Dose–response Curve Prediction Using GPFR Models with Applications to the Control of Renal Anaemia

J. Q. Shi[1], B. Wang[2], E. J. Will[3], and R. M. West[4]

June 12, 2008

[1]Statistics Department, Newcastle University, Newcastle, NE1 7RU, UK.

[2]Department of Mathematics, University of York, York, YO10 5DD, UK

[3]Department of Renal Medicine, St James's University Hospital, Beckett Street, Leeds, LS9 7TF, UK

[4]Centre for Epidemiology & Biostatistics, University of Leeds, Worsley Building, Leeds, LS2 9JT, UK

Summary

Dose–response curves describe changes in subjects responses for differing levels of the dose of a drug or agent, and have a wide application in many areas particularly in medicine, pharmacology and toxicology. Based on a nonparametric Gaussian process regression model, we proposed a functional regression model to study the dose–response

relationship. This model enables modelling of a nonlinear functional regression relationship between a functional response curve and a set of high-dimensional functional covariates. Mean structure and covariance structure are modelled simultaneously, combining the information borrowed from other subjects and the information collected from each individual subject. The methodology has been demonstrated for the management of renal anaemia. The individual dose–response curve can be obtained and can be improved when more information is gathered from each individual patient over time, which enables a patient-specific treatment regime.

Keywords: Control of renal anaemia, Gaussian process functional regression model, Individual dose–response curve, Nonparametric approach, Patient-specific treatment regime.

1 Introduction

The major component of dose–response studies is to characterise the change in subject response for differing levels of the dose of a drug or agent. This has wide application in medicine, epidemiology, pharmacology and other areas. We will first use an example to illustrate the problems we discuss in this paper, as the stimulus to the theoretical development of random-effects Gaussian Process Functional Regression.

In the UK, for a hospital with a catchment population of around one million, approximately 100 patients each year will need to start replacement treatment for renal failure. Patients with reduced kidney function not only require dialysis to remove waste products from their blood, but they 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 of some degree. This can be very effectively treated by injection of exogenous epoetin. Injections are given subcutaneous or intravenously with either a synthetic EPO for example Erythropoetin Beta (EB) or a modified epoetin such as Darbepoetin Alpha (DA).

The dose of epoetin to be given to each patient is determined by monitoring the haemoglobin (Hb) concentration from a blood sample taken typically every two or four weeks. Over the last decade in Leeds, blood samples are taken on a monthly basis and the epoetin dose calculated using a strict management algorithms (Will *et al*, 2007). This represents a clinical decision support system (CDSS). Patients need to have their Hb levels controlled within relatively narrow limits (Volkova and Arab, 2006). If Hb levels are too low then patients become symptomatic of anaemia and if too high then there may be pro-thrombotic risks to their dialysis treatment and vascular tree. The primary therapeutic concern is how to maintain the Hb level by giving a suitable level of dose of epoetin for each patient.

In this example, the response is Hb level for a patient, shown as curves for 151 patients in Figure 1(a). One of covariates is the dose level of epoetin, which is also shown as a curve for each of the patients in Figure 1(b). 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 renal data presented here, measurements are taken at different time points, which cannot be regarded as independent observations since measurements on the same subject will be strongly correlated. Fitting smooth curves through the measurement series provides curves suited to functional data

analysis (see Ramsay and Silverman, 2005). In this paper, we will focus on the study of a functional regression model between a response curve and a set of covariate curves. In practice, for our renal data and many other examples, we often have little information about the real physical relationship between the response curves and covariate curves, it is therefore not realistic to use a parametric model or a linear model such as the linear functional regression models discussed in Ramsay and Silverman (2005). The approach in this paper is to use a Gaussian Process Functional Regression (GPFR) model (Shi, *et al.*, 2005 and 2007), which is regarded as a nonlinear nonparametric model and will be extended to include random effects.

As argued in Shi, *et al.* (2007), one major advantage of the GPFR model is not only to reveal the model structure from the data collected from all subjects, but also to predict for each individual, based on the common model structure and the data collected from that particular individual. This is particular useful in dose–response studies such that we can construct dose–response curve for each individual patient, while the most current studies can only provide a common dose–response curve (see e.g. the discussion in Gönen, 2005). Our method therefore enables the planning of a patient-specific treatment regime and that treatment regime can be improved over time as more data is collected for each of the individual patients. To the best of our knowledge this is the first reporting of such methodology in the literature.

The paper is organised as follows. Section 2 proposes a mixed-effects GPFR model with random effects between different subjects. The related theory and methodology for calculating estimates and prediction are also presented in this section. In Section 3, we demonstrate the methodology for dose–response study with the Leeds renal data.

Discussion and further developments are given in Section 4.

2 Methodology

2.1 GPFR models for dose–response curves

Let $y_m(t)$ be the response curve for m-th subject for $m = 1, \ldots, M$, and $\boldsymbol{x}_m(t)$ and $\boldsymbol{z}_m(t)$ be two sets of functional covariates which are used in mean and covariance models respectively. Here t is time or another one-dimensional covariate. In the Leeds renal example (see the detailed discussion in Section 3), the response curve is the measurement of hemoglobin (Hb) concentration for renal anaemia patients, which is a functional response variable changing over time t. The related functional covariates are applied dosage of the epoetin agent (either EB or DA), and other covariates such as iron dose-level and the Hb level measured in the previous month, which are all changed along time. The aim of the research is to explore the dose–response relationship and to use the input covariates to predict the response curve Hb and then maintain Hb at the desired level. Observations of Hb and all the covariates are taken once a month over 12 months. The data is shown in Figure 1. For such data, the measurements on the same subject are strongly correlated. So we will define a functional regression model to model both mean and covariance structures. The model is defined as follows

$$y_m(t) = \mu_m(\boldsymbol{u}_m, \boldsymbol{x}_m(t)) + \tau_m(\boldsymbol{z}_m(t)) + \varepsilon_m(t),\ \varepsilon_m(t) \sim N(0, \sigma_\varepsilon^2), \tag{1}$$

where $\varepsilon_m(t)$ are random errors and are independent at different times. The vector $\boldsymbol{u}_m$ is a set of curve-based scalar covariates, giving information for the m-th curve or subject,

for example patient's age, sex etc. So, the mean structure $\mu_m(\boldsymbol{u}_m, \boldsymbol{x}_m(t))$ may depend on scalar covariates $\boldsymbol{u}_m$ and l-dimensional functional inputs $\boldsymbol{x}_m(t)$. It is usually difficult to use and justify a very general functional mean structure, some special models are therefore defined and discussed in this paper. The second term $\tau_m(\boldsymbol{z}_m(t))$ is used to model covariance structure for correlated data. A nonparametric Gaussian Process (GP) regression model is used to model the covariance structure (see Rasmussen and Williams, 2006). The GP model is defined as follows:

