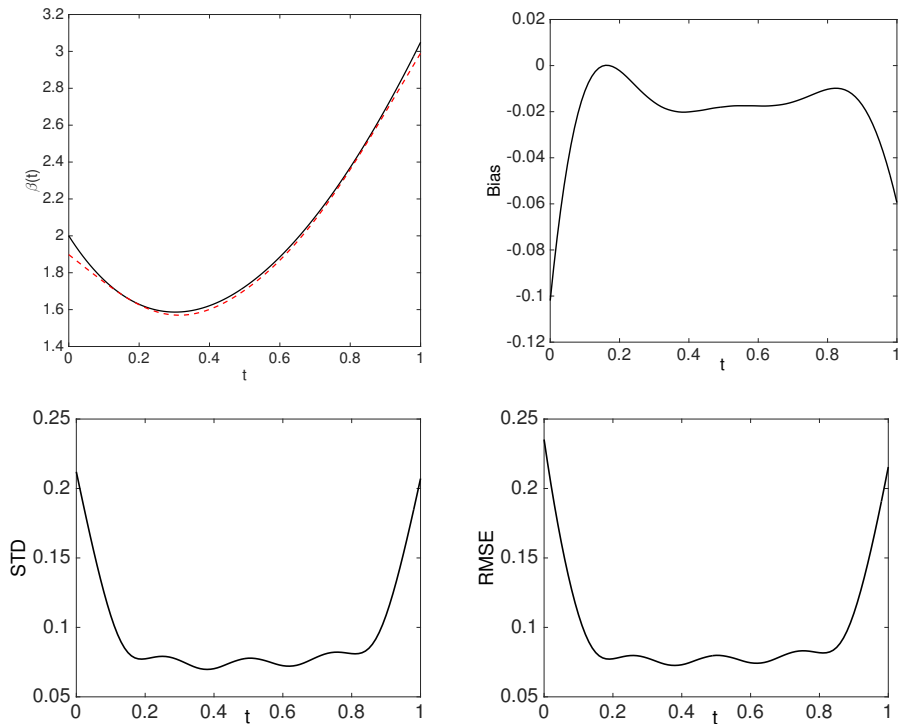


Figure 1: The pointwise mean, bias, standard deviation (STD) and root mean squared error (RMSE) of the estimated population slope function  $\hat{\beta}(t)$  in 1,000 simulation replicates when  $m_i = 20$  and  $\sigma_\epsilon = 1.0$  in our simulation studies. The dashed line in the top left panel is the true population slope function  $\beta(t) = 1 + 2t^2 + \exp(-3t)$ .

for  $\beta(t)$  and  $b_i(t)$ ; while in case 2, we choose 35 Fourier basis functions for  $\beta(t)$  and  $b_i(t)$ . Figure 1 and 2 display the pointwise mean, bias, standard deviation and root mean squared error of the estimated population slope function  $\beta(t)$  in 1,000 simulation replicates when  $(m_i, n) = (20, 100)$  and  $(\sigma_e, \sigma_\epsilon) = (0.5, 1.0)$ . It shows that the pointwise mean of the estimated population slope function  $\hat{\beta}(t)$  is very close to the true function  $\beta(t)$  in both of cases.

The estimation results for all simulation setups are summarized in Table 1-2. As expected, there is a substantial decrease in RMISE when more visits are observed per subject. When the functional covariate  $X(t)$  is observed directly without measurement errors (i.e.  $\sigma_e = 0$ ), the mean and median of RMISE for

