Research Article

Statistics in Medicine

Received 28 January 2011, Accepted 24 November 2011 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.4502

Mixed-effects Gaussian process functional regression models with application to dose–response curve prediction

J. Q. Shi,[a] B. Wang,[b] E. J. Will[c] and R. M. West[d*†]

We propose a new semiparametric model for functional regression analysis, combining a parametric mixed-effects model with a nonparametric Gaussian process regression model, namely a mixed-effects Gaussian process functional regression model. The parametric component can provide explanatory information between the response and the covariates, whereas the nonparametric component can add nonlinearity. We can model the mean and covariance structures simultaneously, combining the information borrowed from other subjects with the information collected from each individual subject. We apply the model to dose–response curves that describe changes in the responses of subjects for differing levels of the dose of a drug or agent and have a wide application in many areas. We illustrate the method for the management of renal anaemia. An individual dose–response curve is improved when more information is included by this mechanism from the subject/patient over time, enabling a patient-specific treatment regime. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: management of renal anaemia; functional/longitudinal data; functional regression analysis; Gaussian process functional regression model; individual dose–response curve; mixed-effects model; patient-specific treatment regime; semiparametric model

1. Introduction

Data in many biomedical situations are collected in the form of curves, so it is natural to consider curves as observations and covariates. Methodology focusing on the curves themselves, as the objects of interest, is termed functional data analysis. In many situations, such as the management of renal anaemia discussed in Section 3, measurements are taken at different time points and cannot be regarded as independent observations because measurements on the same subject will be strongly correlated. Fitting smooth curves through the measurement series provides curves suited to functional data analysis (e.g. [1]). In practice, we often have little information about the real physical relationship between the response curves and most of the function-valued covariates, so we propose a new semiparametric model, combining a parametric mixed-effects (ME) model with a nonparametric Gaussian process regression (GPR) model.

Mixed-effects models provide a flexible parametric method for the analysis of multilevel or grouped data, which have applications in many different disciplines particularly in medical research (e.g. [2]). This model can provide explanatory information between the response and the covariates and describe the heterogeneity among different subjects. In functional data analysis, the use of a pure parametric method may be unsatisfactory should there be little information linking the covariates with the response not be strongly justified through physical theory. We need to use a nonparametric or a semiparametric

[a] *School of Mathematics and Statistics, Newcastle University, Newcastle, NE1 7RU, U.K.*
[b] *Department of Mathematics, University of Leicester, Leicester, LE1 7RH, U.K.*
[c] *Department of Renal Medicine, St James's University Hospital, Beckett Street, Leeds, LS9 7TF, U.K.*
[d] *Leeds Institute of Health Sciences, University of Leeds, Leeds, LS2 9JT, U.K.*
**Correspondence to: R. M. West, Centre for Epidemiology and Biostatistics, University of Leeds, Worsley Building, Leeds, LS2 9JT, U.K.*
[†] *E-mail: r.m.west@leeds.ac.uk*

method in this case. A semiparametric method combines a parametric model with a nonparametric counterpart; the former provides some explanatory information, whereas the latter provides a good fit for nonlinear data (e.g. [3]). Because most conventional nonparametric regression models, such as a local polynomial model or a spline smoothing model, suffer from the *curse of dimensionality* and the applications are limited to problems with lower dimensional covariates (usually one-dimensional or two-dimensional) unless some specific structures are assumed, for example, the additive model and the varying coefficient model (e.g. [4, 5]), we will use a nonparametric GPR model combined with a ME model.

The nonparametric GPR model can cope with high-dimensional covariates and dependent data (e.g. [6, 7]). The nonparametric structure is modelled by a covariance kernel, leading to a consistent prediction of the response curve. The proposed semiparametric ME-GPFR models keep those advantages; the regression relationship between the functional response and the covariates not explained by the parametric model can be modelled by the nonparametric GPR model. Note that the covariates could be function valued and of high dimension. Moreover, as we will show in Section 2, the common mean structure and the covariance structure of the ME-GPFR models are learned from the data collected from all subjects, but the prediction for each individual includes two parts. The first part is based on the common mean structure, and the second part is calculated based on the data collected from that particular individual via the covariance kernel. This leads to a consistent prediction for each individual response curve, which is a major advantage of GPR related models, whereas a parametric model usually can only provide consistent estimates of the parameters involved. This property is particularly useful when it is used in dose–response studies.

The primary aim of dose–response studies is to characterize the change in subject response for differing doses of a drug or agent. This has wide applications in medicine, epidemiology, pharmacology, toxicology, nutrition and other areas. When the ME-GPFR models are used, we can construct a dose–response curve for each individual patient, whereas most of the current studies can only provide a common dose–response curve (i.e. the common curve for an 'average' patient classified by age group, sex and others, but it can not cope with circumstances of each individual patient, which may change over time); see, for example, the discussion in [8]. Our method therefore enables the planning of a patient-specific treatment regime, and that treatment regime can be improved over time as more data are collected for each of individual patient. We provide more discussion in Section 3 when the method is applied to a real data example for the management of renal anaemia.

We organize this paper as follows. We define and discuss the ME-GPFR models are defined and discussed in Section 2.1. We discuss an empirical Bayes learning approach and prediction in Sections 2.2 and 2.3, respectively. We also discuss the methodology for constructing dose–response curves by using the ME-GPFR models in Section 2.3. In Section 3, we apply the methodology for dose–response study to the Leeds renal anaemia data. We provide discussion and further developments in Section 4.

2. Methodology

2.1. Mixed-effects Gaussian process functional regression models

We now consider a problem of functional regression analysis with M replications (each replication is associated with a subject). We consider two types of variables: one is functional (or function-valued) variables and the other is scalar variables. A functional variable $x(t)$ (also called a curve) denotes a variable associated with t, which is time in this paper. For example, the haemoglobin (Hb) levels shown in Figure 1(a) for Leeds renal anaemia data are measurements from a functional variable. A scalar variable denotes an ordinary variable, providing information for each individual subject such as patient's age and sex. We can find the details in [1, 7].

Let $y_m(t)$ be the response curve of the mth subject for $m = 1, \ldots, M$, $\boldsymbol{x}_m(t)$ be a set of function-valued covariates and $\boldsymbol{u}_m$ a set of scalar covariates. Here, t is time or any other one-dimensional function-valued covariate, and it could be one of covariates in $\boldsymbol{x}$. We now consider a functional regression problem with functional response variable and mixed functional and scalar covariates [1]; that is, we want to find a functional regression method to model $y_m(t)$ by $\boldsymbol{x}_m(t)$ and $\boldsymbol{u}_m$. We propose to use the following ME-GPFR model:

$$y_m(t) = \boldsymbol{u}_m'\boldsymbol{\beta}(t) + \boldsymbol{v}_m'(t)\boldsymbol{\gamma} + \boldsymbol{w}_m'(t)\boldsymbol{b}_m + \tau_m(\boldsymbol{z}_m(t)) + \varepsilon_m(t), \tag{1}$$

$$\boldsymbol{b}_m \sim N(0, \boldsymbol{\Sigma}),\ \ \varepsilon_m(t) \sim N(0, \sigma_\varepsilon^2), \tag{2}$$

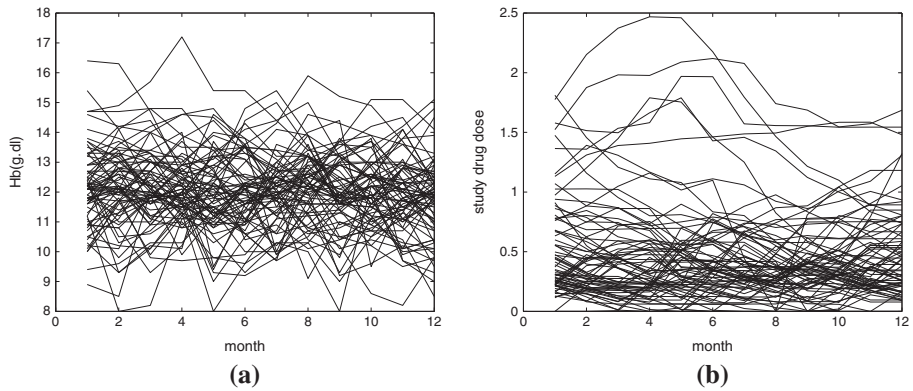


Figure 1. Leeds renal data: (a) the Hb levels recorded for 74 patients for the period of 12 months; and (b) dose levels of the agent DA received by the patients in the same period.

where $\varepsilon_m(t)$ are random errors and are independent at different times and $\boldsymbol{v}_m(t)$, $\boldsymbol{w}_m(t)$ and $z_m(t)$ are subsets of function-valued covariates $\boldsymbol{x}_m(t)$ with dimensions r, k and Q, respectively. Here, the first term on the right-hand side in (1), that is, $\boldsymbol{u}_m'\boldsymbol{\beta}(t)$, depends on a p-dimensional nonfunctional covariate $\boldsymbol{u}_m$. This is a linear functional regression model discussed in [1]. Here, $\boldsymbol{\beta}(t)$ is a set of functional coefficients. To calculate their estimates, we use a B-spline approximation; see the details in the next subsection.

