

## MODEL SELECTION FOR FUNCTIONAL MIXED MODEL VIA GAUSSIAN PROCESS REGRESSION

Misumi, Toshihiro  
Astellas Pharma Inc.

<https://doi.org/10.5109/1798144>

---

出版情報 : Bulletin of informatics and cybernetics. 46, pp.23-35, 2014-12. Research Association  
of Statistical Sciences

バージョン :

権利関係 :



MODEL SELECTION FOR FUNCTIONAL MIXED MODEL VIA  
GAUSSIAN PROCESS REGRESSION

by

Toshihiro MISUMI

---

*Reprinted from the Bulletin of Informatics and Cybernetics  
Research Association of Statistical Sciences, Vol.46*

---

FUKUOKA, JAPAN  
2014

# MODEL SELECTION FOR FUNCTIONAL MIXED MODEL VIA GAUSSIAN PROCESS REGRESSION

By

Toshihiro MISUMI\*

## Abstract

In recent years, a functional mixed model (FMM) has attracted considerable attention in longitudinal data analysis, because of its flexibility. The FMM consists of a fixed effect or a population mean function and some subject-specific functional random effects. In this paper, we introduce the FMM constructed by using a basis expansion technique and a Gaussian process regression, and consider the model evaluation and selection problem for the estimated model. When estimating the unknown parameters included in the FMM by the maximum penalized marginal likelihood method, the FMM is extremely sensitive to the choice of tuning parameters. In order to appropriately select them, we derive two model selection criteria for the FMM based on the perspective of information or Bayesian theories by using a marginalization approach. We conduct Monte Carlo simulations to investigate the effectiveness of our proposed modeling procedures. The proposed modeling procedures for the FMM are applied to the analysis of a longitudinal gene expression data.

*Key Words and Phrases:* Gaussian process regression, Functional random effect, Longitudinal gene expression data, Mixed effects model, Model selection.

## 1. Introduction

Longitudinal data analysis has made a significant progress over the last three decades. Linear mixed models (Laird and Ware, 1982; Verbeke and Molenberghs, 2000) are quite practicable and have achieved a number of successful outcomes in medical and behavioral sciences. Nonlinear regression models, such as the smoothing spline models (Green and Silverman, 1994), the semi-parametric regression (Ruppert *et al.*, 2003), and the functional data analysis (Ramsay and Silverman, 2005) are flexible modeling approaches and there are many reports of their effectiveness.

When the longitudinal heterogeneity between subjects is quite substantial in the data, more appropriate models are required, for which linear mixed models are useful, but not enough to capture the time dependent heterogeneity. In order to overcome this issue, Rice and Wu (2001) and Guo (2002) introduced a subject specific functional random effect as a time dependent function, while the nonlinear regression was used as the population mean function, in their proposed functional mixed model (FMM). In addition, Wu and Zhang (2006), and Chen and Wang (2011) illustrated the functional

---

\* Astellas Pharma Inc., 2-5-1, Nihonbashi-Honcho, Chuo-ku, Tokyo 103-8411, Japan, and Graduate School of Science and Engineering, Chuo University, 1-13-27, Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan. toshihiro.misumi@astellas.com

random effects by using a basis expansion technique and a Gaussian process regression (GPR). The GPR is a type of Bayesian regression, and its essential idea is that the nonlinear regression function is given by a Gaussian process prior defined by a mean and covariance function, for which any finite dimensional subset has a multivariate normal distribution (Shi and Choi, 2011). In the machine learning context, many articles revealed the relationship between the Gaussian process and a Mercer kernel. In particular, several useful covariance functions are proposed (Schölkopf and Smola, 2001; Rasmussen and Williams, 2006; Bishop, 2006). On the other hand, the property of GPR from the Bayesian perspective matches up with the time dependent random effect in the mixed model framework. Accordingly, by combining the nonlinear regression modeling approach for constructing the population mean function and the GPR technique for functional random effects, the FMM enables us to flexibly estimate both the population mean curve and the subject specific prediction curves.