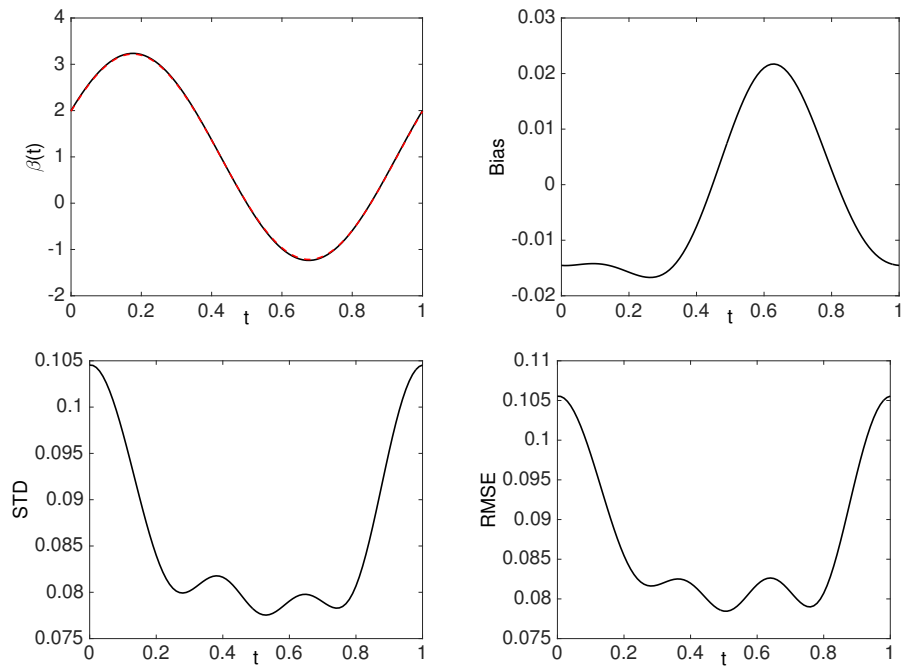


Figure 2: The pointwise mean, bias, standard deviation (STD) and root mean squared error (RMSE) of the estimated population slope function  $\hat{\beta}(t)$  in 1,000 simulation replicates when  $m_i = 20$  and  $\sigma_\epsilon = 1.0$  in our simulation studies. The dashed line in the top left panel is the true population slope function  $\beta(t) = 1 + 2\sin(2\pi t) + \cos(2\pi t)$ .

the estimated population slope function  $\hat{\beta}(t)$  and individual slope function  $\hat{\beta}_i(t)$  is close, which indicates that the estimation is stable. In this case, when the number of replicated measurements for each individual increases from  $m_i = 5$  to  $m_i = 10$ , the mean of RMISE of the estimated population slope function  $\hat{\beta}(t)$  decreases 27%, and the mean of RMISE of the estimated individual slope function  $\hat{\beta}_i(t)$  decreases 8%. When the functional covariate  $X(t)$  is observed with measurement errors with the standard deviation  $\sigma_e = 0.5$ , the mean of RMISE of the estimated population slope function  $\hat{\beta}(t)$  increases 36%, and the mean of RMISE of the estimated individual slope function  $\hat{\beta}_i(t)$  increases 25%, in comparison with the case when the functional covariate  $X(t)$  is observed directly without measurement errors.

#### 4. Applications

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

##### 4.1. Ozone Pollution Analysis

The first study is to re-visit the air pollution study introduced in Section 1. The functional linear mixed-effects model (2) is used to study the effect of the 24-hour nitrogen dioxide  $NO_2(t)$  on the daily maximum ozone concentration. We fit the following mixed-effect model

$$Y_{ij} = \alpha_0 + a_i + \int_0^{23} [\beta(t) + b_i(t)]X_{ij}(t)dt + \epsilon_{ij},$$

$$a_i \sim N(0, \sigma_a^2), \epsilon_{ij} \sim N(0, \sigma_\epsilon^2), i = 1, \dots, n, j = 1, \dots, m_i,$$

where  $Y_{ij}$  is the daily maximum ozone within 24 hours for the  $j$ -th day in the  $i$ -th city,  $X_{ij}(t)$  is the 24-hour nitrogen dioxide  $NO_2(t)$  observations. The data are collected for  $n = 62$  cities from April 13 to September 30, 1996.

For computational facilities, we have considered the penalized spline estimator and expand the functional coefficients in cubic B-splines basis functions with  $K = 26$  equispaced interior knots for the population slope function  $\beta(t)$  and the random slope function  $b_i(t)$ .

