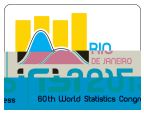

Generalized Gaussian Process Regression Model for Non-Gaussian Functional Data

Bo Wang
Department of Mathematics, University of Leicester, UK, bw77@leicester.ac.uk

Jian Qing Shi*
School of Mathematics & Statistics, Newcastle University, UK, j.q.shi@ncl.ac.uk

Abstract

In this paper we discuss a concurrent generalized Gaussian process regression model for functional data where the functional response variable has a binomial, Poisson or other non-Gaussian distribution from an exponential family while the covariates are mixed functional and scalar variables. The proposed model offers a nonparametric concurrent generalized functional regression method for functional data with multi-dimensional covariates, and provides a natural framework on modeling common mean structure and covariance structure simultaneously for repeatedly observed functional data. The mean structure provides an overall information about the observations, while the covariance structure can be used to catch up the characteristic of each individual batch. The prior specification of covariance kernel enables us to accommodate a wide class of nonlinear models. The definition of the model, the inference and the implementation as well as its asymptotic properties are discussed. Several numerical examples with different non-Gaussian response variables are presented.

Keywords: Covariance kernel, Exponential family, Functional regression analysis, Nonparametric regression.

1. Introduction

A concurrent functional regression model with functional response variable can be defined by

$$y_m(t) = f(\boldsymbol{x}_m(t), \boldsymbol{u}_m) + \epsilon_m(t), \tag{1}$$

where $y_m(t)$ $(m = 1, \ldots, M)$ stands for M batches (or curves) of functional data, $f(\cdot)$ is an unknown nonlinear function, depending on a set of functional covariates $\boldsymbol{x}_m(t)$ and a set of scalar covariates $\boldsymbol{u}_m$, and $\epsilon_m(t)$ is the random error. Shi et al. (2007) proposed a Gaussian process functional regression (GPFR) model, which is defined by

$$f(\boldsymbol{x}_m(t), \boldsymbol{u}_m) = \mu_m(t) + \tau_m(\boldsymbol{x}_m(t)), \tag{2}$$

where $\mu_m(t)$ is the mean structure of the functional data and $\tau_m(\boldsymbol{x}_m(t))$ represents a Gaussian process regression (GPR) model having zero mean and covariance kernel $k(\cdot, \cdot; \boldsymbol{\theta})$ (for the detailed definition of Gaussian process regression models, see Rasmussen and Williams, 2006; Shi and Choi, 2011). This nonparametric functional regression model can address the regression problem with multi-dimensional functional covariates and model the mean structure and covariance structure simultaneously; see the detailed discussion in Shi et al. (2007).

The aim of this paper is to extend the GPFR model (1) and (2) to situations where the response variable, denoted by $z(t)$, is known to be non-Gaussian. The work is motivated by the following example, concerning data collected during standing-up manoeuvres of paraplegic patients. The outputs are the human body's standing-up phases during rising from sitting position to standing position. Specifically, $z(t)$ takes value of either 0, 1 or 2, corresponding to the phases of 'sitting', 'seat unloading and ascending' or 'stablising' respectively, required for feeding back to a simulator control system. Since it is usually difficult to measure the body position in practice, the aim of the example is to develop a model for reconstructing the position of the human body by using some easily measured quantities such as motion kinematic, reaction forces and torques, which are functional covariates denoted by $\boldsymbol{x}(t)$. This is to investigate the regression relationship between the non-Gaussian functional response variable $z(t)$ and a set of functional covariates $\boldsymbol{x}(t)$. Since the standing-up phases are irreversible, $z(t)$ is an ordinal response variable, taking value from three ordered categories. If we assume that there exists an unobservable latent process $\eta(t)$ associated with $\boldsymbol{x}(t)$ and the

response variable $z(t)$ depends on this latent process, then by using a probit link function, we can define a model as follows:

$$z(t) = j \quad \text{if } b_j < \eta(t) \le b_{j+1}, \quad j \in \{0, 1, 2\},$$

where $b_0 = -\infty$, $b_3 = \infty$, and $b_1, b_2 \in \mathcal{R}$ are the thresholds. Now the problem becomes how to model $\eta(\cdot)$ by the functional covariates $\boldsymbol{x}(t)$, or how to find a function f such that $\eta(t) = f(\boldsymbol{x}(t))$.