$$\tau_m(\boldsymbol{z}_m, t) \sim N(0, C(\cdot, \cdot; \boldsymbol{\theta})), \tag{2}$$

where $C(\cdot, \cdot; \boldsymbol{\theta})$ is a selected kernel covariance function for m-th curve, depending on a Q-dimensional functional covariates $\boldsymbol{z}_m$. The GP regression model is a nonparametric model and can be used for fitting any shape of curves in terms of a large dimension functional covariates $\boldsymbol{z}_m$, while most of conventional nonparametric methods are limited to a small dimension of covariates (usually $Q < 3$). Used Larhunen-Loeve expansion to the GP (Wahba, 1990), the model is expressed as

$$\tau_m(\boldsymbol{z}_m, t) = \sum_{j=1}^{\infty} \xi_j \phi_j(\boldsymbol{z}_m(t)), \tag{3}$$

where ξ_j's ($j = 1, 2, \ldots, $) are independent Gaussian random variables with zero mean and variance λ_j which is the j-th eigen-value of kernel covariance function $C(\cdot, \cdot)$, and ϕ_j is the related eigen-function. Therefore, selecting a kernel covariance function in (2) is equivalent to selecting a set of basis functions in a conventional nonparametric model. Rasmussen and Williams (2006) discussed several different types of kernel covariance functions.

We now discuss the mean model in (1). The simplest model is to define a constant

mean

$$\mu_m(\boldsymbol{u}_m, \boldsymbol{x}_m(t)) = \mu_m. \tag{4}$$

We can usually replace μ_m by the sample mean of the observations collected from m-th curve, or the sample mean of the observations collected from all the observations if we assume $\mu_m = \mu$. As pointed out in Shi *et al.* (2007), this model cannot cope with the curve-based (or subject-based) information and the heterogeneity caused whereas a Gaussian Process Functional Regression (GPFR) model enables this extension. The mean structure is modelled by

$$\mu_m(\boldsymbol{u}_m, \boldsymbol{x}_m(t)) = \boldsymbol{u}_m{}'\boldsymbol{\beta}(t). \tag{5}$$

Thus, a GPFR model is defined by (1) with covariance structure (2) and the above mean structure (5). This mean model is a linear functional regression model with scalar covariates as discussed in Ramsay and Silverman (2005). The covariance structure is modelled by a nonparametric Gaussian Process Regression Model. The relationship between the response curve and the functional covariates are mainly modelled by the covariance model nonparametrically (see the detailed discussion in Shi *et al.*, 2007). For the problem where there is little or no information for the parametric relationship between functional response variable and functional covariates, this model performs very well.

2.2 Mixed-effects GPFR models

For the renal data, the overall sample linear correlation coefficient between Hb and dose level is about 0.4, meaning there may be a linear relationship between the response curve and the functional covariate. Motivated by this example, we propose a mixed-effects

GPFR model. The GPFR model is still defined by (1) with covariance structure (2), but with the following mixed-effects mean model.

$$\mu_m(\boldsymbol{u}_m, \boldsymbol{x}_m(t)) = \boldsymbol{u}_m{}'\boldsymbol{\beta}(t) + \boldsymbol{v}_m{}'(t)\boldsymbol{\gamma} + \boldsymbol{w}_m{}'(t)\boldsymbol{b}_m, \tag{6}$$

$$\boldsymbol{b}_m \sim N(0, \boldsymbol{\Sigma}), \tag{7}$$

where $\boldsymbol{u}_m$ is a p-dimensional non-functional covariate, $\boldsymbol{v}_m(t)$ and $\boldsymbol{w}_m(t)$ are r-dimensional and k-dimensional subsets of functional covariates $\boldsymbol{x}_m(t)$ respectively, and $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \ldots, \sigma_k^2)$. The random effects term associated with covariates $\boldsymbol{w}_m(t)$ is used to model the heterogeneity for different subjects.

We propose mixed-effects GPFR models (1) with (2), (6) and (7) here. From (3), the model can be expressed as

$$y_m(t) = \boldsymbol{u}_m{}'\boldsymbol{\beta}(t) + \boldsymbol{v}_m{}'(t)\boldsymbol{\gamma} + \boldsymbol{w}_m{}'(t)\boldsymbol{b}_m + \sum_{j=1}^{\infty} \xi_j \phi_j(\boldsymbol{z}_m(t)) + \varepsilon_m(t). \tag{8}$$

In the right hand side of the above model, the first item stands for the common mean structure which can be learnt from the data collected from all subjects. The rest of items stands for patient-specific part, where the second and third items are linear parametric model. The fourth item is linked to a Gaussian process, which is a nonparametric model and is the main part to model the covariance structure and model the nonlinear relationship between $y_m(t)$ and $\boldsymbol{z}_m(t)$. In theory, we can merge the second and the third items into the fourth item (with a new kernel covariance function), we listed them separately here because this may offer some information of interpretation in applications. Model (8) is a nonparametric model which can be used to fit response curves with any shape in terms of a set of covariate curves although the performance may depend on the choice of kernel covariance function $C(\cdot, \cdot)$.

Suppose that N_m observations are obtained for the m-th subject. All the data collected for the m-th subject are

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

All the observed data is denoted as $\mathcal{D}$. In (6) and (7), $\boldsymbol{v}_m$ and $\boldsymbol{w}_m$ are subsets of $\boldsymbol{x}_m$.

Since the mean structure involves a functional coefficient $\boldsymbol{\beta}(t)$, it is therefore not straightforward to estimate the mean functions. In this paper we propose to use B-spline approximation. Let $\boldsymbol{\Phi}(t) = (\boldsymbol{\Phi}_1(t), \ldots, \boldsymbol{\Phi}_D(t))'$ be a set of B-spline basis functions, the coefficient function $\boldsymbol{\beta}(t)$ can be approximated by $\boldsymbol{B}'\boldsymbol{\Phi}(t)$, and the mean function is represented as

$$\mu_m(\boldsymbol{u}_m, \boldsymbol{x}_m(t)) = \boldsymbol{u}_m{}'\boldsymbol{B}'\boldsymbol{\Phi}(t) + \boldsymbol{v}_m{}'(t)\boldsymbol{\gamma} + \boldsymbol{w}_m{}'(t)\boldsymbol{b}_m. \tag{9}$$