The fixed effects  $\{\alpha_0, \beta(t)\}$  and random effects  $\{a_i, b_i(t)\}$  are estimated by minimizing

$$\begin{aligned} & H(\boldsymbol{\theta}, \boldsymbol{\xi}) \\ = & \sum_{i=1}^n \sum_{j=1}^{m_i} \frac{1}{2\sigma_\epsilon^2} \left( Y_{ij} - \alpha_0 - a_i - \int_0^{23} [\beta(t) + b_i(t)] X_{ij}(t) dt \right)^2 + \frac{1}{2} \sum_{i=1}^n \mathbf{b}'_i \mathbf{D}^{-1} \mathbf{b}_i \\ & + \left[ \frac{\lambda_\beta}{2} \int_0^{23} \left\{ \frac{d^2\beta(t)}{dt^2} \right\}^2 dt + \frac{\lambda_b}{2} \sum_{i=1}^n \int_0^{23} \left\{ \frac{d^2b_i(t)}{dt^2} \right\}^2 dt \right] + \frac{1}{2\sigma_a^2} \sum_{i=1}^n a_i^2. \end{aligned}$$

The smoothing parameters are chosen as  $\lambda_\beta = 10^{2.0}$  and  $\lambda_b = 10^{0.5}$  by GCV criterion. We implement the REML-based EM algorithm proposed in Section 2.4. The estimate for the intercept  $\alpha_0$  is  $\hat{\alpha}_0 = 3.8262$  with the estimated standard error 0.0094, and the 95% confidence interval of  $\alpha_0$  is [3.8078, 3.8446].

Figure 3 displays the estimate population slope function  $\hat{\beta}(t)$  and its approximate 95% pointwise confidence interval. We can see a positive correlation between the maximum ozone concentration and the nitrogen dioxide before 11 am and after 8 pm but negative correlation between 11 am and 8 pm. Due to the stopping of the sun lighting in the night, a lot of nitrogen dioxide is accumulated; with the sunrise at about 6 am, more and more nitrogen dioxide is reacted with the sun light to generate ozone, so more ozone is generated with the decreasing of nitrogen dioxide. This process will last until the sunset at about 7-8 pm, then the nitrogen dioxide is accumulated again.

Figure 4 displays the estimated individual slope function  $\hat{\beta}_i(t) = \hat{\beta}(t) + \hat{b}_i(t)$  for four cities: Baton Rouge, Buffalo, Johnstown, and Tampa. The individual slope function  $\hat{\beta}_i(t)$  for Buffalo is lower than the population slope function in the whole day, which indicates that the hourly nitrogen dioxide has a lower effect on maximum ozone concentration in the whole day. On the other hand, the individual slope function  $\hat{\beta}_i(t)$  for Tampa is higher than the population slope function in the whole day. This interesting phenomenon cannot be found from the regular functional linear regression model.