Generally, letting $h^{-1}(\cdot)$ be a given link function, a generalized linear regression model is defined as $E(z_m(t)) = h(\boldsymbol{x}_m^T(t)\boldsymbol{\beta})$. Breslow and Clayton (1993) proposed a generalized linear mixed model to deal with heterogeneity: $E(z_m(t)|\boldsymbol{\gamma}_2) = h(\boldsymbol{x}_{m1}^T(t)\boldsymbol{\gamma}_1 + \boldsymbol{x}_{m2}^T(t)\boldsymbol{\gamma}_2)$, where $\boldsymbol{\gamma}_1$ is the coefficient for the fixed effect and $\boldsymbol{\gamma}_2$ is a random vector representing random effect. However, if we have little practical knowledge on the relationship between the response variable and the covariates (such as the case in the above Paraplegia example), it is more sensible to use a nonparametric model. In this paper, we propose to use a Gaussian process regression model to define such a nonparametric model, namely a generalized Gaussian process functional regression (GGPFR) model. Similar to GPFR model (Shi et al., 2007), the advantages of this model include: (1) it offers a nonparametric generalized functional regression model for functional data with functional response and multi-dimensional functional covariates; (2) it provides a natural framework on modeling mean structure and covariance structure simultaneously and the latter can be used to model the individual characteristic for each batch; and (3) the prior specification of covariance kernel enables us to accommodate a wide class of nonlinear functions.

2. Generalized Gaussian process functional regression model

2.1 The Model

Let $\{z_m(t), t \in \mathcal{T}\}$ be a functional or longitudinal response variable for the m-th subject, namely the m-th batch. We assume that $z_m(t)$'s are independent for different batches $m = 1, \ldots, M$, but within the batch, $z_m(t_i)$ and $z_m(t_j)$ are dependent at different points. We suppose that $z_m(t)$ has a distribution from an exponential family with the following density function

$$p(z_m(t)|\alpha_m(t), \phi_m(t)) = \exp\left\{\frac{z_m(t)\alpha_m(t) - b(\alpha_m(t))}{a(\phi_m(t))} + c(z_m(t), \phi_m(t))\right\} \tag{3}$$

where $\alpha_m(t)$ and $\phi_m(t)$ are canonical parameter and dispersion parameter respectively, both functional. We have $\mathrm{E}(z_m(t)) = b'(\alpha_m(t))$ and $\mathrm{Var}(z_m(t)) = b''(\alpha_m(t))a(\phi_m(t))$, where $b'(\alpha)$ and $b''(\alpha)$ are the first two derivatives of $b(\alpha)$ with respect to α.

Suppose that $\boldsymbol{x}_m(t)$ is a Q-dimensional vector of functional covariates. Nonparametric generalized Gaussian process functional regression (GGPFR) models are defined by (3) and the following

$$\begin{aligned} \mathrm{E}(z_m(t)|\tau_m(t)) &= h(\mu_m(t) + \tau_m(t)), \\ \tau_m(t) = \tau_m(\boldsymbol{x}_m(t)) &\sim GPR(0, k(\cdot,\cdot;\boldsymbol{\theta})|\boldsymbol{x}_m(t)). \end{aligned} \tag{4}$$