Let $\boldsymbol{y}_m$ be the vector of $\{y_{mi}, i = 1, \ldots, N_m\}$, $\boldsymbol{t}_m = \{t_{mi}, i = 1, \ldots, N_m\}$, $\boldsymbol{V}_m$ be the $N_m \times r$ matrix with the i-th row $\boldsymbol{v}_m{}'(t_{mi})$ and $\boldsymbol{W}_m$ be an $N_m \times k$ matrix with the i-th row $\boldsymbol{w}_m{}'(t_{mi})$, the model for N_m observations for m-th subject is therefore:

$$\boldsymbol{y}_m = \boldsymbol{\Phi}_m\boldsymbol{B}\boldsymbol{u}_m + \boldsymbol{V}_m\boldsymbol{\gamma} + \boldsymbol{W}_m\boldsymbol{b}_m + \boldsymbol{\tau}_m + \boldsymbol{\varepsilon}_m, \tag{10}$$

where $\boldsymbol{\Phi}_m$ is an $N_m \times D$ matrix with (i,d)-th element $\boldsymbol{\Phi}_d(t_{mi})$, $\boldsymbol{\varepsilon}_m \sim N(\boldsymbol{0}, \sigma_\varepsilon^2\boldsymbol{I})$, $\boldsymbol{\tau}_m \sim N(\boldsymbol{0}, \boldsymbol{C}(\boldsymbol{\theta}))$, and $\boldsymbol{C}(\boldsymbol{\theta})$ is an $N_m \times N_m$ covariance matrix with (i,j)-th element $C(\boldsymbol{z}_{mi}, \boldsymbol{z}_{mj}; \boldsymbol{\theta})$ given by (2). The unknown parameters involved in the mixed-effects GPFR models include B-spline coefficient $\boldsymbol{B}$, fixed effect coefficient $\boldsymbol{\gamma}$, random effect covariance matrix $\boldsymbol{\Sigma}$, the parameters $\boldsymbol{\theta}$ involved in covariance structure and variance σ_ε^2 of measurement error. These parameters are denoted collectively as

$$\boldsymbol{\Theta} = (\boldsymbol{B}, \boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \sigma_\varepsilon^2).$$

2.3 Estimation

We use maximum likelihood method to calculate the estimates of $\boldsymbol{\Theta}$. The log-likelihood of $\boldsymbol{\Theta} = (\boldsymbol{B}, \boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \sigma_\varepsilon^2)$ is

$$\begin{aligned} & L(\boldsymbol{B}, \boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \sigma_\varepsilon^2) \\ = & \sum_{m=1}^{M} \Big\{ -\frac{1}{2} N_m \log(2\pi) - \frac{1}{2} \log |\boldsymbol{\Omega}_m| \\ & -\frac{1}{2} (\boldsymbol{y}_m - \boldsymbol{\Phi}_m \boldsymbol{B} \boldsymbol{u}_m - \boldsymbol{V}_m \boldsymbol{\gamma})' \boldsymbol{\Omega}_m^{-1} (\boldsymbol{y}_m - \boldsymbol{\Phi}_m \boldsymbol{B} \boldsymbol{u}_m - \boldsymbol{V}_m \boldsymbol{\gamma}) \Big\}, \end{aligned} \tag{11}$$

where $\boldsymbol{\Omega}_m = \boldsymbol{W}_m \boldsymbol{\Sigma} \boldsymbol{W}_m{}' + \boldsymbol{C}_m + \sigma_\varepsilon^2 \boldsymbol{I}$ and $\boldsymbol{C}_m = (C_m^{ij}),\ i, j = 1, \ldots, N_m$, is the covariance matrix whose element C_m^{ij} is given in (2).

From Appendix I, the first derivatives of $L(\boldsymbol{B}, \boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \sigma_\varepsilon^2)$ in terms of vec$(\boldsymbol{B})$ and $\boldsymbol{\gamma}$ are given by (18) and (19), where vec$(\boldsymbol{B})$ denotes the stacked columns of $\boldsymbol{B}$. Letting $\partial L / \partial \text{vec}(\boldsymbol{B}) = 0$ 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

$$\text{vec}(\boldsymbol{B}) = F_1(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \sigma_\varepsilon^2), \quad \boldsymbol{\gamma} = F_2(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \sigma_\varepsilon^2). \tag{12}$$

The derivation and the formulae of F_1 and F_2 are given in Appendix I. Thus, we used the following iterative procedure. Each of iteration includes two steps:

(i) Update $\boldsymbol{B}$ and $\boldsymbol{\gamma}$ by (12) given the current values of $\boldsymbol{\theta}$, $\boldsymbol{\Sigma}$, σ_ε^2;

(ii) Update $\boldsymbol{\theta}$, $\boldsymbol{\Sigma}$, σ_ε^2 by maximising L in (11) given $\boldsymbol{B}$ and $\boldsymbol{\gamma}$.

To speed up convergence, we usually repeat the above two steps several times within each iteration. There are no explicit expression in step (ii), but we can use the gradient to speed up the maximisation. The expressions of the gradient are given in (22) to (24) in

Appendix I. We denote the maximum likelihood estimates by $\hat{\boldsymbol{\Theta}}$. The estimate of the functional coefficient is given by $\hat{\boldsymbol{\beta}}(t) = \hat{\boldsymbol{B}}'\boldsymbol{\Phi}(t)$.

2.4 Prediction

Suppose that we have already observed some training data for a subject (e.g. a new patient), the $(M+1)$-th subject say, and want to predict the output y^* for a new set of inputs $(t^*, \boldsymbol{x}^*, \boldsymbol{z}^*)$ with $\boldsymbol{x}^* = \boldsymbol{x}(t^*)$ and $\boldsymbol{z}^* = \boldsymbol{z}(t^*)$ 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), or could be at a future time. The latter is of most interest, meaning we want to predict the response curve for future times (this is so called *extrapolation*). Thus, in addition to the training data observed from the first M subjects, we assume that N observations are obtained for this $(M+1)$-th new subject, providing data

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

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

In fact, from (11), the random effect part, random measurement error, and Gaussian process error can be integrated into a new Gaussian process, denoted by $\tilde{\tau}(\boldsymbol{z}_m, \boldsymbol{w}_m, t)$, with zero mean and covariance structure

$$\tilde{C}(\boldsymbol{z}_i, \boldsymbol{z}_j, \boldsymbol{w}_i, \boldsymbol{w}_j; \boldsymbol{\theta}, \boldsymbol{\Sigma}, \sigma_\varepsilon^2) = C(\boldsymbol{z}_i, \boldsymbol{z}_j; \boldsymbol{\theta}) + \sum_{q=1}^{k} \sigma_q^2 w_{iq} w_{jq} + \sigma_\varepsilon^2 \delta_{ij}, \tag{13}$$

where δ_{ij} is the Kronecker delta and $C(\boldsymbol{z}_i, \boldsymbol{z}_j; \boldsymbol{\theta})$ is the kernel covariance function in (2). The models (1), (2), (6) and (7) can then be written as

$$y_m(t) = \tilde{\mu}_m(t) + \tilde{\tau}(\boldsymbol{z}_m, \boldsymbol{w}_m, t), \quad \text{with} \quad \tilde{\mu}_m(t) = \boldsymbol{u}_m{}'\boldsymbol{\beta}(t) + \boldsymbol{v}_m{}'(t)\boldsymbol{\gamma}. \tag{14}$$