The regression relationship between the response $y_m(t)$ and the function-valued covariates $\boldsymbol{x}_m(t)$ is described by the other terms in (1), including a parametric ME part of $\boldsymbol{v}_m'(t)\boldsymbol{\gamma} + \boldsymbol{w}_m'(t)\boldsymbol{b}_m$ and a nonparametric part $\tau_m(z_m(t))$ or $\tau_m(z_m)$. In the ME part, the first term stands for the fixed effect and the coefficient $\boldsymbol{\gamma}$ is the same for all subjects, whereas the second one stands for the random-effect part. The heterogeneity among different subjects is quantified by $\boldsymbol{\Sigma}$. We assume it is a diagonal matrix of $\text{diag}(\sigma_1^2, \ldots, \sigma_k^2)$. The parametric part is very useful in practice particularly in medical research because it describes the explicit relationship between the functional response and those function-valued covariates and provides some explanatory information. However, because of the complexity of many problems, a parametric model is usually difficult to fully describe such a functional regression relationship. In (1), we use a nonparametric GPR model to describe the part, which is unexplained by the parametric model. The GPR model is defined by

$$\tau_m(z_m) \sim \text{GPR}[0, k(\cdot,\cdot;\boldsymbol{\theta})|z_m]. \tag{3}$$

Specifically, the aforementioned model asserts that $\tau(z_1)$ and $\tau(z_2)$ have a normal distribution for any z_1 and z_2 (we omit the subscript m temporarily for easy explanation). Without loss of generality, we can assume zero means. The covariance between $\tau(z_1)$ and $\tau(z_2)$ is determined by a covariance kernel $k(\cdot,\cdot)$: $\text{Cov}(\tau(z_1), \tau(z_2)) = k(z_1, z_2; \boldsymbol{\theta})$.

One example is the following squared exponential covariance function

$$k(z_1, z_2; \boldsymbol{\theta}) = v_0 \exp\left\{-\frac{1}{2}\sum_{q=1}^{Q} w_q(z_{1q} - z_{2q})^2\right\} + a_0, \tag{4}$$

where z_q is the qth element of z and $\boldsymbol{\theta} = (a_0, v_0, w_1, \ldots, w_Q)$ is the vector of the hyperparameters. Some other commonly used covariance functions include the powered exponential, the rational quadratic and the Matérn covariance function [7]. The hyperparameter $\boldsymbol{\theta}$ involved in the covariance kernel plays a role similar to the smoothing parameters in a spline model. We can use generalized cross-validation, empirical Bayesian learning or other methods to choose its value. On the basis of the aforementioned GPR model, if we have observed a set of data including n pairs (τ_i, z_i), $i = 1, \ldots, n$, then $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_n)'$ has an n-variate normal distribution $N(\boldsymbol{0}, \boldsymbol{C}(\boldsymbol{\theta}))$, where the (i, j)th element of $\boldsymbol{C}$ is calculated by $k(z_i, z_j; \boldsymbol{\theta})$ using the covariance kernel given in (4). We can find a detailed discussion of the GPR model and the selection of covariance kernels in, for example, [6, 7].

The ME-GPFR models defined in (1)–(3) have several important features. First, it is a semiparametric model, combining the advantages of providing explanatory information based on the parametric part and providing a good-fit to the data based on the nonparametric part. Second, the model can cope with

mixed functional and scalar covariates. Third, it provides a natural framework on simultaneously modelling common mean structure across different subjects and covariance structure for each individual; the feature is particularly important for dose–response study because the covariance structure can be used to model the individual characteristic for each patient nonparametrically. This enables the planning of a patient-specific treatment regime. We will give the detailed discussion in the following subsections. Moreover, compared with other conventional nonparametric regression models, the GPR-related models have some peculiar advantages. It can cope with multidimensional functional covariates, and the prior specification of covariance kernel enables the model to accommodate a wide class of nonlinear functions.

The ME-GPFR model defined previously is indeed very flexible. If we exclude the first term and the term of $\tau_m(z_m(t))$ in (1), it is a conventional ME model. If we exclude the terms in (1) other than the first one and the random errors, it is the linear functional regression model discussed in Chapter 13 in [1]. If the ME part is excluded, it is the GPFR model discussed in [9, 10].

In the next subsection, we will discuss how to estimate all the unknown parameters involved in both the mean and the GP covariance kernel.

2.2. *Empirical Bayes approach*

Suppose that N_m observations are obtained for the mth subject, and the data collected for the mth subject are

$$\mathcal{D}_m = \{(y_{mi}, t_{mi}, \boldsymbol{x}_{mi}) \text{ for } i = 1, \ldots, N_m;\ \boldsymbol{u}_m\}.$$

The observed data for all subjects are denoted as $\mathcal{D}$. Recall that $\boldsymbol{v}_m(t)$, $\boldsymbol{w}_m(t)$ and $z_m(t)$ are subsets of $\boldsymbol{x}_m(t)$.

It is not straightforward to estimate the functional coefficient $\boldsymbol{\beta}(t)$ involved in the mean function. Here, we use a B-spline approximation; see, for example, [11]. Let $\boldsymbol{\Phi}(t) = (\boldsymbol{\Phi}_1(t), \ldots, \boldsymbol{\Phi}_D(t))'$ be a set of B-spline basis functions, then the coefficient function $\boldsymbol{\beta}(t)$ can be approximated by $\boldsymbol{B}'\boldsymbol{\Phi}(t)$. Let $\boldsymbol{y}_m$ be a vector of $\{y_{mi}, i = 1, \ldots, N_m\}$, $\boldsymbol{t}_m = \{t_{mi}, i = 1, \ldots, N_m\}$, $\boldsymbol{V}_m$ be an $N_m \times r$ matrix with the ith row $\boldsymbol{v}'_m(t_{mi})$ and $\boldsymbol{W}_m$ be an $N_m \times k$ matrix with the ith row $\boldsymbol{w}'_m(t_{mi})$, then the discrete form of the ME-GPFR models based on the N_m observations for the mth subject is expressed by the following:

$$\boldsymbol{y}_m = \boldsymbol{\Phi}_m \boldsymbol{B} \boldsymbol{u}_m + \boldsymbol{V}_m \boldsymbol{\gamma} + \boldsymbol{W}_m \boldsymbol{b}_m + \boldsymbol{\tau}_m + \boldsymbol{\varepsilon}_m, \tag{5}$$

where $\boldsymbol{\Phi}_m$ is an $N_m \times D$ matrix with the (i, d)th element $\boldsymbol{\Phi}_d(t_{mi})$, $\boldsymbol{\varepsilon}_m \sim N(\boldsymbol{0}, \sigma_\varepsilon^2 \boldsymbol{I})$, $\boldsymbol{\tau}_m \sim N(\boldsymbol{0}, \boldsymbol{C}_m(\boldsymbol{\theta}))$, and $\boldsymbol{C}_m(\boldsymbol{\theta})$ is an $N_m \times N_m$ covariance matrix with (i, j)th element calculated by $k(z_{mi}, z_{mj}; \boldsymbol{\theta})$ in (4) if we use the squared exponential covariance kernel. The unknown parameters involved in the ME-GPFR models include the B-spline coefficient $\boldsymbol{B}$, the fixed effect coefficient $\boldsymbol{\gamma}$, the random-effect covariance matrix $\boldsymbol{\Sigma}$ and the variance σ_ε^2 of measurement error. As discussed in the previous section, $\boldsymbol{\theta}$ is a set of hyperparameters involved in the covariance kernel. We may use an empirical Bayes approach to select their values. The idea is to choose the value of $\boldsymbol{\theta}$ by maximizing its marginal likelihood; see the detailed discussion in [10, 12]. The hyperparameters as well as the other parameters can therefore be estimated by maximizing the following log likelihood:

$$\begin{aligned} l(\boldsymbol{B}, \boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \sigma_\varepsilon^2) = \sum_{m=1}^{M} \Big\{ & -\frac{1}{2} N_m \log(2\pi) - \frac{1}{2} \log |\boldsymbol{\Omega}_m| \\ & -\frac{1}{2} (\boldsymbol{y}_m - \boldsymbol{\Phi}_m \boldsymbol{B} \boldsymbol{u}_m - \boldsymbol{V}_m \boldsymbol{\gamma})' \boldsymbol{\Omega}_m^{-1} (\boldsymbol{y}_m - \boldsymbol{\Phi}_m \boldsymbol{B} \boldsymbol{u}_m - \boldsymbol{V}_m \boldsymbol{\gamma}) \Big\}, \end{aligned} \tag{6}$$

where $\boldsymbol{\Omega}_m = \boldsymbol{W}_m \boldsymbol{\Sigma} \boldsymbol{W}'_m + \boldsymbol{C}_m(\boldsymbol{\theta}) + \sigma_\varepsilon^2 \boldsymbol{I}$.

In Appendix A, the first order derivatives of $l(\boldsymbol{B}, \boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \sigma_\varepsilon^2)$ in terms of $\mathrm{vec}(\boldsymbol{B})$ and $\boldsymbol{\gamma}$ are respectively given by (12) and (13), where $\mathrm{vec}(\boldsymbol{B})$ denotes the stacked columns of $\boldsymbol{B}$. Letting $\partial l / \partial \mathrm{vec}(\boldsymbol{B}) = 0$ and $\partial l / \partial \boldsymbol{\gamma} = 0$, we can get the explicit forms for $\boldsymbol{B}$ and $\boldsymbol{\gamma}$ given $\boldsymbol{\theta}$, $\boldsymbol{\Sigma}$, σ_ε^2 as

$$\mathrm{vec}(\boldsymbol{B}) = F_1(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \sigma_\varepsilon^2), \quad \boldsymbol{\gamma} = F_2(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \sigma_\varepsilon^2). \tag{7}$$

We also provide the derivation and the formulae of F_1 and F_2 in Appendix A. We propose to use an iterative procedure in which the parameters $\boldsymbol{B}$ and $\boldsymbol{\gamma}$ and the remainders are estimated in turn. Each iteration includes the following two steps:

(i) Update $\boldsymbol{B}$ and $\boldsymbol{\gamma}$ by (7) given the current values of $\boldsymbol{\theta}$, $\boldsymbol{\Sigma}$ and σ_ε^2; and
(ii) Update $\boldsymbol{\theta}$, $\boldsymbol{\Sigma}$ and σ_ε^2 by maximizing l in (6) given $\boldsymbol{B}$ and $\boldsymbol{\gamma}$.

We can repeat the aforementioned two steps several times within each iteration. This speeds up convergence from our experience in empirical studies. There are no explicit expressions in step (ii), but we can use the gradients to speed up the maximization. We provide the expressions of the gradients in (16) to (18) in Appendix A. We denote the maximum likelihood estimates by $\hat{\boldsymbol{\Theta}}$, where $\boldsymbol{\Theta} = (\boldsymbol{B}, \boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \sigma_\varepsilon^2)$. The estimate of the functional coefficient is given by $\hat{\boldsymbol{\beta}}(t) = \hat{\boldsymbol{B}}'\boldsymbol{\Phi}(t)$.

After all the parameters are estimated by the aforementioned empirical Bayes approach, the common mean part is given by $\tilde{\mu}_m(t) = \boldsymbol{u}_m'\hat{\boldsymbol{B}}'\boldsymbol{\Phi}(t) + \boldsymbol{v}_m'(t)\hat{\boldsymbol{\gamma}}$. The covariance structure is determined by $\hat{\boldsymbol{\Sigma}}$, $\hat{\boldsymbol{\theta}}$ and the covariance kernel given in (3). Both mean structure and covariance structure are learned from the data collected from all subjects. The mean part can be used as a prediction. The second part is calculated on the basis of the data collected from each individual subject using the aforementioned learned covariance model, which improves the curve prediction and leads to a consistent prediction. We provide the details in the next subsection.

2.3. *Prediction and dose–response curve*

We now consider the prediction problem by assuming that the mean and covariance structures of the ME-GPFR models have been given. Suppose that we have already observed some data for a subject (e.g. for a new patient), the $(M+1)$th subject say, and want to predict the output y^* for a new set of inputs $(t^*, \boldsymbol{x}^*)$ with $\boldsymbol{x}^* = \boldsymbol{x}(t^*)$ at a new test point t^*. This new time point could be at a past time but without an observation (i.e. a missing observation) but more often will be a future time. The latter is certainly of most interest, meaning we want to predict the response curve for future times (this is so-called *extrapolation*). Thus, in addition to the data observed from the first M subjects, we assume that N observations are obtained for this $(M+1)$th new subject, providing data

$$\mathcal{D}_{M+1} = \{(y_{M+1,i}, t_{M+1,i}, \boldsymbol{x}_{M+1,i}) \text{ for } i = 1, \ldots, N; \text{and } \boldsymbol{u}_{M+1}\}.$$

We therefore now have data $\mathcal{D} = \{\mathcal{D}_1, \ldots, \mathcal{D}_M, \mathcal{D}_{M+1}\}$.

From (6), the random-effect part, random measurement error and Gaussian process error can be integrated into a new Gaussian process, denoted by $\tilde{\tau}(\boldsymbol{z}_m, \boldsymbol{w}_m)$, with zero mean and covariance structure

$$\tilde{C}(\boldsymbol{z}_i, \boldsymbol{z}_j, \boldsymbol{w}_i, \boldsymbol{w}_j; \boldsymbol{\theta}, \boldsymbol{\Sigma}, \sigma_\varepsilon^2) = k(\boldsymbol{z}_i, \boldsymbol{z}_j; \boldsymbol{\theta}) + \sum_{q=1}^{k} \sigma_q^2 w_{iq} w_{jq} + \sigma_\varepsilon^2 \delta_{ij}, \tag{8}$$

where δ_{ij} is the Kronecker delta and $k(\boldsymbol{z}_i, \boldsymbol{z}_j; \boldsymbol{\theta})$ is the kernel covariance function in (3). We can then write models (1) and (2) as

$$y_m(t) = \mu_m(t) + \tilde{\tau}(\boldsymbol{z}_m(t), \boldsymbol{w}_m(t)), \quad \text{with } \mu_m(t) = \boldsymbol{u}_m'\boldsymbol{\beta}(t) + \boldsymbol{v}_m'(t)\boldsymbol{\gamma}. \tag{9}$$

For the $(M+1)$th subject, an estimate of the mean at time point t is given by

$$\tilde{\mu}_{M+1}(t) = \boldsymbol{u}_{M+1}'\hat{\boldsymbol{B}}'\boldsymbol{\Phi}(t) + \boldsymbol{v}_{M+1}'(t)\hat{\boldsymbol{\gamma}}. \tag{10}$$

If we assume that the covariance structures at data points $t_{M+1,i}$ for $i = 1, \ldots, N$ and at the new data point t^* are the same, that is, both $\{y_{M+1,i}, i = 1, \ldots, N\}$ and y^* are associated with the same Gaussian process with zero mean and covariance kernel (8), the prediction at t^* is given by

$$\hat{y}^* = \mathrm{E}(y^*|\mathcal{D}) = \tilde{\mu}_{M+1}(t^*) + \boldsymbol{H}'(\boldsymbol{y}_{M+1} - \tilde{\boldsymbol{\mu}}_{M+1}(\boldsymbol{t})), \tag{11}$$

where $\boldsymbol{t} = (t_{M+1,1}, \ldots, t_{M+1,N})$. We provide the proof in Appendix B. The predictive variance is given by (20). We obtain the first term on the right-hand side in the aforementioned equation from the mean model, whereas the second term depends on the given covariance kernel (8) and the data collected from the $(M+1)$th subject. This part is used to model the individual characteristic for each subject. It is equal to zero if no data are observed for the subject.