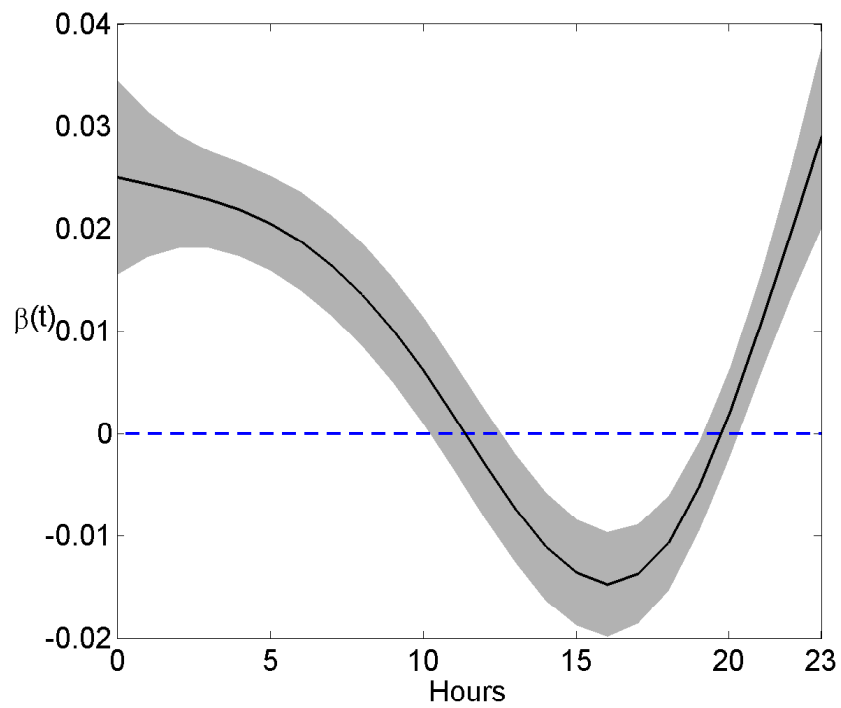


Figure 3: The estimated population slope function  $\hat{\beta}(t)$  for predicting the log maximum of ozone concentration from the hourly  $NO_2(t)$ . The shaded area indicates the pointwise 95% confidence interval for  $\hat{\beta}(t)$ .



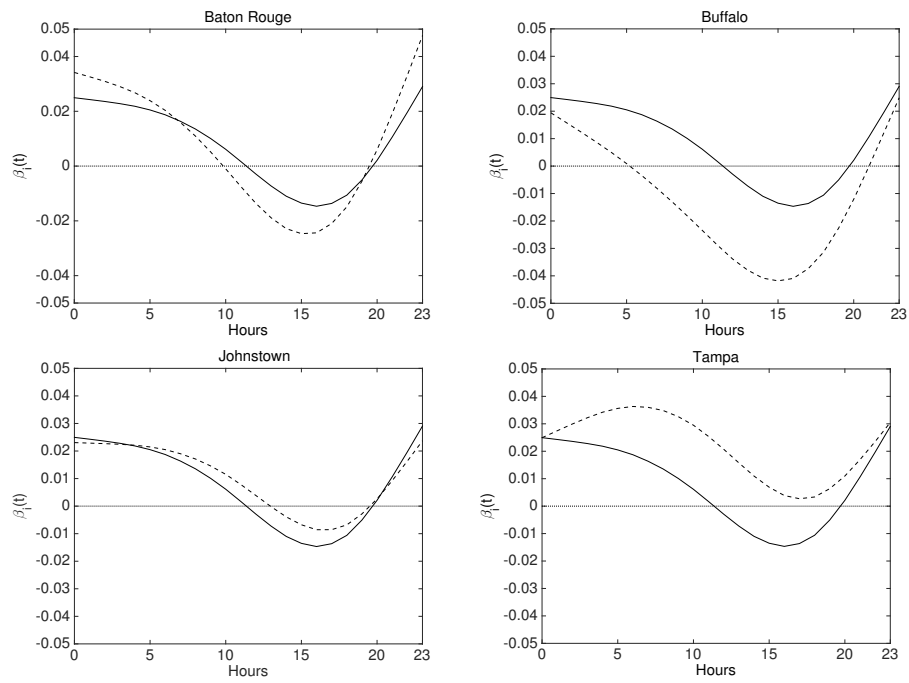


Figure 4: The estimated individual slope function  $\hat{\beta}_i(t) = \hat{\beta}(t) + \hat{b}_i(t)$  for predicting the log maximum ozone concentration from the hourly  $NO_2(t)$  for four cities: Baton Rouge, Buffalo, Johnstown, and Tampa. The solid line is the estimated population slope function  $\hat{\beta}(t)$ , and the dashed line is the individual slope function  $\hat{\beta}_i(t)$ .