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

$$\tilde{\mu}_{M+1}(t) = \boldsymbol{u}_{M+1}{}'\hat{\boldsymbol{\beta}}(t) + \boldsymbol{v}'_{M+1}\hat{\boldsymbol{\gamma}}. \tag{15}$$

Now, we assume that the covariance structure at data points $t_{M+1,i}$ for $i = 1, \ldots, N$ and at new data point t^* are the same, the prediction at t^* is given by

$$\hat{y}^* = \mathrm{E}(y^*|\mathcal{D}) = \tilde{\mu}_{M+1}(t^*) + \boldsymbol{H}'(\boldsymbol{y}_{M+1} - \tilde{\boldsymbol{\mu}}_{M+1}(\boldsymbol{t})). \tag{16}$$

The proof is given in Appendix II. The predictive variance is given by (26). Here we assume that the correlation between the response and input variables are the same at different times (including future times). This assumption is reasonable in practice.

The predictive mean (16) includes two parts. The first item comes from the common mean model, which is learnt based on the information collected from all the subjects. If we have not obtained any data for the new patient, we use it as a prediction. The second item is estimated from the data collected from the particular subject, i.e. it is estimated based on the patient-specific information. Thus, the predictive mean (16) can be used to construct individual dose–response curve and to plan patient-specific treatment regime as we will discuss in the next section.

2.5 Consistency

We used an iterated maximum likelihood approach to estimate the unknown parameters in the previous sections. All the data collected from M subjects are used on estimating parameters. This leads to consistent estimates under some regularity conditions (see Peters and Walker, 1978).

The major problem is the consistency of the prediction. The common mean structure is estimated from the data collected from all M subjects. It is a consistent estimate of the true mean structure under some regularity conditions (see the details in Ramsay and Silverman, 2005). If we have not observed any data for a new subject, we can only used the common mean as a prediction of $y^*(t)$, which is obviously not a consistent estimate of $y^*(t)$. When we have obtained some observations for a new subject, the prediction $\hat{y}^*(t)$ given in (16) is a posterior mean based on a Gaussian process prior. Choi (2005) proved the posterior consistency of $\hat{y}^*(t)$ under some regularity conditions if the selected kernel covariance function is stationary and the sample size of the observations collected in the new subject is sufficiently large. A more useful result is given in Seeger *et al.* (2008). They proved the information consistency under the Kullback–Leibler divergence between the true function $y^*(t)$ and the prediction $\hat{y}^*(t)$, and also provided the error bound. However, it is still an interesting problem to explore what the error bound is when the sample size of the observations obtained from each individual is not sufficiently large and the kernel covariance structure is not stationary.

The selection of input covariates in the model is an important issue for prediction. In this paper we use Bayes factors. Since its exact value is difficult to calculate, we utilise its approximate form of Bayesian Information Criterion (BIC) (Schwarz, 1978). Recalling that $L(\cdot)$ is the log-likelihood function as defined in (11) and letting $\hat{\boldsymbol{\Theta}}$ be the maximum likelihood estimate, the BIC value is given by

$$\text{BIC} = -2L(\hat{\boldsymbol{\Theta}}) + G\log(N),$$

where G is the total number of parameters and $N = N_1 + \cdots + N_M$ is the sample size.

3 Dose-response study with applications to the management of renal anaemia

We will use our Leeds renal data to illustrate how to use the models discussed in the above section to conduct a dose–response study.

The main purpose of the experiment is to assess the control of haemoglobin (Hb) levels in patients each dosed with one of two epoetin agents. This however is not entirely straightforward. Contributions to the variation in Hb levels also come from further deterioration in kidney function, iron status, and inter-current complications such as infection, surgery, and hemorrhage. Iron status is usually assured with sufficient doses of iron but there is little control over other aspects.

A recent study was undertaken in Leeds (Tolman *et al.*, 2005) in order to compare the effectiveness of the weekly administration of two epoetin agents: EB and DA. Consented patients were randomly allocated to treatment with one of the two agents, then monitored monthly for their Hb levels. A total of 151 patients were followed for the original study period of 9 months and a further 3 months 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 CDSS.

Taking a functional data analysis approach to the data from this trial, West *et al.* (2007) 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, the primary

aim being to determine if equivalent control could be achieved by the two agents given the same clinical environment: the same frequency of epoetin management and administration (monthly review and weekly injections), and using the same CDSS.

In this section, we will first calculate the prediction and assess the performance of the models, and then demonstrate how to obtain individual dose–response curves.

3.1 Prediction of Hb

As a consequence of randomisation in the Tolman randomised controlled trial, 77 patients received EB and 74 DA. It is possible, and of considerable interest to explore, that control of patients under these two agents differs. Therefore, the mixed-effects GPFR model will be fitted independently within the two arms of the trial.

We use model (1) with mean model (6) and covariance structure model (2). As an example, we use the following kernel covariance function

$$C(\boldsymbol{z}_i, \boldsymbol{z}_j; \boldsymbol{\theta}) = v_0 \exp\left\{-\frac{1}{2}\sum_{q=1}^{Q} w_q (z_{iq} - z_{jq})^2\right\} + a_0, \tag{17}$$

We will first consider the following three mean models.

M1: with constant mean $\mu_m(\boldsymbol{x}_m, t) = 11.8$;

M2: with simple mean: $\mu_m(\boldsymbol{x}_m, t) = \mu(t)$, i.e., no covariates are included in the mean structure; and

M3: with mixed effect mean: $\mu_m(\boldsymbol{x}_m, t) = \boldsymbol{v}_m{}'(t)\boldsymbol{\gamma} + \boldsymbol{w}_m{}'(t)\boldsymbol{b}_m$, where the fixed effect part involves the covariates $\boldsymbol{v}_m(t) = \{1, t, d_m(t-2), d_m^2(t-2)\}$ and the random effect part involves the covariates $\boldsymbol{w}_m(t) = \{1, d_m(t-2)\}$. Here, t is the time and $d_m(t-2)$ is the dose level for subject m at time $t-2$. In this project, we found that the Hb level $y_m(t)$

has the highest correlation with dose-level $d_m(t-2)$ (i.e. the dosage of the drug taken between 30–60 days ago), although it is also related to $d_m(t-1)$ and $d_m(t-3)$. So for clarity of exposition we use $d_m(t-2)$ in this illustrative example. A more comprehensive study is being undertaken and will be reported at a later date.