We now discuss briefly the problem of consistency. There are two issues here. One is related to the common mean in (10), and the other is related to the $(M+1)$th curve $y_{M+1}(t)$ itself. We estimate the common mean structure from the data collected from all M subjects. It is a consistent estimator of

the true mean structure under some regularity conditions; see the details in [1]. If we have not observed any data for the new subject, we can only use the common mean $\tilde{\mu}_{M+1}(t)$ as a prediction, which is obviously not a consistent estimator of $y_{M+1}(t)$, the response curve for the particular subject.

The consistency of $\hat{y}_{M+1}(t)$ as given in (11) to $y_{M+1}(t)$ can only be achieved by the observations collected from the $(M+1)$th subject through the GPR model defined in (3) or the second term in Equation (11). With the theory developed in [13] and [14], if N, the sample size of the data collected from the subject, is sufficiently large and if the kernel covariance function satisfies certain regularity conditions (e.g. the kernel covariance function in (4)), $\hat{y}_{M+1}(t)$ is consistent to $y_{M+1}(t)$.

The GPR model is a nonparametric approach. The achievement of consistency does not depend on the common mean structure and the choice of the values of hyperparameters involved in the covariance kernel, although the empirical Bayes approach discussed in the previous subsection leads to the consistency of common mean structure and would improve the accuracy of prediction in (11) for finite sample size. With fewer data for a particular subject, a prediction will be mainly given on the basis of the common mean structure. If more data is collected for that subject, the prediction given in (11) would be improved tending to its own response curve $y_{M+1}(t)$. It can therefore be used to construct an individual dose–response curve and so to plan a patient-specific treatment regime.

We construct the dose–response curve as follows. We first use the data to learn the mean and covariance structures as discussed earlier. Then, for a particular subject at a certain time point t, we assume that the data collected so far is $\mathcal{D}_{M+1}$, in which $\boldsymbol{x}_{M+1}(t)$ includes a covariate that is the dosage for a certain drug namely $d(t)$. Then, for each possible value of $d(t)$ along with the values of other covariates (i.e. all the information collected for this particular subject up to time t), we predict the value of y^* at a future time t^*. For different dose levels of $d(t)$, we may obtain a different value of prediction, and indeed, we anticipate this to be so. We call the curve of prediction $\hat{y}^*$ against the different dose levels the *dose–response curve*.

Note that the dose–response curve will be tailored for each subject and to a specified time point as well as the amount of data available. Hence, this development of theory delivers important refinements to the prescription of drugs and agents.

3. Dose–response study applied to the management of renal anaemia

Patients with reduced kidney function not only require dialysis to remove waste products from their blood but also produce less erythropoietin (EPO), the natural stimulus to the production of red blood cells in the bone marrow. As a consequence, most dialysis patients suffer renal anaemia to some degree. This can be effectively treated with an exogenous epoetin. Injections are given subcutaneously or intravenously with either a synthetic EPO, for example, EPO beta, or a modified epoetin such as darbepoetin alpha (DA).

The dose of epoetin to be given to each patient is determined by monitoring the Hb concentration from a blood sample taken typically every 2 or 4 weeks. Over the last decade in Leeds, blood samples are taken on a monthly basis and the epoetin dose is calculated using a strict management algorithm [15, 16]. This represents a clinical decision support system. Patients need to have their Hb levels controlled within relatively narrow limits [17]. If Hb levels are too low, then patients become symptomatic of anaemia and, if too high, then there may be prothrombotic risks to their dialysis treatment and vascular tree. The primary therapeutic concern is how to maintain the Hb level of each patient by giving a suitable dose of exogenous epoetin.

We looked at the data collected from 74 patients who received DA. They were followed-up for the original study period of 9 months with a further 3-month extension. With initial (month 0) readings, this gave 13 measurements of Hb, epoetin dose and administered iron supplements. Throughout, all doses were determined by the same clinical decision support system. Note that an Hb of around 11.8 g/dl is considered to be the 'target' level for dialysis patients, a little lower than the normal range for healthy patients.

In this example, the response is Hb level for a patient. Figure 1(a) shows the monthly Hb measurements for the 74 patients. One of the covariates is the dose level of epoetin DA, the measurements of which are shown in Figure 1(b) for all the patients. We will apply the ME-GPFR models and conduct a dose–response study. The main purpose of the experiment is to assess the control of Hb levels in patients each dosed with the agent DA. This is not entirely straightforward. Contributions to the variation in Hb levels may also arise from further deterioration in kidney function, iron status and intercurrent complications such as infection, surgery and hemorrhage. Iron status is usually assured with sufficient doses

of iron, but there is often little control over other aspects. Taking a functional data analysis approach to the data from this trial, West *et al.* [18] demonstrated that control of renal anaemia can be assessed graphically. Phase plots relating the derivatives of the Hb trajectories identified those patients that were well controlled by the management regime and those that were less well controlled. The current study aims to extend that analysis and quantify the degree of control.

We will provide predictions of Hb levels and assess the performance of the models and then demonstrate how to obtain individual dose–response curves.

3.1. Prediction of haemoglobin

We illustrate the basic model structure in Figure 2, where $y(t)$ is the response variable—Hb level at time t, depending on the Hb level at time $t-1$, dose level d before time t and other input covariates $\boldsymbol{z}(t)$. We use the ME-GPFR models defined in (1)–(3) with the covariance function (4). For the purpose of comparison, we consider six different models as listed in Table I, in which FE and ME are the conventional fixed-effects and mixed-effects models, respectively, and the others are the related semiparametric models with GPR. The details are given as follows:

$$\begin{aligned}
\text{FE}: &\quad y_m(t) = \boldsymbol{v}_m'(t)\boldsymbol{\gamma} + \epsilon_m(t);\\
\text{ME}: &\quad y_m(t) = \boldsymbol{v}_m'(t)\boldsymbol{\gamma} + \boldsymbol{w}_m'(t)\boldsymbol{b}_m + \epsilon_m(t);\\
\text{FE-GPFR}: &\quad y_m(t) = \boldsymbol{v}_m'(t)\boldsymbol{\gamma} + \tau_m(\boldsymbol{z}_m(t)) + \epsilon_m(t);\\
\text{ME-GPFR}: &\quad y_m(t) = \boldsymbol{v}_m'(t)\boldsymbol{\gamma} + \boldsymbol{w}_m'(t)\boldsymbol{b}_m + \tau_m(\boldsymbol{z}_m(t)) + \epsilon_m(t).
\end{aligned}$$

The FE part involves the covariate $\boldsymbol{v}_m(t) = \{1, t, d_m(t-2), d_m^2(t-2)\}$. Here, t is the month and $d_m(t-2)$ is the dose level for subject m at month $t-2$. We found that the Hb level $y_m(t)$ has the highest correlation with dose $d_m(t-2)$ (i.e. the dosage of the drug taken between 30–60 days ago), although it is also related to $d_m(t-1)$ and $d_m(t-3)$. For clarity of exposition, we restrict our analysis to regression upon $d_m(t-2)$ in this paper. The random-effect part involves the covariate $\boldsymbol{w}_m(t)$, which are listed in Table I. In the related covariance structure, the covariates include those given in ME2.

We implemented the experiment using leave-one-out cross-validation (CV). We used all the data collected from 73 (of 74) patients and those up to month m from the remaining patient to train the models (i.e. estimate all the unknown parameters involved in the models), then predicted the value of Hb level at month $m+2$ for this specific patient, using the true dose level at month m. This is common practice in a dynamic model; see, for example, [19]. We calculated the predictions from months 8 to 12 for each subject, that is, for $m = 6, \ldots, 10$. The root of the mean squared errors (RMSE) between the predicted

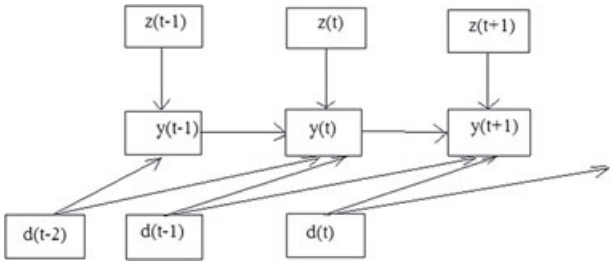


Figure 2. Diagram of the model structure for Leeds renal data: $y(t)$, $d(t)$ and $z(t)$ represent the Hb level, the dose level and other input covariates at time t, respectively.