In this paper, we introduce the FMM which is constructed based on the GPR and the spline based Gaussian basis functions (Kawano and Konishi, 2007), and subsequently present that our model can be written as the mixed model representation. We also consider the problem of model selection for the proposed FMM. We utilize a penalized maximum likelihood method for estimating the FMM. The estimated FMM is extremely sensitive to the choice of tuning parameters such as the smoothing parameters and the number of basis functions. In order to appropriately select them, we derive two model selection criteria for the FMM based on the perspective of information and Bayesian theories: a generalized information criterion (GIC, Konishi and Kitagawa, 1996) and a generalized Bayesian information criterion (GBIC, Konishi *et al.*, 2004). The GIC and GBIC were derived as an extension of the AIC (Akaike, 1974) and BIC (Schwarz, 1978), respectively, for evaluating the candidate models given by the maximum penalized likelihood method. Meanwhile, the FMM has some random effects in common with a usual linear mixed model, that is, it has some random variables in addition to an error term. Thus, we have to take this point into account in considering the GIC/GBIC. In some efforts to overcome such an issue, a marginal AIC or a conditional AIC has been proposed in the linear mixed framework (e.g., Burnham and Anderson, 2002; Vaida and Blanchard, 2005; Liang *et al.*, 2008). We focus on simply considering the marginalization approach for the FMM in this article as a first step because the research of model selection for the FMM is very limited. After integrating random effects out from the model, we derive a marginal GIC (mGIC) and a marginal GBIC (mGBIC) for the FMM. The effectiveness of these model selection criteria is investigated through some Monte Carlo simulation studies by comparing with the conventional model selection criteria such as the marginal AIC/BIC and the marginal generalized cross validation (GCV) based on the effective degrees of freedom (Hastie and Tibshirani, 1990). We apply the proposed modeling procedures to the analysis of the yeast cell cycle data which is a longitudinal gene expression data (Spellman *et al.*, 1998).

The rest of the article is organized as follows. In Section 2, we present the FMM based on the GPR and the basis expansion technique, and its estimation procedure within the mixed model framework. Section 3 provides the conventional model selection criteria and the derivations of the mGIC and mGBIC. In Section 4, we conduct some Monte Carlo simulation studies to examine the effectiveness of our proposed model selection criteria, and apply the FMM to the analysis of the yeast cell cycle data. Summary and discussion are given in Section 5.

## 2. Functional mixed modeling

### 2.1. Functional mixed model via Gaussian process regression

Suppose we have  $n$  independent repeated measurements  $\{y(t_{ij}), t_{ij}; i = 1, \dots, n, j = 1, \dots, J_i\}$ , where  $y(t_{ij})$  denotes a response variable at  $t_{ij}$  which intends each subject  $i$  and time-point  $j$ , that is, we consider the unbalanced data situation. For representing the relationship between these measurements, concurrently with considering the longitudinal heterogeneity between subjects, the FMM via GPR is defined as

$$\begin{aligned} y(t_{ij}) &= f(t_{ij}) + \gamma_i(t_{ij}) + \varepsilon(t_{ij}), \\ \gamma_i(t) &\sim GP(0, r), \quad \varepsilon_i = (\varepsilon(t_{i1}), \dots, \varepsilon(t_{iJ_i}))' \sim N_{J_i}(\mathbf{0}, \mathbf{\Omega}_i), \end{aligned} \quad (1)$$

where  $f(t)$  is a fixed effect or a population mean function,  $\gamma_i(t)$  is a random effect function for subject  $i$  ( $i = 1, \dots, n$ ),  $\varepsilon_i$  is measurement errors, and  $\mathbf{\Omega}_i$  is a variance-covariance matrix assuming  $\mathbf{\Omega}_i = \sigma_\varepsilon^2 \mathbf{I}_{J_i}$  with a  $J_i$ -dimensional identical matrix  $\mathbf{I}_{J_i}$ . The *GP* stands for a Gaussian process with mean function  $m(t) = 0$ , and a covariance function  $r(s, t)$  which represents the variability between subjects for the times  $s, t \in [0, \infty)$ . It is assumed that  $f(t)$  and  $\gamma_1(t), \dots, \gamma_n(t)$  are expressed as

$$f(t) = \sum_{k=1}^K v_k \varphi_k(t) = \boldsymbol{\varphi}(t)' \boldsymbol{\alpha}, \quad \gamma_i(t) = \sum_{l=1}^L w_l^{(i)} \psi_l(t) = \boldsymbol{\psi}(t)' \mathbf{b}_i \quad (i = 1, \dots, n),$$

where  $\boldsymbol{\varphi}(t) = (\varphi_1(t), \dots, \varphi_K(t))'$  and  $\boldsymbol{\psi}(t) = (\psi_1(t), \dots, \psi_L(t))'$  are the basis functions, and  $\boldsymbol{\alpha} = (v_1, \dots, v_K)'$  and  $\mathbf{b}_i = (w_1^{(i)}, \dots, w_L^{(i)})'$  are their coefficients. We apply the spline based Gaussian basis functions (Kawano and Konishi, 2007) to  $\varphi_1(t), \dots, \varphi_K(t)$  and  $\psi_1(t), \dots, \psi_L(t)$  given by