In the related covariance structure models, we use all the covariates given in *M3* and the values of Hb observed up to date.

We first used the data collected from all patients to train the model (i.e. estimate all the unknown parameters involved in the models), and then considered the problem of prediction for each of the individual patients. We used the data up to current month, and assumed a level of dosage which was used from the date to the next 30 days to predict the value of Hb level in 60 days. We started prediction from Month 8 for each subject, i.e., we use all the data collected from the patient in the first six months, including Hb level, to predict their Hb level in Month 8. For each subject, we calculated the predictions from Month 8 to Month 12. The root of mean squared errors (*rmse*) between the prediction and the real observations were used to judge the performance of the model.

Table 1 gives the values of *rmse* for each of the models under the two epoetin agents. It can be seen that the performance of the mixed effects model M3 is preferred: it has the lowest value of *rmse* for both arms of the trial. Figure 2 shows the predictions and the real observations for two typical patients.

Having established that the random-effects model performs best, the issue of equality of prediction under the two agents can be tackled. A Mann–Whitney test applied to the observed *rmse*'s indicates no statistically significant difference ($p = 0.1301$).

The covariance structure involved in model (1) plays an important role in prediction.

Table 1: The values of *rmse* for the different models

Model	Epoetin agent	
	DA	EB
M1	1.0500	0.7993
M2	0.8773	0.7983
M2(mean only)	1.1452	1.0939
M3	0.7412	0.6982
M3(mean only)	0.8427	0.7521
M4	0.7573	0.6993

If we ignore the covariance structure, and just use the mean structures defined in M2 and M3, the related average values of *rmse* are also given in Table 1, which are much larger than the *rmse* obtained from Model M2 and M3 respectively. For M3 although the related mean model includes a quadratic term, the GPFR model performs much better than the model with mean structure only. This is the evidence that GPFR can model the unknown nonlinear functional regression relationship, and it therefore improves the model fit and the prediction. This finding is consistent with the results reported in Shi *et al.* (2007). The mean model synthesises the information from different subjects at each time point, but the covariance model tunes the prediction based on the information collected for each individual. Therefore, Models (1) give very accurate values of prediction.

We have also considered some other forms of mixed-effects GPFR model, it seems that the performance of M3 cannot be improved further. For example, we consider the model with the following mean structure

M4: with the mean model in M3 but with $\boldsymbol{v}_m(t) = \{1, t, t^2, d_m(t-2), d_m^2(t-2)\}$ and $\boldsymbol{w}_m(t) = \{1, d_m(t-2), d_m^2(t-2)\}$.

Table 2: Estimates of coefficients for model M4

Fixed-effect part			Random-effect part		
Covariate	Estimate	Std error	Covariate	Estimate	Std error
constant	11.1246	0.1583	constant	0.6665	0.1745
t	-0.0061	0.0145	$d_m(t-2)$	2.6722	0.9626
t^2	-0.0095	0.0052	$d_m^2(t-2)$	0.4713	0.4868
$d_m(t-2)$	3.3384	0.4796			
$d_m^2(t-2)$	-0.9350	0.3830			

The average value of *rmse* is 0.7573 which is even slightly larger than the one obtained from M3 although some more covariates are added in M4. The estimates of $\boldsymbol{\gamma}$ in fixed-effect part and σ_i^2 in random-effect part are given in Table 2. We found that the coefficients for t and t^2 in fixed-effect part and the coefficient for $d_m^2(t-2)$ in random-effect part are not significant. This explains why the performance of M3 and M4 are almost the same, and the performance of M3 cannot be improved further by model M4.

3.2 Application to dose–response study

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 levels of dose. The models discussed in this paper can be used to construct a dose–response curve for each *individual*, and therefore support clinicians to prescribe suitable doses to control the response value within a certain range for each *individual* patient. This allows a so-called *patient-specific treatment regime*.

We now discuss two elements of dose–response study illustrated by our renal anaemia data. One is to obtain a predictive individual dose–response curve, i.e., predicting the

values of Hb level if different dose levels are applied. We take 11 different dose levels from 0 to 2.5, and then calculate the predicted values for each of those dose levels. An example is given in Figure 3 for a representative patient. We present the predicted values of Hb for Month 8 based on the data collected at the first 6 months in panel (a) and the predicted values for Month 12 based the data collected at the first 10 months in panel (b). The x-axis stands for the different levels of drug DA, while the y-axis stands for the related predicted values of Hb level. This is a typical dose–response curve. Most studies can only give a common curve without considering the special feature for each individual, but 3 (a) gives the individual dose–response curve based on the common model structure learned from the data collected from all patients and the particular information for the individual collected in the first six months; while 3 (b) is analogue to 3 (a) but based on 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 level based on the figure. In Figure 3, the dotted line is the target control level of Hb. Therefore, a suitable dose level of drug DA taken by the selected patient after Month 6 should be around 1.4 unit, and be around 0.7 unit after Month 10.

Another interesting problem is to fix the dose level unchanged at the current level and to see how the predicted values of the response (Hb) will change in the next few months. For the patient we selected in Figure 3, if the dose level taken for the patient is fixed from Month 7 to Month 12 as the same dose level used in Month 6. The predicted values of Hb is shown as the dashed line in Figure 4(a). As comparisons, the real observations (the solid line) and the prediction based on actual dose levels (the dotted line) are also presented. For this patient, it seems appropriate to keep the dose level unchanged for the

next six months.

3.3 Test for agent EB

Now we consider the prediction of Hb values by using agent EB. We also use the model M3 as described in the previous subsections, but use the data collected from those taken drug EB here. We calculated the predicted values of Hb from Months 8 to 12 analogous to the discussion in the previous subsections for agent DA. The values of $rmse$ between the predicted values and the corresponding real observations are obtained. The average value of $rmse$ is 0.6982 for all 77 patients; the values of $rmse$ are given in Table 1 for other models.

The predicted individual dose–response curve can be constructed similarly. One example is given in Figure 5. The panel (a) gives the prediction of Hb level for Month 8 for different dose levels using agent EB based on the data collected up to Month 6; and the panel (b) gives the predictions based on the data collected up to Month 10. Thus, the patient needs to take about $1.1 * 200 = 220$ unit of EB after month 6 and should take about $.8 * 200 = 160$ unit after Month 10.

4 Discussion

In this paper, we proposed a mixed-effects GPFR model to construct individual dose–response curves for a patient-specific treatment regime. A nonparametric Gaussian Process Functional Regression model is used, which can fit a response curve with any 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 structure and covariance structure 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 becomes available. A comparison of the dose–response curves constructed for Month 8 and Month 12 were given in Figure 3.

An interesting problem worth further exploration is how to find an optimal solution for dose level when we want in order 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. This is an active area of research.