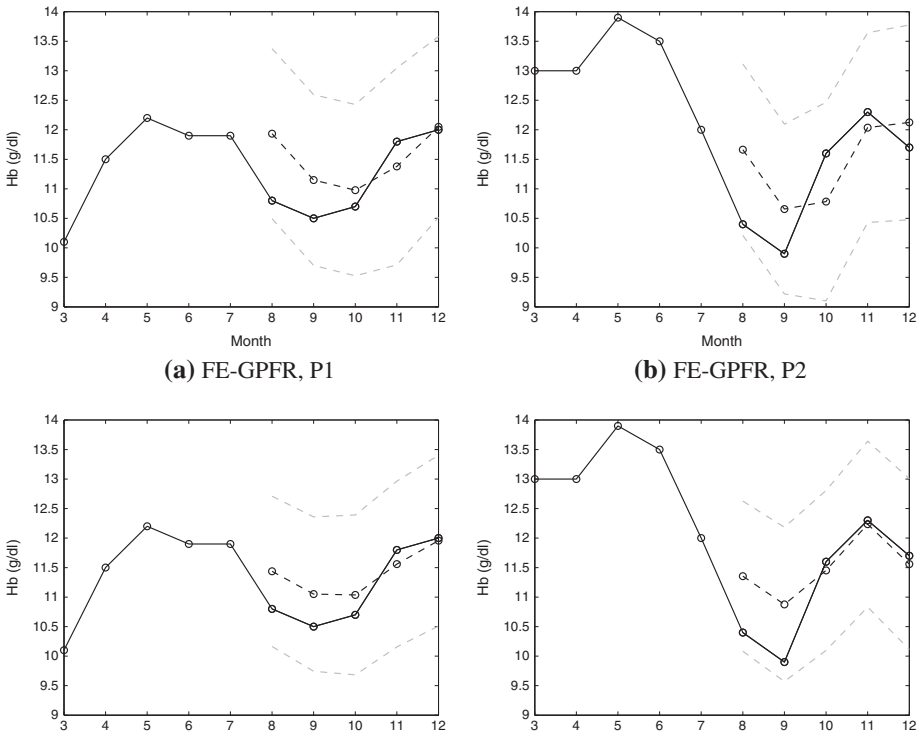
Table I. The values of RMSE by leave-one-out CV and BIC for the different models.

Models	$\boldsymbol{w}_m(t)$	RMSE	BIC ($\times 10^3$)
FE		1.1509	2.4811
ME	$1, d_m(t-2)$	1.0349	2.3403
ME2	$1, d_m(t-2), y_m(t-1), y_m(t-2)$	0.9652	2.3536
FE-GPFR		0.8873	2.1917
ME-GPFR	$1, d_m(t-2)$	0.8732	1.7242
ME-GPFR2	$1, d_m(t-2), y_m(t-1), y_m(t-2)$	0.8732	1.7308

and the real observations was used to judge the performance of the model; see, for example, [6]. The results reported in Table I are the average values of RMSE for all the patients calculated by using the aforementioned procedure. We also report the values of BIC [20]. Figure 3 shows the predictions by FE-GPFR and ME-GPFR and the real observations for two randomly selected patients.

The nonparametric GPR involved in model (1) plays an important role in prediction. By comparing conventional FE and ME models (ME and ME2) with the semiparametric GPFR (FE-GPFR, ME-GPFR and ME-GPFR2) models, the average values of RMSE for the models FE, ME and ME2 are much larger than those obtained from the related semiparametric models. This is the evidence that the nonparametric GPR can model the unknown nonlinear functional regression relationship and therefore improves model fit and prediction. This finding is consistent with the results reported in [10]. The mean model synthesizes the information from different subjects at each time point, whereas the covariance component of the model tunes the prediction on the basis of the information collected for each individual. Therefore, model (1) gives very accurate values of prediction.

With the results given in Table I, it is obvious that the model ME-GPFR, which has the smallest RMSE and BIC values, is preferable. By using all the data in model training, the value of RMSE given by ME-GPFR is even smaller (0.7412). Table II shows the estimates of $\boldsymbol{\gamma}$ in the FE part and σ_i^2 in random-effect part. We also report the P-values to test whether the coefficients are equal to zero. It shows that almost



(a) FE-GPFR, P1 **(b)** FE-GPFR, P2

(c) ME-GPFR, P1 **(d)** ME-GPFR, P2

Figure 3. Plots of the prediction of Hb against time (in month) for two randomly selected patients P1 and P2: the solid lines stand for real observations, and the dashed lines stand for predictions and the 95% predictive intervals.

Table II. Estimates of coefficients for the ME-GPFR model.

Fixed-effect part				Random-effect part			
Covariate	Estimate	Standard error	P-value	Covariate	Estimate	Standard error	P-value
Constant	11.0924	0.1494	0.0000	Constant	0.6392	0.1680	0.0001
t	−0.0147	0.0135	0.2778	$d_m(t-2)$	2.6269	0.7860	0.0008
$d_m(t-2)$	3.1033	0.4314	0.0000				
$d_m^2(t-2)$	−0.7786	0.2756	0.0047				

all the coefficients are significant. We also examined models with more functional covariates but found that the model cannot be further improved.

3.2. Dose–response relationships

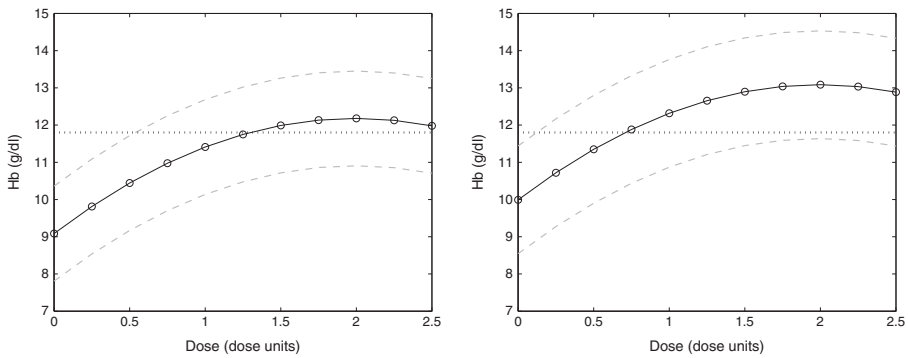
The dose–response relationship describes the change in effect on an organism or patient, for example, the Hb level in our renal anaemia case study, caused by differing doses. The proposed models discussed can be used to construct a dose–response curve for each *individual* and therefore support clinicians to prescribe suitable doses and so control the response value within a certain range for each *individual* patient. This allows a so-called *patient-specific treatment regime*.

We now discuss the dose–response study illustrated by our renal anaemia data. This is to obtain a predictive individual dose–response curve, that is, prediction of responses for different doses. We take 11 different doses from 0 to 2.5 and then calculate the predicted values for each of those dose levels. We provide an example in Figure 4 for a representative patient. We present the predicted values of Hb for month 8 on the basis of the data collected at the first 6 months in panel (a) and the predicted values for month 12 on the basis of the data collected at the first 10 months in panel (b). The x-axis stands for the different levels of drug DA, whereas the y-axis stands for the related predicted values of Hb. This is a typical dose–response curve. Most studies can only provide a common curve without considering special features for each individual, but Figure 4(a) gives the individual dose–response curve on the basis of the common model structure learned from the data collected from all patients and the particular information for the individual collected in the first 6 months, whereas Figure 4(b) is analogous to Figure 4(a) but on the basis of the information collected in the first 10 months for the patient. This is very useful in practice, and the clinician may prescribe a suitable dose based on the figures. In Figure 4, the dotted lines are the target control level of Hb. Thus, a suitable dose level of drug DA taken by the selected patient after month 6 should be around 1.4 units and be around 0.7 units after month 10.

The change in the relationship shown in Figure 4 might have arisen for the following reasons and illustrates advantages of the ME-GPFR model:

- Initially, the dose–response curve is similar to the mean curve for all patients, but more data from that specific patient enables a 'correction' that enables the dose to be reduced.
- The circumstances of the patient may change over time, with perhaps improved management of iron status or recovery from infection.