Here, the unobserved latent variable $\tau_m(t)$ is modeled by a nonparametric GPR model via a Gaussian process prior, depending on the functional covariates $\boldsymbol{x}_m(t)$. The GPR model is specified by a covariance kernel $k(\cdot,\cdot;\boldsymbol{\theta})$ and by the Karhunen-Loève expansion

$$\tau(\boldsymbol{x}) = \sum_{j=1}^{\infty} r_j \psi_j(\boldsymbol{x}),$$

where $r_j \sim N(0, \lambda_j)$, $\lambda_1 \ge \lambda_2 \ge \cdots \ge 0$ are the eigenvalues and $\psi_1(\boldsymbol{x}), \psi_2(\boldsymbol{x}), \ldots$ are the associated eigenfunctions of the covariance kernel. One example of $k(\cdot,\cdot;\boldsymbol{\theta})$ is the following squared exponential covariance function with a nonstationary linear term:

$$\begin{aligned} &\mathrm{Cov}(\tau(t_i), \tau(t_j)) = k(\boldsymbol{x}(t_i), \boldsymbol{x}(t_j); \boldsymbol{\theta}) \\ =&v_1 \exp\Big(-\frac{1}{2}\sum_{q=1}^{Q} w_q(x_q(t_i) - x_q(t_j))^2\Big) + a_1 \sum_{q=1}^{Q} x_q(t_i)x_q(t_j), \end{aligned} \tag{5}$$

where $\boldsymbol{\theta} = (w_1, \ldots, w_Q, v_1, a_1)$ is a set of hyper-parameters involved in the Gaussian process prior. We can use generalized cross-validation (GCV) or empirical Bayesian method to choose the value of $\boldsymbol{\theta}$. When Q is large, GCV approach is usually inefficient. We will use the empirical Bayesian method in this paper; the details are given in the next subsection. Some other covariance kernels such as powered exponential and Matérn covariance functions can also be used; see more discussion on the choice of covariance function in Rasmussen and Williams (2006) and Shi and Choi (2011).

In the model given by (4) the response variable $z_m(t)$ depends on $\boldsymbol{x}_m(t)$ at the current time only. This type of models are often called the concurrent regression models in the literature. Therefore (4) can be regarded as a generalization of the concurrent functional linear model discussed in Ramsay and Silverman (2005). In this model the common mean structure across M batches is given by $\mu_m(t)$. If we use a linear mean function which depends on a set of p scalar covariates $\boldsymbol{u}_m$ only, (4) can be expressed as

$$\mathrm{E}(z_m(t)|\tau_m(t)) = h(\mu_m(t) + \tau_m(t)) = h(\boldsymbol{u}_m^T\boldsymbol{\beta}(t) + \tau_m(t)). \tag{6}$$

In this case the regression relationship between the functional response $z_m(t)$ and the functional covariates $\boldsymbol{x}_m(t)$ is modeled by the covariance structure $\tau_m(\boldsymbol{x}_m)$. Other mean structures, including concurrent form of functional covariates, can also be used.

The proposed model has some features worth noting. In addition to those discussed in Section 1, we highlight that the GGPFR model is actually very flexible. It can model the regression relationship between the non-Gaussian functional response and the multi-dimensional functional covariates nonparametrically. Moreover, if we had known some prior information between $z_m(t)$ (or $\mathrm{E}(z_m(t))$) and some of the functional covariates, we could easily integrate it by adding a parametric mean part. For example we may define

$$\mu_m(t) = \boldsymbol{u}_m^T\boldsymbol{\beta}(t) + \boldsymbol{x}_{m1}^T(t)\boldsymbol{\gamma}_1 + \boldsymbol{x}_{m2}^T(t)\boldsymbol{\gamma}_2,$$

i.e. including a term in the GGPFR similar to the generalized linear mixed model (Breslow and Clayton, 1993); an example of this type model is provided in Appendix G. The nonparametric part can still be modeled by $\tau_m(t)$ via a GPR model. Other nonparametric covariance structure can also be considered; some examples can be found in Rice and Silverman (1991), Hall et al. (2008) and Leng et al. (2009). However, most of these methods are limited to small (usually one) dimensional $\boldsymbol{x}(t)$ or the covariance matrix with a special structure.