A comprehensive study for management of renal anaemia is also being undertaken. Data are available from a larger multi-centre dataset and our experience from the development of methodology here can be further established. A hierarchical model will be used to model and cluster the data collected from different resources (see e.g. Shi and Wang, 2008), and then to construct a more accurate dose–response curve for each individual patient. The sensitivity of patients to the epoetin agents can be assessed as reflected in their dose–response curves.

Generally this methodology has a wide range of applications wherever medications are used for the long-term maintenance of patient health: for example warfarin for management of anti-coagulation 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 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.

References

Choi, T. (2005). Posterior consistency in nonparametric regression problems under Gaussian process priors. PhD Thesis. Carnegie Mellon University.

Gönen, M. (2005). Planning a dose–response study with subject-specific doses. *Statistics in Medicine*, **24**, 2613–2623.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, **38**, 963–974.

Peters, B. C. and Walker, H. F. (1978). An iterative procedure for obtaining maximum-likelihood estimates of the parameters for a mixture of normal distributions. *SIAM J. Appl. Math.*, **35**, 362–378.

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

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

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–464.

Seeger, M. W., Kakade, S. M. and Foster, D. P. (2008). Information consistency of nonparametric Gaussian process methods. Technical report. Max

Planck Institute for Biological Cybernetics, Germany.

Shi, J. Q., Murray-Smith, R. and Titterington, D. M. (2005). Hierarchical Gaussian process mixtures for regression. *Statistics and Computing*, **15**, 31–41.

Shi, J. Q. and Wang, B. (2008). Curve prediction and clustering with mixtures of Gaussian process functional regression models. *Statistics and Computing*, in press.

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.

Tolman, C., Richardson, D., Bartlett, C. and Will, E. J. (2005). 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*, **16**, 1463–1470.

Volkova, N. and Arab, L. (2006). Evidence-based systematic literature review of hemoglobin/ hematocrit and all-cause mortality in dialysis patients. *American Journal of Kidney Disease*, **47**, 24–36.

West, R. M., Harris, K., Gilthorpe, M. S., Tolman, C. and Will, E. J. (2007). 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*, **18**, 2371–2376.

Will, E. J., Richardson, D., Tolman, C. and Bartlett, C. (2007). Development

and exploitation of a clinical decision support system for the management of renal anaemia. *Nephrology, Dialysis and Transplantation*, **22** [suppl 4], iv31–iv36.

Appendix I: 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 (11). After a straightforward calculation and simplification, we obtain

$$\frac{\partial L}{\partial \mathrm{vec}(\boldsymbol{B})} = \sum_{m=1}^{M} (\boldsymbol{y}_m - \boldsymbol{\Phi}_m \boldsymbol{B} \boldsymbol{u}_m - \boldsymbol{V}_m \boldsymbol{\gamma})' \boldsymbol{\Omega}_m^{-1} (\boldsymbol{u}_m \otimes \boldsymbol{\Phi}_m')', \tag{18}$$

$$\frac{\partial L}{\partial \boldsymbol{\gamma}} = \sum_{m=1}^{M} (\boldsymbol{y}_m - \boldsymbol{\Phi}_m \boldsymbol{B} \boldsymbol{u}_m - \boldsymbol{V}_m \boldsymbol{\gamma})' \boldsymbol{\Omega}_m^{-1} \boldsymbol{V}_m, \tag{19}$$

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

$$\mathrm{vec}(\boldsymbol{B}) = F_1(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \sigma_\varepsilon^2) = (A_{11} - A_{12} A_{22}^{-1} A_{21})^{-1} (\boldsymbol{y}_{(1)} - A_{12} A_{22}^{-1} \boldsymbol{y}_{(2)}), \tag{20}$$

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

where,

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

Those give the formulae to update vec($\boldsymbol{B}$) and $\boldsymbol{\gamma}$ in (12).

Denote the j-th element of $\boldsymbol{\theta}$ by θ_j. The gradient in terms of $(\theta_j, \sigma_\varepsilon^2, \sigma_j^2)$ are given by

$$\begin{aligned}
\frac{\partial L}{\partial \theta_j} &= \sum_{m=1}^{M} \Big\{ -\frac{1}{2}\mathrm{tr}(\boldsymbol{\Omega}_m^{-1}\frac{\partial \boldsymbol{C}_m}{\partial \theta_j}) \\
&\quad + \frac{1}{2}(\boldsymbol{y}_m - \boldsymbol{\Phi}_m\boldsymbol{B}\boldsymbol{u}_m - \boldsymbol{V}_m\boldsymbol{\gamma})'\boldsymbol{\Omega}_m^{-1}\frac{\partial \boldsymbol{C}_m}{\partial \theta_j}\boldsymbol{\Omega}_m^{-1}(\boldsymbol{y}_m - \boldsymbol{\Phi}_m\boldsymbol{B}\boldsymbol{u}_m - \boldsymbol{V}_m\boldsymbol{\gamma})\Big\}, \qquad (22) \\
\frac{\partial L}{\partial \sigma_\varepsilon^2} &= \sum_{m=1}^{M} \Big\{ -\frac{1}{2}\mathrm{tr}(\boldsymbol{\Omega}_m^{-1}) \\
&\quad + \frac{1}{2}(\boldsymbol{y}_m - \boldsymbol{\Phi}_m\boldsymbol{B}\boldsymbol{u}_m - \boldsymbol{V}_m\boldsymbol{\gamma})'\boldsymbol{\Omega}_m^{-1}\boldsymbol{\Omega}_m^{-1}(\boldsymbol{y}_m - \boldsymbol{\Phi}_m\boldsymbol{B}\boldsymbol{u}_m - \boldsymbol{V}_m\boldsymbol{\gamma})\Big\}, \qquad (23) \\
\frac{\partial L}{\partial \sigma_i^2} &= \sum_{m=1}^{M} \Big\{ -\frac{1}{2}\mathrm{tr}(\boldsymbol{\Omega}_m^{-1}\boldsymbol{W}_{mi}\boldsymbol{W}_{mi}{}') \\
&\quad + \frac{1}{2}(\boldsymbol{y}_m - \boldsymbol{\Phi}_m\boldsymbol{B}\boldsymbol{u}_m - \boldsymbol{V}_m\boldsymbol{\gamma})'\boldsymbol{\Omega}_m^{-1}\boldsymbol{W}_{mi}\boldsymbol{W}_{mi}{}'\boldsymbol{\Omega}_m^{-1}(\boldsymbol{y}_m - \boldsymbol{\Phi}_m\boldsymbol{B}\boldsymbol{u}_m - \boldsymbol{V}_m\boldsymbol{\gamma})\Big\}, \\
&\quad i = 1, \ldots, k, \text{ with } \boldsymbol{W}_{mi} \text{ being the } i\text{-th column of } \boldsymbol{W}_m. \qquad (24)
\end{aligned}$$

The above gradients are used such that the maximisation procedure step (ii) in Section 2.3 is implemented most efficiently.