Note that the fixed effect for dose is restricted to have a quadratic form. The results, typically illustrated in Figure 4, confirm that this is a sufficient representation, consistent with clinical opinion. Specifically, there is a plateau effect at around 1.5 dose units—almost no benefit is gained by prescribing above 1.5 units. The slight dip above 2 units is not relevant because it is negligible and is for a dose level outside of the usual prescription range. More important is the plateau feature that the quadratic form does capture efficiently in terms of the number of parameters.



(a) The 8-th month **(b)** The 12-th month

Figure 4. The dose–response curves for two randomly selected patients. The y-axis is the prediction of the Hb response and the x-axis the different dose levels for agent DA (in dose unit)—the solid lines stand for predictions with different dose levels, the dashed lines stand for their 95% predictive intervals, and the dotted lines stand for the target control level of Hb = 11.8.

4. Discussion

We proposed a semiparametric ME-GPFR model to construct individual dose–response curves for a patient-specific treatment regime. We used a nonparametric GPR model, which can model a response curve with a general shape in terms of a set of functional covariate curves. The model can therefore be used to address a wide class of problems with high-dimensional functional or longitudinal data. Mean and covariance structures are modelled simultaneously and so combine the information acquired from other subjects and the information collected from each individual. The subject-specific prediction and dose–response can be constructed and improved over time as more individual data become available. We advocated the use of the empirical Bayes learning approach as discussed in Section 2.3 (the MATLAB (The MathWorks Inc., Natick, MA, USA) codes are available for download at http://www.staff.ncl.ac.uk/j.q.shi). We can also use a fully Bayes approach; see, for example, Chapter 3 in [7] and [21].

Although we only considered the ME model in this paper, the semiparametric models defined in (1)–(3) can cope with other parametric models. The model learning and prediction discussed in subsections 2.2 and 2.3 can be extended accordingly without significant difficulties. The GPR model for $\tau(z_m)$ can also be replaced by other nonparametric models such as the additive model and the varying-coefficient model, although the GPR model is preferable when the dimension of z_m is large based on the arguments made early.

For the illustrative example of renal anaemia, we provide a comparison of the dose–response curves constructed for months 8 and 12 in Figure 4 and interestingly illustrate how the dose for an individual changes. Initially, the common dose–response curve dominates the dose decision, but this changes as more information on that patient becomes available. Further, in the example, there is a change in required dose over time: a temporal effect in the response.

An interesting problem worthy of further exploration is how to find an optimal solution for dose level to maintain the response curve in a given range under certain conditions, such as minimising the change of dosage in successive months. This is important so that dose titration is not applied to stable patients, with the risk of provoking or worsening a periodicity of response. The solutions depend on the choice of decision criterion. This is an active area of research; see, for example, [22–25].

Generally, the proposed methodology has a wide range of applications wherever medications are used for the long-term maintenance of patient health, for example, warfarin (or new Boehringer drug Pradaxa (Boehringer, Ingelheim am Rhein, Germany)) for management of anticoagulation and numerous diabetic medicines for glucose control. All these patients are individually managed. To date, information from the response of other patients is incorporated through clinician experience and management algorithms are based on a population-level response. There is an opportunity to refine management of many long-term conditions by tailoring dose–response curves to individuals from the combination of group and individual data.

If the data are collected from different sources, hierarchical mixture models may be considered (e.g. [26]). This presents the opportunity to cluster the patient responses by longitudinal latent class and use response-curve clustering to inform individual patient response.

APPENDIX A. Derivation of the formulae for calculating the maximum likelihood estimates

The log likelihood for $\boldsymbol{\Theta} = (\boldsymbol{B}, \boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \sigma_\varepsilon^2)$ is given by (6). After a straightforward calculation and simplification, we obtain

$$\frac{\partial l}{\partial \mathrm{vec}(\boldsymbol{B})} = \sum_{m=1}^{M} (\boldsymbol{y}_m - \boldsymbol{\Phi}_m \boldsymbol{B} \boldsymbol{u}_m - \boldsymbol{V}_m \boldsymbol{\gamma})' \boldsymbol{\Omega}_m^{-1} (\boldsymbol{u}_m \otimes \boldsymbol{\Phi}_m')', \tag{12}$$

$$\frac{\partial l}{\partial \boldsymbol{\gamma}} = \sum_{m=1}^{M} (\boldsymbol{y}_m - \boldsymbol{\Phi}_m \boldsymbol{B} \boldsymbol{u}_m - \boldsymbol{V}_m \boldsymbol{\gamma})' \boldsymbol{\Omega}_m^{-1} \boldsymbol{V}_m, \tag{13}$$

where $\otimes$ denotes Kronecker product. Letting $\partial l / \partial \mathrm{vec}(\boldsymbol{B}) = 0$ and $\partial l / \partial \boldsymbol{\gamma} = 0$, we can get the explicit forms for $\boldsymbol{B}$ and $\boldsymbol{\gamma}$ as

$$\mathrm{vec}(\boldsymbol{B}) = F_1(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \sigma_\varepsilon^2) = (\boldsymbol{A}_{11} - \boldsymbol{A}_{12} \boldsymbol{A}_{22}^{-1} \boldsymbol{A}_{21})^{-1} (\boldsymbol{y}_{(1)} - \boldsymbol{A}_{12} \boldsymbol{A}_{22}^{-1} \boldsymbol{y}_{(2)}), \tag{14}$$

$$\boldsymbol{\gamma} \quad = \quad F_2(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \sigma_\varepsilon^2) = (\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})^{-1}(\boldsymbol{y}_{(2)} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{y}_{(1)}), \tag{15}$$

where

$$\begin{aligned}
\boldsymbol{y}_{(1)} &= \sum_{m=1}^{M} (\boldsymbol{u}_m \otimes \boldsymbol{\Phi}'_m)\boldsymbol{\Omega}_m^{-1}\boldsymbol{y}_m, \\
\boldsymbol{y}_{(2)} &= \sum_{m=1}^{M} \boldsymbol{V}'_m\boldsymbol{\Omega}_m^{-1}\boldsymbol{y}_m, \\
\boldsymbol{A}_{11} &= \sum_{m=1}^{M} (\boldsymbol{u}_m \otimes \boldsymbol{\Phi}'_m)\boldsymbol{\Omega}_m^{-1}(\boldsymbol{u}'_m \otimes \boldsymbol{\Phi}_m), \\
\boldsymbol{A}'_{21} = \boldsymbol{A}_{12} &= \sum_{m=1}^{M} (\boldsymbol{u}_m \otimes \boldsymbol{\Phi}'_m)\boldsymbol{\Omega}_m^{-1}\boldsymbol{V}_m, \\
\boldsymbol{A}_{22} &= \sum_{m=1}^{M} \boldsymbol{V}'_m\boldsymbol{\Omega}_m^{-1}\boldsymbol{V}_m.
\end{aligned}$$

Those give the formulae to update vec($\boldsymbol{B}$) and $\boldsymbol{\gamma}$ in (7).