$$\varphi_k(t) = \exp \left\{ -\frac{(t - \tau_{k+2})^2}{2h_\varphi^2} \right\}, \quad \psi_l(t) = \exp \left\{ -\frac{(t - \tau_{l+2})^2}{2h_\psi^2} \right\},$$

where  $\tau_k$  are equally spaced knots so that they satisfy  $\tau_1 < \dots < \tau_4 = \min(t) < \dots < \tau_{K+2} = \max(t) < \dots < \tau_{K+4}$  and  $h_\varphi = (\tau_{k+2} - \tau_k)/3$ , and the same is true for  $\tau_l$  and  $h_\psi$ . For the covariance function  $r(s, t)$ , the basis expansions for  $\gamma_i(t)$  give the Gaussian process (Wu and Zhang, 2006; Chen and Wang, 2011) as follows:

$$r(s, t) = \text{Cov}(\gamma_i(s), \gamma_i(t)) = \boldsymbol{\psi}(s)' \text{Cov}(\mathbf{b}_i) \boldsymbol{\psi}(t),$$

where  $\text{Cov}(\mathbf{b}_i)$  is an  $L \times L$  variance-covariance matrix, and let  $\text{Cov}(\mathbf{b}_i) = \mathbf{\Delta}$  for all  $i$ .

Then, the FMM in (1) can be given as the mixed model representation:

$$\begin{aligned} \mathbf{y}_i &= \mathbf{X}_i \boldsymbol{\alpha} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i, \\ \mathbf{b}_i &\sim N_L(\mathbf{0}, \mathbf{\Delta}), \quad \boldsymbol{\varepsilon}_i \sim N_{J_i}(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}_{J_i}), \end{aligned} \quad (2)$$

where  $\mathbf{y}_i = (y(t_{i1}), \dots, y(t_{iJ_i}))'$ ,  $\mathbf{X}_i = (\boldsymbol{\varphi}(t_{i1}), \dots, \boldsymbol{\varphi}(t_{iJ_i}))'$  and  $\mathbf{Z}_i = (\boldsymbol{\psi}(t_{i1}), \dots, \boldsymbol{\psi}(t_{iJ_i}))'$ . From this derivation, we can estimate the unknown parameters included in the FMM within the framework of linear mixed models.

## 2.2. Model estimation

Unknown parameters, such as the coefficient vectors  $\alpha$ , the variance parameter  $\sigma_\varepsilon^2$  and the variance-covariance matrix  $\Delta$  are estimated and the random effect vectors  $\mathbf{b}_1, \dots, \mathbf{b}_n$  are predicted by the maximum penalized likelihood method. The penalized marginal log-likelihood function in (2) is given as follows:

$$\ell_m^{PEN}(\alpha, \sigma_\varepsilon^2, \tilde{\Delta}|\mathbf{y}) = \ell_m(\alpha, \sigma_\varepsilon^2, \tilde{\Delta}|\mathbf{y}) - \frac{1}{2}n\lambda_\alpha\alpha'G_\alpha\alpha, \quad (3)$$

where  $\mathbf{y} = (\mathbf{y}'_1, \dots, \mathbf{y}'_n)'$ ,  $\tilde{\Delta} = \text{diag}(\Delta, \dots, \Delta)$ , the second term represents the penalty for the roughness of the population mean function,  $\lambda_\alpha (> 0)$  is the smoothing parameter which controls the degree of the penalty and  $G_\alpha$  is a  $K \times K$  positive semi-definite matrix. The marginal log-likelihood function  $\ell_m(\alpha, \sigma_\varepsilon^2, \tilde{\Delta}|\mathbf{y})$  can be written by

$$\begin{aligned} \ell_m(\alpha, \sigma_\varepsilon^2, \tilde{\Delta}|\mathbf{y}) &= \log \int L(\mathbf{b}, \alpha, \sigma_\varepsilon^2, \tilde{\Delta}|\mathbf{y}) d\mathbf{b} \\ &= \log \int \frac{1}{(2\pi\sigma_\varepsilon^2)^{(J+nL)/2} |\tilde{\Delta}|^{1/2}} \\ &\quad \times \exp \left[ -\frac{1}{2\sigma_\varepsilon^2} \{ \mathbf{y} - (\mathbf{X}\alpha + \mathbf{Z}\mathbf{b}) \}' \{ \mathbf{y} - (\mathbf{X}\alpha + \mathbf{Z}\mathbf{b}) \} - \frac{1}{2} \mathbf{b}' \tilde{\Delta} \mathbf{b} \right] d\mathbf{b} \\ &= \log \left[ \frac{1}{(2\pi)^{J/2} |\mathbf{V}|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{X}\alpha)' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\alpha) \right\} \right] \\ &= -\frac{J}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} (\mathbf{y} - \mathbf{X}\alpha)' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\alpha), \end{aligned}$$