2.2. Empirical Bayesian Learning

Now suppose that we have M batches of data from M subjects or experimental units. In the m-th batch, N_m observations are collected at $\boldsymbol{T}_m = \{t_{m1}, \ldots, t_{mN_m}\}$. We denote $z_m(t_{mi})$, $\tau_m(t_{mi})$ and $\boldsymbol{x}_m(t_{mi})$ by z_{mi}, τ_{mi} and $\boldsymbol{x}_{mi}$, respectively, for $i = 1, \ldots, N_m$ and $m = 1, \ldots, M$. Collectively, we denote $\boldsymbol{Z}_m = (z_{m1}, \ldots, z_{mN_m})^T$ and $\boldsymbol{Z} = \{\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M\}$, and denote $\boldsymbol{\tau}_m$, $\boldsymbol{\tau}$, $\boldsymbol{X}_m$ and $\boldsymbol{X}$ in the same way. They are the realizations of $z_m(t)$, $\tau_m(t)$ and $\boldsymbol{x}_m(t)$ at $\boldsymbol{T}_m$. A discrete GGPFR model is therefore given by

$$\begin{aligned} z_{mi}|\alpha_{mi}, \phi, \tau_{mi} &\sim EF(\alpha_{mi}, \phi), \quad i = 1, \ldots, N_m, && (7)\\ \mathrm{E}(z_{mi}|\tau_{mi}) &= b'(\alpha_{mi}) = h(\boldsymbol{u}_m^T\boldsymbol{\beta}(t_i) + \tau_{mi}), && (8)\\ \boldsymbol{\tau}_m &= (\tau_{m1}, \ldots, \tau_{mN_m})^T \sim N(0, \boldsymbol{C}_m) && (9) \end{aligned}$$

for $m = 1, \ldots, M$, where $EF(\cdot, \cdot)$ is a distribution from the exponential family (3) and $\alpha_{mi} = \alpha_m(t_i)$. $\boldsymbol{\tau}_m$ has an N_m-variate normal distribution with zero mean and covariance matrix $\boldsymbol{C}_m = (C_m^{ij})$ for $i, j = 1, \ldots, N_m$. Here we assume a fixed dispersion parameter ϕ, but the method developed in this paper can be applied to more general cases.

We consider the estimation of $\boldsymbol{\beta}(t)$ first. To estimate the functional coefficient $\boldsymbol{\beta}(t)$, we expand it by a set of basis functions (see e.g. Ramsay and Silverman, 2005). In this paper, we use B-spline approximation. Let $(t) = (\Phi_1(t), \ldots, \Phi_D(t))^T$ be the B-spline basis functions, then the functional coefficient $\boldsymbol{\beta}(t)$ can be represented as $\boldsymbol{\beta}(t) = \boldsymbol{B}^T (t)$, where the j-th column of $\boldsymbol{B}$, $B_j = (B_1^j, \ldots, B_D^j)^T$, is the B-spline coefficients for $\beta_j(t)$. Thus, at the observation point $\boldsymbol{T}_m$, we have $\boldsymbol{\mu}_m = {}_m\boldsymbol{B}\boldsymbol{u}_m$, where ${}_m$ is an $N_m \times D$ matrix with the (i, d)-th element $\Phi_d(t_{mi})$.

The covariance matrix $\boldsymbol{C}_m = (C_m^{ij})$ of $\boldsymbol{\tau}_m$ depends on $\boldsymbol{X}_m$ and the unknown hyper-parameter $\boldsymbol{\theta}$. If we use covariance kernel (5), its element C_m^{ij} is given by

$$C_m^{ij} = v_1 \exp\Big(-\frac{1}{2}\sum_{q=1}^{Q} w_q(x_{miq} - x_{mjq})^2\Big) + a_1\sum_{q=1}^{Q} x_{miq}x_{mjq}. \tag{10}$$

The covariance matrix involves the hyper-parameter $\boldsymbol{\theta} = \{w_1, \ldots, w_Q, v_1, a_1\}$, whose value is given based on the prior knowledge in conventional Bayesian analysis. As discussed in Shi and Choi (2011), empirical Bayesian learning method is preferable for GPR models when the dimension of $\boldsymbol{\theta}$ is large.