#### 4.2. Weather Data Analysis

In this study, we are exploring the effect of the daily temperature in each year on the annual precipitation. We use the dataset consisting of the annual precipitation and the corresponding daily temperature measurements for 38 Canadian weather stations in 1961-1991. There are a lot of missing data in the year 1979, thus we delete the data in the year 1979. The functional linear mixed-effects model of our interest is

$$Y_{ij} = \alpha_0 + a_i + \int_0^{365} [\beta(t) + b_i(t)] X_{ij}(t) dt + \epsilon_{ij}, \quad a_i \sim N(0, \sigma_a^2), \quad \epsilon_{ij} \sim N(0, \sigma_\epsilon^2),$$

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

Due to the periodicity of weather data, we choose 35 Fourier basis functions to represent the population slope function  $\beta(t)$  and the individual slope function  $b_i(t)$ . We follow the suggestion of Ramsay and Silverman (2005) to use the harmonic acceleration operator to define the roughness penalty for the population slope function  $\beta(t)$ . The harmonic acceleration operator is defined as  $L\beta(t) = d^3\beta(t)/dt^3 + \omega^2 d\beta(t)/dt$ , where  $\omega = \frac{2\pi}{365}$  is the period of the non-parametric function. Therefore, the 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 slope function  $b_i(t)$ .

The fixed effects  $\{\alpha_0, \beta(t)\}$  and random effects  $\{a_i, b_i(t)\}$  are estimated by minimizing

$$\begin{aligned} & H(\boldsymbol{\theta}, \boldsymbol{\xi}) \\ = & \sum_{i=1}^n \sum_{j=1}^{m_i} \frac{1}{2\sigma_\epsilon^2} \left( Y_{ij} - \alpha_0 - a_i - \int_0^{365} [\beta(t) + b_i(t)] X_{ij}(t) dt \right)^2 + \frac{1}{2} \sum_{i=1}^n \mathbf{b}'_i \mathbf{D}^{-1} \mathbf{b}_i \\ & + \left[ \frac{\lambda_\beta}{2} \int_0^{365} \left\{ L\beta(t) \right\}^2 dt + \frac{\lambda_b}{2} \sum_{i=1}^n \int_0^{365} \left\{ Lb_i(t) \right\}^2 dt \right] + \frac{1}{2\sigma_a^2} \sum_{i=1}^n a_i^2. \end{aligned}$$

The smoothing parameters are chosen as  $\lambda_\beta = 10^{13.75}$  and  $\lambda_b = 10^{12.25}$  by GCV criterion. We implement the REML-based EM algorithm proposed in Section

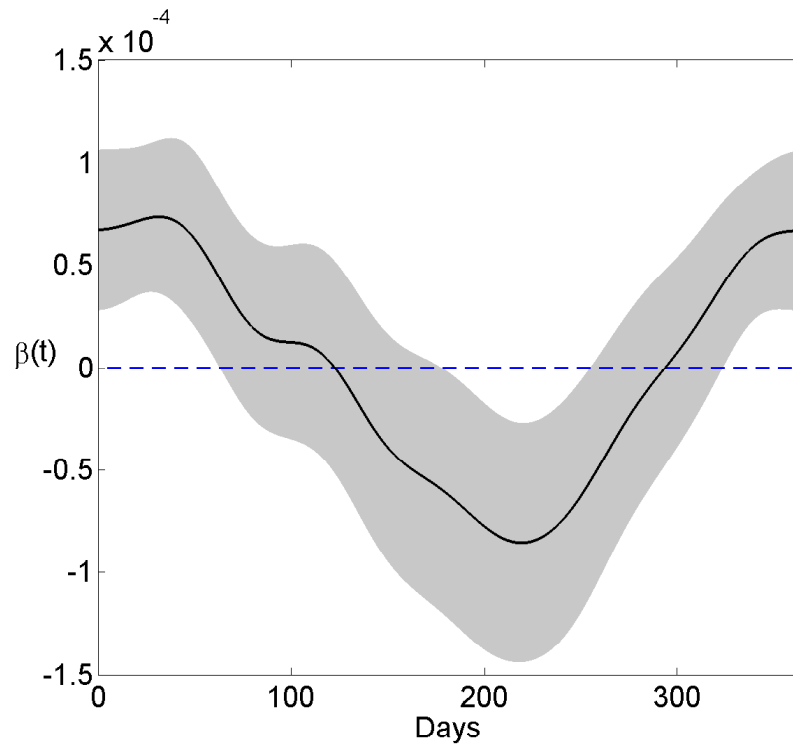


Figure 5: The estimated population slope function  $\hat{\beta}(t)$  for predicting the log total annual precipitation from the daily temperature. The shaded area indicates the pointwise 95% confidence interval for  $\hat{\beta}(t)$ .