Appendix II: derivation of predictive mean and predictive variance

In Section 2.4, we assume that the covariance structure is the same at the different time points. From (14), when the mean is given, we assume that $(\tilde{\tau}_1, \ldots, \tilde{\tau}_N, \tilde{\tau}^*)$ has a $(N+1)$-dimensional normal distribution $N(\mathbf{0}, \tilde{\boldsymbol{C}}_{N+1})$ with

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

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

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

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 (13). The conditional mean and variance of $\tilde{\tau}^*$ given $\tilde{\boldsymbol{\tau}}_{M+1}$ and the mean $\boldsymbol{\mu}(t)$ are given by

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

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

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 $\tilde{\boldsymbol{\mu}}_{M+1}(\boldsymbol{t}) = (\tilde{\mu}_{M+1}(t_1), \ldots, \tilde{\mu}_{M+1}(t_N))'$ is the vector of means at data points $\boldsymbol{t} = (t_1, \ldots, t_N)$ and is given by (15). This is (16) 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

$$\boldsymbol{\Lambda}_m = \begin{pmatrix} \boldsymbol{u}_m{}' \otimes \boldsymbol{\Phi}_m & \boldsymbol{V}_m \end{pmatrix} \quad \text{and} \quad \boldsymbol{\alpha} = \begin{pmatrix} \mathrm{vec}(\boldsymbol{B}) \\ \boldsymbol{\gamma} \end{pmatrix}, \tag{25}$$

the estimates (20) and (21) can be expressed as

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

where

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

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

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

where, $\hat{\boldsymbol{\Omega}}_m$ means $\boldsymbol{\Omega}_m$ with replacing the parameters by their estimates; see, for example, Section 3 in Laird and Ware (1982). As noted in Section 3.2 of this paper, these formulae do not take into account the uncertainty of the estimation of the parameters involved in $\boldsymbol{\Omega}_m$.

Therefore, we have

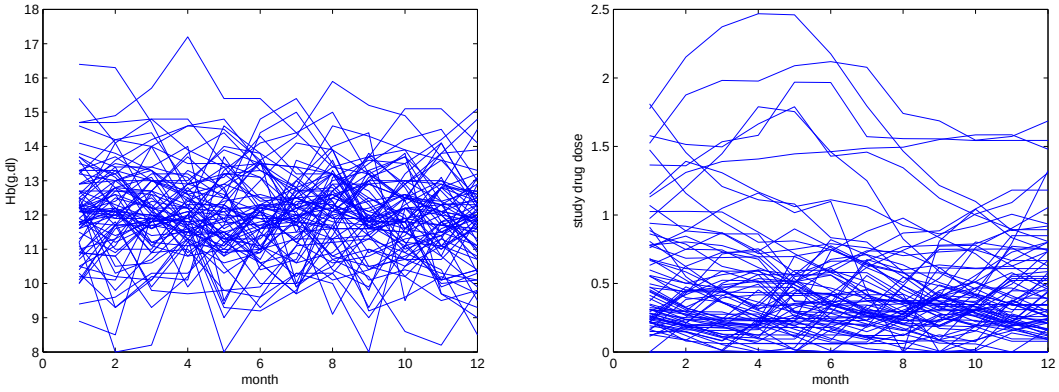
$$\begin{aligned}\mathrm{Var}[\tilde{\mu}_{M+1}(t)|\mathcal{D}] &= \mathrm{Var}[(\boldsymbol{u}_{M+1}{}'\hat{\boldsymbol{B}}{}'\boldsymbol{\Phi}(t) + \boldsymbol{v}_{M+1}{}'(t)\hat{\boldsymbol{\gamma}})|\mathcal{D}] \\ &= \mathrm{Var}[\boldsymbol{\Lambda}_{M+1}\hat{\boldsymbol{\alpha}}|\mathcal{D}] \\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

where each column of $\boldsymbol{\Lambda}_{M+1}$ is defined as in (25) 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 above results, we have

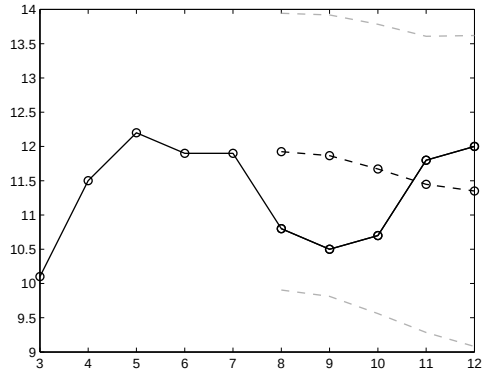
$$\mathrm{Var}(y^*|\mathcal{D}) = \hat{\sigma}^{*2} + (\boldsymbol{\Lambda}^*_{M+1} - \boldsymbol{H}'\boldsymbol{\Lambda}_{M+1})'\hat{A}^{-1}(\boldsymbol{\Lambda}^*_{M+1} - \boldsymbol{H}'\boldsymbol{\Lambda}_{M+1}). \tag{26}$$

where $\boldsymbol{\Lambda}^*_{M+1}$ is defined in (25) but evaluated at $t = t^*$. This is used to calculate the predictive variance.

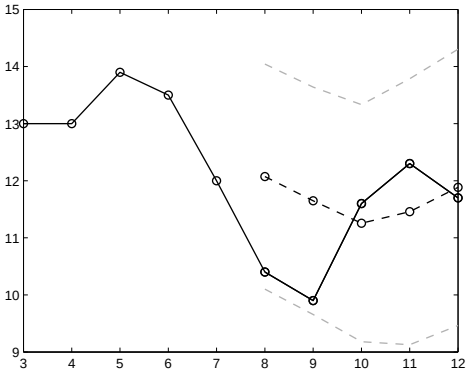


(a) (b)

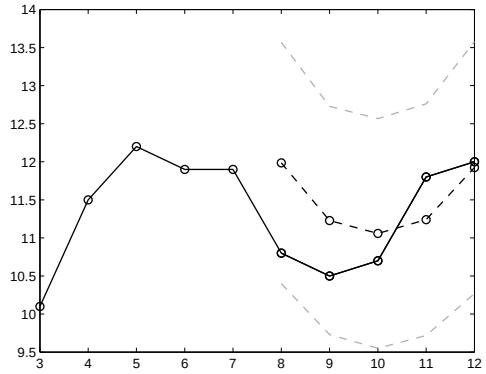
Figure 1: Renal data: (a) Hb for all patients (b) Dose level