The idea of empirical Bayesian learning is to choose the value of the hyper-parameter $\boldsymbol{\theta}$ by maximizing the marginal density function. Thus, $\boldsymbol{\theta}$ as well as the unknown parameter $\boldsymbol{B}$ can be estimated at the same time by maximizing the following marginal log-likelihood

$$\begin{aligned} l(\boldsymbol{B}, \boldsymbol{\theta}) &= \sum_{m=1}^{M} \log\{p(\boldsymbol{Z}_m | \boldsymbol{B}, \boldsymbol{\theta}, \boldsymbol{X}_m)\} \\ &= \sum_{m=1}^{M} \log \int \Big\{ \prod_{i=1}^{N_m} p(z_{mi} | \tau_{mi}, \boldsymbol{B}) \Big\} (2\pi)^{-\frac{N_m}{2}} |\boldsymbol{C}_m|^{-\frac{1}{2}} \exp\Big\{ -\frac{1}{2} \boldsymbol{\tau}_m^T \boldsymbol{C}_m^{-1} \boldsymbol{\tau}_m \Big\} d\boldsymbol{\tau}_m, \end{aligned} \tag{11}$$

where $p(z_{mi}|\tau_{mi}, \boldsymbol{B})$ is derived from the exponential family as defined in (7). Obviously the integral involved in the above marginal density is analytically intractable unless $p(z_{mi}|\tau_{mi}, \boldsymbol{B})$ has a special form such as the density function of normal distribution. One method to address this problem is to use Laplace approximation. However, as pointed out in Section 4.1 in Rue et al. (2009), the error rate of the Laplace approximation may be $O(1)$ since the dimension of $\boldsymbol{\tau}_m$ increases with the sample size N_m. A better method is to approximate $p(\boldsymbol{Z}_m|\boldsymbol{B}, \boldsymbol{\theta})$ in (11) (here and in the rest of the section the conditioning on $\boldsymbol{X}_m$ is omitted for simplicity) by

$$\tilde{p}(\boldsymbol{Z}_m | \quad), \left. \frac{p(\boldsymbol{\tau}_m, \boldsymbol{Z}_m | \quad)}{\tilde{p}_G(\boldsymbol{\tau}_m | \boldsymbol{Z}_m, \quad)} \right|_{\boldsymbol{\tau}_m = \boldsymbol{\tau}_m(\Theta)}, \tag{12}$$

where $\quad = (\boldsymbol{B}, \boldsymbol{\theta})$, $\tilde{p}_G(\boldsymbol{\tau}_m|\boldsymbol{Z}_m, \quad)$ is the Gaussian approximation to the full conditional density $p(\boldsymbol{\tau}_m|\boldsymbol{Z}_m, \quad)$, and $\boldsymbol{\tau}_m$ is the mode of the full conditional density of $\boldsymbol{\tau}_m$ for a given $\quad$. Here,

$$\begin{aligned} p(\boldsymbol{\tau}_m, \boldsymbol{Z}_m | \quad) &= p(\boldsymbol{Z}_m | \boldsymbol{\tau}_m, \quad) p(\boldsymbol{\tau}_m | \quad) \\ &= \exp\Big\{ \log p(\boldsymbol{\tau}_m | \boldsymbol{\theta}) + \sum_{i=1}^{N_m} \log p(z_{mi} | \tau_{mi}, \boldsymbol{B}) \Big\}. \end{aligned}$$

We approximate $g_{mi}(\tau_{mi}) \text{ , } \log p(z_{mi}|\tau_{mi}, \boldsymbol{B})$ by Taylor expansion to the second order

$$g_{mi}(\tau_{mi}) \approx g_{mi}(\tau_{mi}^{(0)}) + a_{mi}\tau_{mi} - \frac{1}{2} d_{mi} \tau_{mi}^2,$$

where a_{mi} and d_{mi} depend on the first two derivatives of $g_{mi}(\tau_{mi})$ respectively and evaluated at $\tau_{mi}^{(0)}$. Thus,

$$p(\boldsymbol{\tau}_m, \boldsymbol{Z}_m | \quad) \propto \exp\Big\{ -\frac{1}{2} \boldsymbol{\tau}_m^T \boldsymbol{C}_m^{-1} \boldsymbol{\tau}_m - \frac{1}{2} \boldsymbol{\tau}_m^T \boldsymbol{D}_m \boldsymbol{\tau}_m + \boldsymbol{a}_m^T \boldsymbol{\tau}_m \Big\},$$

where $\boldsymbol{D}_m = \mathrm{diag}(d_{m1}, \ldots, d_{mN_m})$ and $\boldsymbol{a}_m^T = (a_{m1}, \ldots, a_{mN_m})$. We can then use the following Fisher scoring algorithm (Fahrmeir and Lang, 2001) to find the Gaussian approximation. Starting with $\tau_{mi}^{(0)}$, the k-th iteration is given by