2.2. The estimate for the intercept  $\alpha_0$  is  $\hat{\alpha}_0 = 2.994$  with the estimated standard error 0.055, and the 95% confidence interval of  $\alpha_0$  is  $[2.886, 3.101]$ .

Figure 5 displays the estimated population slope function  $\beta(t)$  and the 95% pointwise confidence interval. 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 Figure 6. There are some obvious individual variations from the population slope function for each station. For example, the Brandon city is located in

western Manitoba province, on the banks of the Assiniboine River. Figure 6 shows that the individual slope function of Brandon is higher than the population slope function in the whole year, because Brandon has a lower latitude and a large amount of precipitation in most of whole year. This phenomenon cannot be inferred from the regular functional linear regression model.

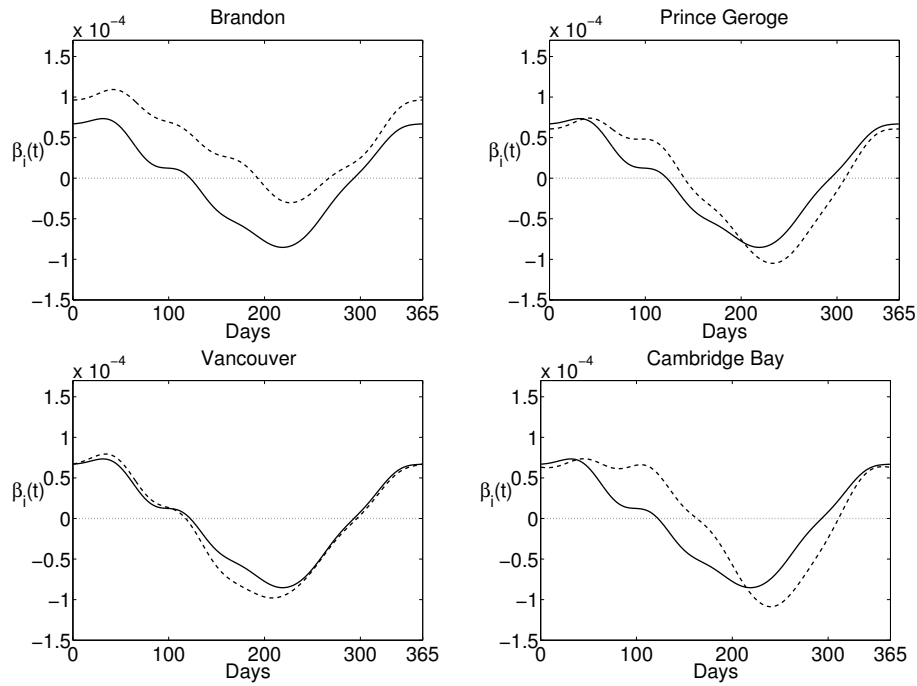


Figure 6: The estimated individual slope function  $\hat{\beta}_i(t) = \hat{\beta}(t) + \hat{b}_i(t)$  for predicting the log total annual precipitation from the daily temperature for four cities: Brandon, Prince George, Vancouver, Cambridge Bay. The solid line is the estimated population slope function  $\hat{\beta}(t)$ , and the dashed line is the individual slope function  $\hat{\beta}_i(t)$ .