Denote the jth element of $\boldsymbol{\theta}$ by θ_j. The gradients in terms of $(\theta_j, \sigma_\varepsilon^2, \sigma_i^2)$ are given by

$$\begin{aligned}
\frac{\partial l}{\partial \theta_j} = \sum_{m=1}^{M} \Bigg\{ & -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Omega}_m^{-1}\frac{\partial \boldsymbol{C}_m}{\partial \theta_j}\right) \\
& + \frac{1}{2}(\boldsymbol{y}_m - \boldsymbol{\Phi}_m\boldsymbol{B}\boldsymbol{u}_m - \boldsymbol{V}_m\boldsymbol{\gamma})'\boldsymbol{\Omega}_m^{-1}\frac{\partial \boldsymbol{C}_m}{\partial \theta_j}\boldsymbol{\Omega}_m^{-1}(\boldsymbol{y}_m - \boldsymbol{\Phi}_m\boldsymbol{B}\boldsymbol{u}_m - \boldsymbol{V}_m\boldsymbol{\gamma})\Bigg\},
\end{aligned} \tag{16}$$

$$\begin{aligned}
\frac{\partial l}{\partial \sigma_\varepsilon^2} = \sum_{m=1}^{M} \Bigg\{ & -\frac{1}{2}\mathrm{tr}(\boldsymbol{\Omega}_m^{-1}) \\
& + \frac{1}{2}(\boldsymbol{y}_m - \boldsymbol{\Phi}_m\boldsymbol{B}\boldsymbol{u}_m - \boldsymbol{V}_m\boldsymbol{\gamma})'\boldsymbol{\Omega}_m^{-1}\boldsymbol{\Omega}_m^{-1}(\boldsymbol{y}_m - \boldsymbol{\Phi}_m\boldsymbol{B}\boldsymbol{u}_m - \boldsymbol{V}_m\boldsymbol{\gamma})\Bigg\},
\end{aligned} \tag{17}$$

$$\begin{aligned}
\frac{\partial l}{\partial \sigma_i^2} = \sum_{m=1}^{M} \Bigg\{ & -\frac{1}{2}\mathrm{tr}(\boldsymbol{\Omega}_m^{-1}\boldsymbol{W}_{mi}\boldsymbol{W}'_{mi}) \\
& + \frac{1}{2}(\boldsymbol{y}_m - \boldsymbol{\Phi}_m\boldsymbol{B}\boldsymbol{u}_m - \boldsymbol{V}_m\boldsymbol{\gamma})'\boldsymbol{\Omega}_m^{-1}\boldsymbol{W}_{mi}\boldsymbol{W}'_{mi}\boldsymbol{\Omega}_m^{-1}(\boldsymbol{y}_m - \boldsymbol{\Phi}_m\boldsymbol{B}\boldsymbol{u}_m - \boldsymbol{V}_m\boldsymbol{\gamma})\Bigg\}, \\
& i = 1, \ldots, k, \text{ with } \boldsymbol{W}_{mi} \text{ being the } i\text{th column of } \boldsymbol{W}_m.
\end{aligned} \tag{18}$$

The aforementioned gradients are used such that the maximization procedure step (ii) in Section 2.2 can be implemented more efficiently.

APPENDIX B. Derivation of predictive mean and predictive variance

In Section 2.3, we assume that the covariance structure is the same at the different time points. From (9), when the mean is given, we assume that $(\tilde{\tau}_1, \ldots, \tilde{\tau}_N, \tilde{\tau}^*)$ has an $(N+1)$-dimensional normal distribution $N(\boldsymbol{0}, \tilde{\boldsymbol{C}}_{N+1})$ with

$$\tilde{\boldsymbol{C}}_{N+1} = \begin{pmatrix} \tilde{\boldsymbol{C}} & \tilde{\boldsymbol{C}}(z^*, z_{M+1}, \boldsymbol{w}^*, \boldsymbol{w}_{M+1}) \\ \tilde{\boldsymbol{C}}'(z^*, z_{M+1}, \boldsymbol{w}^*, \boldsymbol{w}_{M+1}) & \tilde{C}(z^*, z^*, \boldsymbol{w}^*, \boldsymbol{w}^*) \end{pmatrix},$$

which is an $(N+1) \times (N+1)$ matrix. Here,

$$\tilde{\boldsymbol{C}}(z^*, z_{M+1}, \boldsymbol{w}^*, \boldsymbol{w}_{M+1}) = [\tilde{C}(z^*, z_{M+1,1}, \boldsymbol{w}^*, \boldsymbol{w}_{M+1,1}), \ldots, \tilde{C}(z^*, z_{M+1,N}, \boldsymbol{w}^*, \boldsymbol{w}_{M+1,N})]'$$

is the covariance between y^* and $\boldsymbol{y}_{M+1} = (y_{M+1,1}, \ldots, y_{M+1,N})'$; $\tilde{\boldsymbol{C}}$ is the $N \times N$ covariance matrix of $\boldsymbol{y}_{M+1}$ or $\tilde{\boldsymbol{\tau}}_{M+1} = (\tilde{\tau}_1, \ldots, \tilde{\tau}_N)'$, which depends on $\boldsymbol{z}_{M+1}$ and $\boldsymbol{w}_{M+1}$. All of them are calculated by (8). The conditional mean and variance of $\tilde{\tau}^*$ given $\tilde{\boldsymbol{\tau}}_{M+1}$ and the mean $\boldsymbol{\mu}(t)$ are given by

$$\begin{aligned} \mathrm{E}(y^*|\mathcal{D}, \boldsymbol{\mu}) &= \mu_{M+1}(t^*) + \boldsymbol{H}'\boldsymbol{\tau}_{M+1} = \mu_{M+1}(t^*) + \boldsymbol{H}'(\boldsymbol{y}_{M+1} - \boldsymbol{\mu}_{M+1}(\boldsymbol{t})), \\ \hat{\sigma}^{*2} &= \mathrm{Var}(y^*|\mathcal{D}, \boldsymbol{\mu}) = \tilde{C}(z^*, z^*, \boldsymbol{w}^*, \boldsymbol{w}^*) - \boldsymbol{H}'\tilde{\boldsymbol{C}}\boldsymbol{H}, \end{aligned}$$

where $\boldsymbol{H}' = [\tilde{\boldsymbol{C}}(z^*, \boldsymbol{z}_{M+1}, \boldsymbol{w}^*, \boldsymbol{w}_{M+1})]'\tilde{\boldsymbol{C}}^{-1}$.

Thus, the predictive mean for y^* at new test point t^* is given by

$$\hat{y}^* = \mathrm{E}(y^*|\mathcal{D}) = \tilde{\mu}_{M+1}(t^*) + \boldsymbol{H}'(\boldsymbol{y}_{M+1} - \tilde{\boldsymbol{\mu}}_{M+1}(\boldsymbol{t})),$$

where $\tilde{\boldsymbol{\mu}}_{M+1}(\boldsymbol{t}) = (\tilde{\mu}_{M+1}(t_{M+1,1}), \ldots, \tilde{\mu}_{M+1}(t_{M+1,N}))'$ is the vector of means at data points $\boldsymbol{t} = (t_{M+1,1}, \ldots, t_{M+1,N})$ and is given by (10). This is (11) and is used as a prediction of y^*. The predictive variance is given by

$$\begin{aligned} \mathrm{Var}(y^*|\mathcal{D}) &= \mathrm{E}[\mathrm{Var}(y^*|\mathcal{D}, \boldsymbol{\mu})] + \mathrm{Var}[\mathrm{E}(y^*|\mathcal{D}, \boldsymbol{\mu})] \\ &= \hat{\sigma}^{*2} + \mathrm{Var}[\tilde{\mu}_{M+1}(t^*) + \boldsymbol{H}'(\boldsymbol{y}_{M+1} - \tilde{\boldsymbol{\mu}}_{M+1}(\boldsymbol{t}))|\mathcal{D}] \\ &= \hat{\sigma}^{*2} + \mathrm{Var}[\tilde{\mu}_{M+1}(t^*)|\mathcal{D}] + \boldsymbol{H}'\mathrm{Var}[\tilde{\boldsymbol{\mu}}_{M+1}(\boldsymbol{t})|\mathcal{D}]\boldsymbol{H} - 2\boldsymbol{H}'\mathrm{Cov}[\tilde{\mu}_{M+1}(t^*), \tilde{\boldsymbol{\mu}}_{M+1}(\boldsymbol{t})|\mathcal{D}]. \end{aligned}$$