(i) Find the solution $\boldsymbol{\tau}_m^{(k)}$ from $(\boldsymbol{C}_m^{-1} + \boldsymbol{D}_m)\boldsymbol{\tau}_m^{(k)} = \boldsymbol{a}_m$,

(ii) Update $\boldsymbol{a}_m$ and $\boldsymbol{D}_m$ using $\boldsymbol{\tau}_m^{(k)}$ and repeat (i).

After the process converges, say at $\boldsymbol{\tau}_m$, we get the Gaussian approximation $\tilde{p}_G(\boldsymbol{\tau}_m|\boldsymbol{Z}_m, \quad)$ which is the density function of the normal distribution $N\left(\boldsymbol{\tau}_m, (\boldsymbol{C}_m^{-1} + \boldsymbol{D}_m)^{-1}\right)$. We can then calculate $\hat{\quad} = (\hat{\boldsymbol{B}}, \hat{\boldsymbol{\theta}})$ by maximizing (11) using the approximation (12).

2.3 Prediction

Now we consider two types of prediction problems. First suppose that we have already observed some data for a subject, say N observations in the k-th batch, and want to obtain prediction at other points. This can be for one of the batches $1, \ldots, M$ or a completely new one. The observations are denoted by $\boldsymbol{Z}_k = \{z_{ki}, i =$

$1, \ldots, N\}$ which are collected at $\{t_{k1}, \ldots, t_{kN}\}$. The corresponding input vectors are $\boldsymbol{X}_k = \{\boldsymbol{x}_{k1}, \ldots, \boldsymbol{x}_{kN}\}$, and we also know the subject-based covariate $\boldsymbol{u}_k$. It is of interest to predict z^* at a new point t^* for the k-th subject given the test input $\boldsymbol{x}^* = \boldsymbol{x}_k(t^*)$. Secondly we will assume there are no data observed from the subject of interest except the subject-based covariate and want to predict z^* at a new point t^* with the input $\boldsymbol{x}^*$. We use $\mathcal{D}$ to denote all the training data and assume that the model itself has been trained (i.e. all the unknown parameters have been estimated) by the method discussed in the previous section. The main purpose in this section is to calculate $\mathrm{E}(z^*|\mathcal{D})$ and $\mathrm{Var}(z^*|\mathcal{D})$, which are used as the prediction and the predictive variance of z^*.

We now consider the first type of prediction. Let $\tau^* = \tau_k(t^*)$ be the underlying latent variable at t^*, then τ^* (for convenience we ignore the subscript) and $\boldsymbol{\tau}_k = (\tau_{k1}, \ldots, \tau_{kN})^T$ satisfy (9), and the expectation of z^* conditional on τ^* is given by (8):

$$\mathrm{E}(z^*|\tau^*, \mathcal{D}) = h(\boldsymbol{u}_k^T \hat{\boldsymbol{B}}^T \quad (t^*) + \tau^*). \tag{13}$$

It follows that

$$\mathrm{E}(z^*|\mathcal{D}) = \mathrm{E}\big[\mathrm{E}(z^*|\tau^*, \mathcal{D})\big] = \int h(\boldsymbol{u}_k^T \hat{\boldsymbol{B}}^T \quad (t^*) + \tau^*) p(\tau^*|\mathcal{D}) d\tau^*. \tag{14}$$