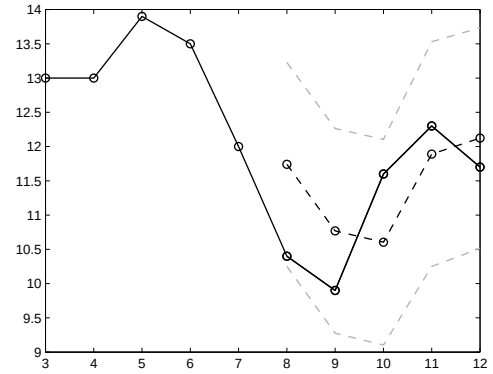
(a) M1, P1



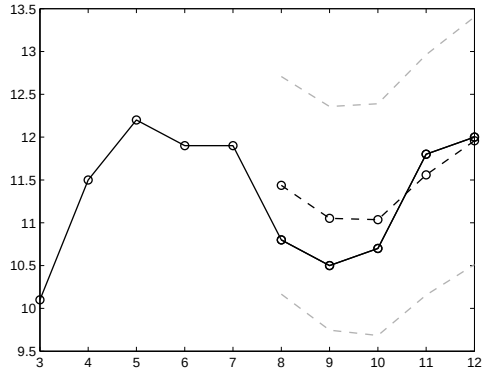
(b) M1, P2



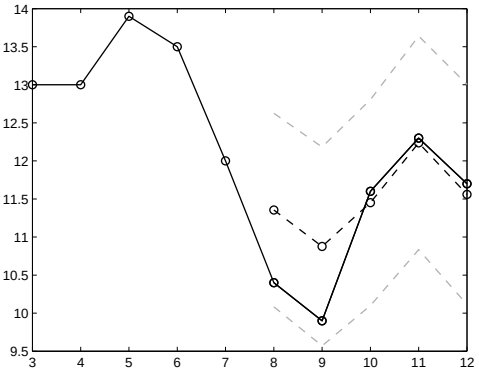
(c) M2, P1



(d) M2, P2

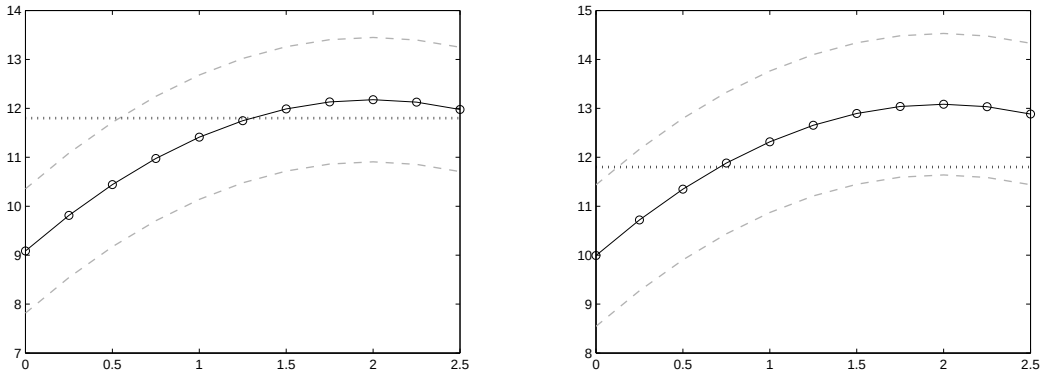


(e) M3, P1



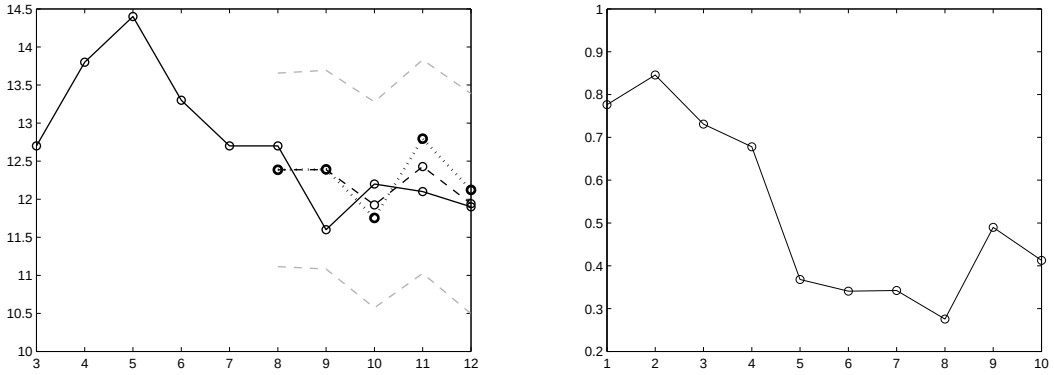
(f) M3, P2

Figure 2: Prediction of Hb for two patients P1 and P2 and the three models: the solid lines stand for real observations, and the dashed lines stand for predictions and the 95% predictive intervals.



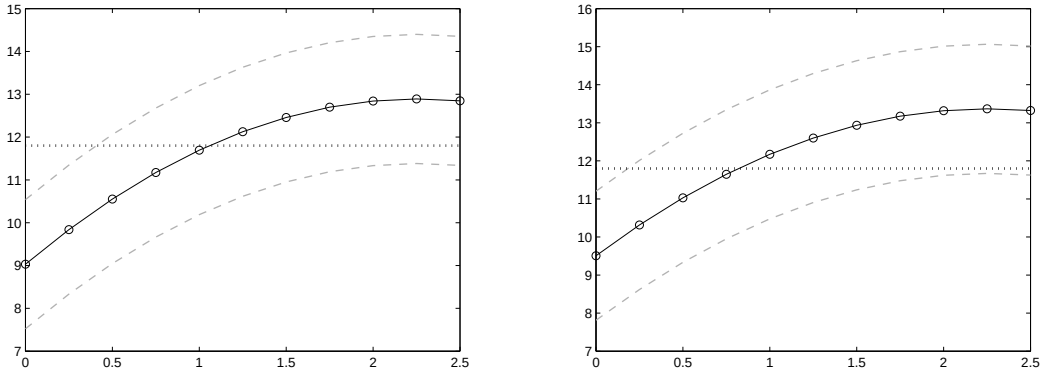
(a) The 8-th month (b) The 12-th month

Figure 3: Hb response for different dose DA level: 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 $Hb = 11.8$



(a) Hb response (b) Actual drug usage

Figure 4: Predicted values of Hb when the same dose level as in Month 6 is used: the dashed lines in (a) stand for the predictions for fixed dose level and their 95% predictive intervals; the dotted line represents the predictions by using the actual dose level as shown in (b); and the solid line represents the real observations



(a) The 8-th month

(b) The 12-th month

Figure 5: Using EB: Hb response for different dose level: 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 $Hb = 11.8$