Defining

$$\boldsymbol{\Lambda}_m = \begin{pmatrix} \boldsymbol{u}_m' \otimes \boldsymbol{\Phi}_m & \boldsymbol{V}_m \end{pmatrix} \text{ and } \boldsymbol{\alpha} = \begin{pmatrix} \mathrm{vec}(\boldsymbol{B}) \\ \boldsymbol{\gamma} \end{pmatrix}, \tag{19}$$

we can express the estimates (14) and (15) as

$$\hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{A}}^{-1} \begin{pmatrix} \boldsymbol{y}_{(1)} \\ \boldsymbol{y}_{(2)} \end{pmatrix} = \left(\sum_{m=1}^{M} \boldsymbol{\Lambda}_m' \hat{\boldsymbol{\Omega}}_m^{-1} \boldsymbol{\Lambda}_m \right)^{-1} \sum_{m=1}^{M} \boldsymbol{\Lambda}_m' \hat{\boldsymbol{\Omega}}_m^{-1} \boldsymbol{y}_m,$$

where

$$\hat{\boldsymbol{A}} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix} = \sum_{m=1}^{M} \boldsymbol{\Lambda}_m' \hat{\boldsymbol{\Omega}}_m^{-1} \boldsymbol{\Lambda}_m$$

with A_{ij} given in Appendix A and with the parameters replaced by their estimates. The variance of $\hat{\boldsymbol{\alpha}}$ is

$$\begin{aligned} \mathrm{Var}(\hat{\boldsymbol{\alpha}}) &= \left(\sum_{m=1}^{M} \boldsymbol{\Lambda}_m' \hat{\boldsymbol{\Omega}}_m^{-1} \boldsymbol{\Lambda}_m \right)^{-1} \sum_{m=1}^{M} \boldsymbol{\Lambda}_m' \hat{\boldsymbol{\Omega}}_m^{-1} \mathrm{Var}(\boldsymbol{y}_m) \hat{\boldsymbol{\Omega}}_m^{-1} \boldsymbol{\Lambda}_m \left(\sum_{m=1}^{M} \boldsymbol{\Lambda}_m' \hat{\boldsymbol{\Omega}}_m^{-1} \boldsymbol{\Lambda}_m \right)^{-1} \\ &= \left(\sum_{m=1}^{M} \boldsymbol{\Lambda}_m' \hat{\boldsymbol{\Omega}}_m^{-1} \boldsymbol{\Lambda}_m \right)^{-1} = \hat{\boldsymbol{A}}^{-1}, \end{aligned}$$

where $\hat{\boldsymbol{\Omega}}_m$ means $\boldsymbol{\Omega}_m$ with replacing the parameters by their estimates; see, for example, Section 3 in [27]. As noted in Section 3.2 of their paper, these formulae do not take into account the uncertainty of the estimation of the parameters involved in $\boldsymbol{\Omega}_m$.

Therefore, we have

$$\begin{aligned} \mathrm{Var}[\tilde{\mu}_{M+1}(t)|\mathcal{D}] &= \mathrm{Var}[(\boldsymbol{u}_{M+1}'\hat{\boldsymbol{B}}'\boldsymbol{\Phi}(t) + \boldsymbol{v}_{M+1}'(t)\hat{\boldsymbol{\gamma}})|\mathcal{D}] \\ &= \mathrm{Var}[\boldsymbol{\Lambda}_{M+1}\hat{\boldsymbol{\alpha}}|\mathcal{D}] \\ &= \boldsymbol{\Lambda}_{M+1} \left(\sum_{m=1}^{M} \boldsymbol{\Lambda}_m' \hat{\boldsymbol{\Omega}}_m^{-1} \boldsymbol{\Lambda}_m \right)^{-1} \boldsymbol{\Lambda}_{M+1}' = \boldsymbol{\Lambda}_{M+1} \hat{\boldsymbol{A}}^{-1} \boldsymbol{\Lambda}_{M+1}', \end{aligned}$$

where each column of $\boldsymbol{\Lambda}_{M+1}$ is defined as in (19) but evaluated at $t = t_{M+1,i}$ for $i = 1, \ldots, N$. And $\mathrm{Var}[\tilde{\mu}_{M+1}(t^*)|\mathcal{D}]$ and $\mathrm{Cov}[\tilde{\mu}_{M+1}(t), \tilde{\mu}_{M+1}(t^*)|\mathcal{D}]$ can be calculated similarly.

Combining the aforementioned results, we have

$$\mathrm{Var}(y^*|\mathcal{D}) = \hat{\sigma}^{*2} + (\mathbf{\Lambda}^*_{M+1} - \boldsymbol{H}'\mathbf{\Lambda}_{M+1})'\hat{\boldsymbol{A}}^{-1}(\mathbf{\Lambda}^*_{M+1} - \boldsymbol{H}'\mathbf{\Lambda}_{M+1}), \tag{20}$$

where $\mathbf{\Lambda}^*_{M+1}$ is defined in (19) but evaluated at $t = t^*$. This is used to calculate the predictive variance.

Acknowledgements

We would like to thank the associate editor and two referees for their helpful comments and suggestions that have greatly improved the manuscript.

References

1. Ramsay JO, Silverman BW. *Functional Data Analysis*, 2nd edition. Springer: New York, 2005.
2. Pinheiro JC, Bates DM. *Mixed-Effects Models in S and S-PLUS*. Springer: New York, 2000.
3. Ruppert D, Wand MP, Carroll RJ. *Semiparametric Regression*. Cambridge University Press: Cambridge, 2003.
4. Hastie T, Tibshirani RJ. *Generalized Additive Model*. Chapman & Hall: London, 1990.
5. Fan J, Zhang W. Statistical estimation in varying coefficient models. *Annals of Statistics* 1999; **27**:1491–1518.
6. Rasmussen CE, Williams CKI. *Gaussian Processes for Machine Learning*. The MIT Press: Cambridge, MA, 2006.
7. Shi JQ, Choi T. *Gaussian Process Regression Analysis for Functional Data*. Chapman & Hall/CRC: London, 2011.
8. Gönen M. Planning a dose–response study with subject-specific doses. *Statistics in Medicine* 2005; **24**:2613–2623.
9. Shi JQ, Murray-Smith R, Titterington DM. Hierarchical Gaussian process mixtures for regression. *Statistics and Computing* 2005; **15**:31–41.
10. Shi JQ, Wang B, Murray-Smith R, Titterington DM. Gaussian process functional regression modelling for batch data. *Biometrics* 2007; **63**:714–723.
11. Faraway J. Regression analysis for a functional response. *Technometrics* 1997; **39**:254–261.
12. Carlin BP, Louis TA. *Bayesian and Empirical Bayes Methods for Data Analysis*. Chapman & Hall/CRC: London, 1996.
13. Choi T. Posterior consistency in nonparametric regression problems under Gaussian process priors. *PhD Thesis*, Carnegie Mellon University, 2005.
14. Seeger MW, Kakade SM, Foster DP. Information consistency of nonparametric Gaussian process methods. *IEEE Transactions on Information Theory* 2008; **54**:2376–2382.
15. Will EJ, Richardson D, Tolman C, Bartlett C. Development and exploitation of a clinical decision support system for the management of renal anaemia. *Nephrology, Dialysis and Transplantation* 2007; **22**. [suppl 4]: iv31–iv36.
16. Tolman C, Richardson D, Bartlett C, Will EJ. Structured conversion from thrice weekly to weekly erythropoietic regimens using a computerized decision-support system: a randomized clinical study. *Journal of the American Society of Nephrology* 2005; **16**:1463–1470.
17. Volkova N, Arab L. Evidence-based systematic literature review of hemoglobin/hematocrit and all-cause mortality in dialysis patients. *American Journal of Kidney Disease* 2006; **47**:24–36.
18. West RM, Harris K, Gilthorpe MS, Tolman C, Will EJ. A description of patient sensitivity to epoetins and control of renal anaemia – functional data analysis applied to a randomized controlled clinical trial in hemodialysis patients. *Journal of the American Society of Nephrology* 2007; **18**:2371–2376.
19. West M, Harrison J. *Bayesian Forecasting and Dynamic Models*, 2nd edition. Springer-Verlag: New York, USA, 1997.
20. Schwarz G. Estimating the dimension of a model. *Annals of Statistics* 1978; **6**:461–464.
21. Nott DJ, Li J. A sign based loss approach to model selection in nonparametric regression. *Statistics and Computing* 2010; **20**:485–498.
22. Murphy S. Optimal dynamic treatment regimes (with discussion). *Journal of the Royal Statistical Society* Series B 2003; **65**:331–366.
23. Robins JM. Optimal structured nested models for optimal sequential decisions. In *Proceedings of the Second Seattle Symposium on Biostatistics*, Lin DY, Heagerty PJ (eds). Springer: New York, 2004; 189–326.
24. Henderson R, Ansell P, Alshibani D. Regret-regression for optimal dynamic treatment regimes. *Biometrics* 2010; **66**:1192–1201.
25. Berger MPF, Wong WK. *Applied Optimal Designs*. John Wiley and Sons: New York, 2005.
26. Shi JQ, Wang B. Curve prediction and clustering with mixtures of Gaussian process functional regression models. *Statistics and Computing* 2008; **18**:267–283.
27. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982; **38**:963–974.