A simple method to calculate the above expectation is to approximate $p(\tau^*|\mathcal{D})$ using a Gaussian approximation $\tilde{p}_G(\boldsymbol{\tau}_k|\mathcal{D})$ as discussed around equation (12), that is,

$$p(\tau^*|\mathcal{D}) = \int p(\tau^*|\boldsymbol{\tau}_k, \mathcal{D}) p(\boldsymbol{\tau}_k|\mathcal{D}) d\boldsymbol{\tau}_k \approx \int p(\tau^*|\boldsymbol{\tau}_k) \tilde{p}_G(\boldsymbol{\tau}_k|\mathcal{D}) d\boldsymbol{\tau}_k. \tag{15}$$

Since it is assumed that both $\boldsymbol{\tau}_k$ and τ^* come from the same Gaussian process with covariance kernel $k(\cdot, \cdot; \boldsymbol{\theta})$, we have

$$(\boldsymbol{\tau}_k^T, \tau^*)^T \sim N\left(\boldsymbol{0}, \boldsymbol{C}_{N+1;N+1}\right), \quad \boldsymbol{C}_{N+1;N+1} = \begin{bmatrix} \boldsymbol{C}_{N;N} & \boldsymbol{C}_N^* \\ \boldsymbol{C}_N^{*T} & k(\boldsymbol{x}^*, \boldsymbol{x}^*; \hat{\boldsymbol{\theta}}) \end{bmatrix}$$

where $\boldsymbol{C}_{N;N}$ is the covariance matrix of $\boldsymbol{\tau}_k$, and $\boldsymbol{C}_N^*$ is a vector of the covariances between $\boldsymbol{\tau}_k$ and τ^*. Thus, $p(\tau^*|\boldsymbol{\tau}_k) = N(\boldsymbol{a}^T \boldsymbol{\tau}_k, \sigma^{*2})$, where $\boldsymbol{a}^T = \boldsymbol{C}_N^{*T} \boldsymbol{C}_{N;N}^{-1}$ and $\sigma^{*2} = k(\boldsymbol{x}^*, \boldsymbol{x}^*; \hat{\boldsymbol{\theta}}) - \boldsymbol{C}_N^{*T} \boldsymbol{C}_{N;N}^{-1} \boldsymbol{C}_N^*$. From the discussion given in the last paragraph in Section 2.2, we have

$$\tilde{p}_G(\boldsymbol{\tau}_k|\mathcal{D}) = N(\boldsymbol{\tau}_k, \quad), \qquad , \ (\boldsymbol{C}_{N;N}^{-1} + \boldsymbol{D}_k)^{-1}.$$

The integrand in (15) is therefore the product of two normal density functions. It is not difficult to prove (see the details in Appendix B of the supplementary materials) that $p(\tau^*|\mathcal{D})$ is still a normal density function

$$p(\tau^*|\mathcal{D}) = N(\boldsymbol{a}^T \boldsymbol{\tau}_k, \boldsymbol{a}^T \quad \boldsymbol{a} + \sigma^{*2}). \tag{16}$$

Then (14) can be evaluated by numerical integration.

To calculate $\mathrm{Var}(z^*|\mathcal{D})$, we use the formula:

$$\mathrm{Var}(z^*|\mathcal{D}) = \mathrm{E}[\mathrm{Var}(z^*|\tau^*, \mathcal{D})] + \mathrm{Var}[\mathrm{E}(z^*|\tau^*, \mathcal{D})]. \tag{17}$$

From the model definition, we have

$$\begin{aligned} \mathrm{Var}[\mathrm{E}(z^*|\tau^*, \mathcal{D})] &= \mathrm{E}\Big[\mathrm{E}(z^*|\tau^*, \mathcal{D})\Big]^2 - \Big(\mathrm{E}[\mathrm{E}(z^*|\tau^*, \mathcal{D})]\Big)^2 \\ &= \int \big[h(\boldsymbol{u}_k^T \hat{B}^T \quad (t^*) + \tau^*)\big]^2 p(\tau^*|\mathcal{D}) d\tau^* - \big[\mathrm{E}(z^*|\mathcal{D})\big]^2, \end{aligned} \tag{18}$$