## 5. Conclusions

The functional linear regression model (1) is a popular tool to analyze the relationship between a scalar response variable and a functional covariate. But when there are repeated measurements on multiple subjects, the same slope function seems to be a too strict assumption. In this article, we propose a

functional linear mixed-effect model (2). This model is more flexible than the regular functional linear regression model in the sense that each subject has their individual slope function while all subjects share a population slope function.

The population and random slope functions are estimated by the 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 show that the 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 the 24-hour nitrogen dioxide on the daily maximum ozone concentration. Some interesting results are found. For example, the hourly nitrogen dioxide has a consistently higher effect on the daily maximum ozone concentration in the whole day in some cities such as Tampa. These insights cannot be gained from the regular functional linear regression models. The similar phenomenon is also found in our second application to investigate the effect of the daily temperature on the annual precipitation.

## References

## References

- Ash, R.B., Gardner, M.F., 1975. Topics in Stochastic Processes. Probability and Mathematical Statistics 27. Academic Press, New York.
- Cai, T.T., Hall, P., 2006. Prediction in functional linear regression. *The Annals of Statistics* 34, 2159–2179.
- Cardot, H., Crambes, C., Sarda, P., 2007. Ozone pollution forecasting using conditional mean and conditional quantiles with functional covariates, in: Hardle, W., Mori, Y., Vieu, P. (Eds.), *Statistical Methods for Biostatistics and Related Fields*. Springer, Berlin, pp. 221–243.

- Chiou, J.M., Müller, H.G., Wang, J.L., 2003. Functional quasi-likelihood regression models with smooth random effects. *J. R. Stat. Soc. Ser. B.* 65, 405–423.
- Crambes, C., Kneip, A., Sarda, P., 2009. Smoothing splines estimators for functional linear regression. *The Annals of Statistics* 37, 35–72.
- Ferraty, F., Vieu, P., 2006. *Nonparametric Functional Data Analysis: Methods, Theory, Applications and Implementations*. Springer-Verlag, London.
- Goldsmith, J., Crainiceanu, C.M., Caffo, B., Reich, D., 2012. Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements. *Journal of the Royal Statistical Society, Series C* 61, 453–469.
- Goldsmith, J., Wand, M.P., Crainiceanu, C., 2011. Functional regression via variational bayes. *Electronic Journal of Statistics* 5, 572–602.
- Morris, J.S., 2015. Functional regression. *Statistics and Its Application* 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. *Journal of the Royal Statistical Society, Series B* 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. *Journal of the Royal Statistical Society, Series B* 53, 233–243.
- Staniswalis, J.G., Lee, J.J., 1998. Nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association* 93, 1403–1418.

- Wu, H.L., Zhang, J.T., 2006. Nonparametric Regression Methods for Longitudinal Data Analysis: Mixed-Effects Modeling Approaches. Wiley, New York.
- Wu, Y.C., Fan, J.Q., Müller, H.G., 2010. Varying-coefficient functional linear regression. *Bernoulli* 16, 730–758.
- Yao, F., Müller, H., Wang, J., 2005. Functional linear regression analysis for longitudinal data. *The Annals of Statistics* 33, 2873–2903.
- Yuan, M., Cai, T.T., 2010. A reproducing kernel hilbert space approach to functional linear regression. *The Annals of Statistics* 38, 3412–3444.

Table 1: The Bias, Standard deviation (STD) and RMSE of the intercept, and the means of RMISE of slope functions obtained by applying the REML-based EM algorithm on the simulated data with 1,000 simulation replicates. The true population slope function is  $\beta(t) = 1 + 2t^2 + \exp(-3t)$ .