and

$$\mathrm{E}[\mathrm{Var}(z^*|\tau^*, \mathcal{D})] = \int \mathrm{Var}(z^*|\tau^*, \mathcal{D}) p(\tau^*|\mathcal{D}) d\tau^* = \int b''(\hat{\alpha}^*) a(\phi) p(\tau^*|\mathcal{D}) d\tau^*, \tag{19}$$

where $\hat{\alpha}^*$ is a function of $h(\boldsymbol{u}_k^T \hat{B}^T \quad (t^*) + \tau^*)$, and $p(\tau^*|\mathcal{D})$ is given by (16). Thus (18) and (19) can also be evaluated by numerical integration.

The posterior density $p(\tau^*|\mathcal{D})$ in (16) is obtained based on the Gaussian approximation $\tilde{p}_G(\boldsymbol{\tau}_k|\mathcal{D})$ to $p(\boldsymbol{\tau}_k|\mathcal{D})$. It usually gives quite accurate results. The methods to improve Gaussian approximation were discussed in Rue et al. (2009). They can also be used to calculate $p(\tau^*|\mathcal{D})$ from (15).
An alternative way is to use the first integral in (15) to replace $p(\tau^*|\mathcal{D})$ in (14) and perform a multi-dimensional integration using, for example, Laplace approximation; see Appendix C of the supplementary materials for the details.
The second type of prediction is to predict a completely new batch with subject-based covariate $\boldsymbol{u}^*$. We want to predict z^* at $(t^*, \boldsymbol{x}^*)$. The details are omitted here.

3. Consistency
Information consistency has been proved. The details are omitted here.

4. Numerical examples
Simulation studies and several real data examples will be presented.

5. Conclusions
The model discussed in the paper is based on a concurrent regression framework. The idea can be extended to so called function-on-function framework, i.e. the response variable is functional, but the response at each point depends on the whole curve of functional predictor. A linear function-on-function model is discussed in Chapter 16 in Ramsay and Silverman (2005). A special case is that the response is scalar and it depends on functional predictors; see e.g. Chapter 15 in Ramsay and Silverman (2005) or Senturk and Müller (2008).

References

Breslow, N. E., and Clayton, D. G. (1993), "Approximate Inference in Generalized Linear Mixed Models," *Journal of the American Statistical Association*, 88, 9–25.

Fahrmeir, L., and Lang, S. (2001), "Bayesian Inference for Generalized Additive Mixed Models Based on Markov Random Field Priors," *Applied Statistics*, 50, 201–220.

Hall, P., Müller, H.-G., and Yao, F. (2008), "Modelling Sparse Generalized Longitudinal Observations with Latent Gaussian Processes," *Journal of Royal Statistical Society*, Ser. B, 70, 703–723.

Leng, C., Zhang, W., and Pan, J. (2009), "Semiparametric Mean-covariance Regression Analysis for Longitudinal Data," *Journal of the American Statistical Association*, 105, 181–193.

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

Rasmussen, C. E., and Williams, C. K. I. (2006), *Gaussian Processes for Machine Learning*, Cambridge, MA: The MIT Press.

Rice, J. A., and Silverman, B. W. (1991), "Estimating the Mean and Covariance Nonparametrically When the Data are Curves," *Journal of Royal Statistical Society*, Ser. B, 53, 233–243.

Rue, H., Martino, S., and Chopin, N. (2009), "Approximate Bayesian Inference for Latent Gaussian Models Using Integrated Nested Laplace Approximations" (with discussion), *Journal of Royal Statistical Society*, Ser. B, 71, 319–392.

Senturk, D., and Müller, H.-G. (2008), "Generalized Varying Coefficient Models for Longitudinal Data," *Biometrika*, 95, 653–666.

Shi, J. Q., and Choi, T. (2011), *Gaussian Process Regression Analysis for Functional Data*, London: Chapman and Hall/CRC.

Shi, J. Q., Wang, B., Murray-Smith, R., and Titterington, D. M. (2007), "Gaussian Process Functional Regression Modelling for Batch Data," *Biometrics*, 63, 714–723.