$\sigma_e$	$\sigma_\epsilon$	$n$	$m_i$	Intercept			RMISE $\{\beta(t)\}$	RMISE $\{\beta_i(t)\}$		
				Bias	STD	RMSE				
0.0	0.5	50	5	-0.0017	0.0986	0.0986	0.0047	0.0210		
			10	0.0019	0.0926	0.0926	0.0032	0.0187		
			20	0.0005	0.0862	0.0862	0.0025	0.0166		
		100	5	-0.0034	0.0638	0.0639	0.0027	0.0180		
			10	-0.0012	0.0611	0.0611	0.0019	0.0161		
			20	-0.0021	0.0609	0.0610	0.0014	0.0145		
		0.0	1.0	50	5	-0.0010	0.1079	0.1079	0.0080	0.0274
					10	0.0000	0.1023	0.1023	0.0060	0.0237
					20	-0.0018	0.0923	0.0923	0.0043	0.0204
100	5			-0.0006	0.0760	0.0760	0.0054	0.0230		
	10			-0.0016	0.0663	0.0663	0.0036	0.0198		
	20			-0.0015	0.0630	0.0630	0.0024	0.0174		
0.5	0.5			50	5	-0.0024	0.1001	0.1002	0.0049	0.0212
					10	0.0003	0.0952	0.0952	0.0035	0.0189
					20	0.0005	0.0890	0.0890	0.0033	0.0177
		100	5	-0.0023	0.0636	0.0637	0.0029	0.0182		
			10	0.0011	0.0620	0.0620	0.0020	0.0162		
			20	-0.0016	0.0623	0.0624	0.0022	0.0155		
		0.5	1.0	50	5	-0.0036	0.1130	0.1131	0.0082	0.0277
					10	0.0026	0.1039	0.1039	0.0057	0.0235
					20	-0.0001	0.0935	0.0935	0.0043	0.0204
100	5			-0.0023	0.0764	0.0764	0.0053	0.0228		
	10			0.0009	0.0698	0.0698	0.0037	0.0199		
	20			-0.0011	0.0633	0.0633	0.0026	0.0176		



Table 2: The Bias, Standard deviation (STD) and RMSE of the intercept, and the means of RMISE of slope functions obtained by applying the REML-based EM algorithm on the simulated data with 1,000 simulation replicates. The true population slope function is  $\beta(t) = 1 + 2 \sin(2\pi t) + \cos(2\pi t)$ .

$\sigma_e$	$\sigma_\epsilon$	$n$	$m_i$	Intercept			RMISE $\{\beta(t)\}$	RMISE $\{\beta_i(t)\}$		
				Bias	STD	RMSE				
0.0	0.5	50	5	0.0039	0.0980	0.0981	0.0049	0.0415		
			10	0.0032	0.0912	0.0912	0.0038	0.0409		
			20	-0.0012	0.0890	0.0890	0.0033	0.0402		
		100	5	-0.0014	0.0643	0.0644	0.0026	0.0405		
			10	-0.0014	0.0630	0.0631	0.0023	0.0396		
			20	-0.0008	0.0629	0.0629	0.0019	0.0395		
		0.0	1.0	50	5	0.0059	0.1106	0.1107	0.0082	0.0447
					10	0.0050	0.1005	0.1006	0.0053	0.0423
					20	0.0006	0.0948	0.0948	0.0038	0.0407
100	5			-0.0014	0.0747	0.0747	0.0039	0.0418		
	10			-0.0006	0.0687	0.0687	0.0028	0.0401		
	20			-0.0003	0.0653	0.0653	0.0022	0.0397		
0.5	0.5			50	5	0.0021	0.0981	0.0981	0.0052	0.0418
					10	0.0042	0.0929	0.0930	0.0039	0.0409
					20	-0.0012	0.0895	0.0895	0.0033	0.0402
		100	5	-0.0030	0.0643	0.0644	0.0027	0.0406		
			10	-0.0016	0.0620	0.0620	0.0023	0.0396		
			20	-0.0013	0.0625	0.0625	0.0019	0.0395		
		0.5	1.0	50	5	0.0026	0.1142	0.1142	0.0082	0.0447
					10	0.0026	0.1008	0.1009	0.0052	0.0422
					20	-0.0020	0.0927	0.0927	0.0039	0.0408
100	5			-0.0034	0.0754	0.0755	0.0040	0.0419		
	10			-0.0015	0.0678	0.0678	0.0029	0.0402		
	20			0.0004	0.0652	0.0652	0.0022	0.0397		