where  $\mathbf{X} = (\mathbf{X}'_1, \dots, \mathbf{X}'_n)'$ ,  $\mathbf{b} = (\mathbf{b}'_1, \dots, \mathbf{b}'_n)'$ ,  $\mathbf{Z} = \text{diag}(\mathbf{Z}_1, \dots, \mathbf{Z}_n)$ ,  $\mathbf{V} = \mathbf{Z}\tilde{\Delta}\mathbf{Z}' + \sigma_\varepsilon^2 \mathbf{I}_J$  and  $J = \sum_{i=1}^n J_i$ . In order to estimate the parameters included in the FMM, we employ an expectation-maximization (EM) algorithm (Laird and Ware, 1982; Davidian and Giltinan, 1995). The EM-algorithm for estimating the proposed model is given in Appendix.

Taking  $\tilde{\Delta} = \lambda_b/\sigma_\varepsilon^2 \mathbf{I}_{nL}$  yields a ridge type penalty  $\lambda_b \mathbf{b}'\mathbf{b}$  for the full log-likelihood  $\log L(\mathbf{b}, \alpha, \sigma_\varepsilon^2, \tilde{\Delta}|\mathbf{y})$ . By expressing the variance parameter for random effect as  $\sigma_b^2 = \lambda_b/\sigma_\varepsilon^2$ , the smoothing parameter can be estimated as  $\hat{\lambda}_b = \hat{\sigma}_b^2/\hat{\sigma}_\varepsilon^2$ , where  $\hat{\sigma}_b^2$  and  $\hat{\sigma}_\varepsilon^2$  are estimated variance parameters. This result is widely known in the context of the semi-parametric regression model (Ruppert *et al.*, 2003).

After estimating the unknown parameters, a 95% pointwise standard deviation band of the population mean function  $f(t)$  can be constructed as  $f(t) \pm 1.96 \sqrt{\varphi(t)' \text{Cov}(\hat{\alpha}) \varphi(t)}$ , for which we use the linear mixed model methodology and the same way with Wu and Zhang (2006). Here, the variance-covariance matrix of  $\hat{\alpha}$  is given as

$$\begin{aligned} \text{Cov}(\hat{\alpha}) &= \left( \sum_{i=1}^n \mathbf{X}'_i \hat{\mathbf{V}}_i^{-1} \mathbf{X}_i + n\lambda_\alpha \mathbf{G}_\alpha \right)^{-1} \left( \sum_{i=1}^n \mathbf{X}'_i \hat{\mathbf{V}}_i^{-1} \mathbf{X}_i \right) \\ &\quad \times \left( \sum_{i=1}^n \mathbf{X}'_i \hat{\mathbf{V}}_i^{-1} \mathbf{X}_i + n\lambda_\alpha \mathbf{G}_\alpha \right)^{-1}, \end{aligned}$$

where  $\hat{\mathbf{V}}_i = \mathbf{Z}_i \hat{\Delta} \mathbf{Z}'_i + \hat{\sigma}_\varepsilon^2 \mathbf{I}_{J_i}$ .

### 3. Model selection

The estimated FMM is extremely sensitive to the choice the predefined number of basis functions  $K$  and  $L$ , and the smoothing parameter  $\lambda_\alpha$ . In order to appropriately select these tuning parameters, we consider the marginalized model approach based on the penalized marginal log-likelihood function (4). We first present the conventional marginal model selection criteria, and then our proposed model selection criteria are introduced.

#### 3.1. Conventional marginal model selection criteria

The conventional marginal model selection criteria, the mAIC, mBIC and mGCV for the FMM, are given as follows:

$$\begin{aligned} \text{mAIC} &= -2\ell_m(\hat{\boldsymbol{\theta}}|\mathbf{y}) + 2\widehat{edf}, \\ \text{mBIC} &= -2\ell_m(\hat{\boldsymbol{\theta}}|\mathbf{y}) + \log(n)\widehat{edf}, \\ \text{mGCV} &= -\frac{1}{n} \frac{\ell_m(\hat{\boldsymbol{\theta}}|\mathbf{y})}{(1 - \widehat{edf}/J)^2}, \end{aligned}$$

where  $\hat{\boldsymbol{\theta}}$  is the estimated vector for  $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \mathbf{b}', (\text{vech}\tilde{\boldsymbol{\Delta}})')$  with an operator  $\text{vech}$  that transforms the upper triangular elements of matrix into a vector, and thus  $\text{vech}\tilde{\boldsymbol{\Delta}}$  is a  $nL(nL+1)/2$ -dimensional vector. The  $\widehat{edf}$  is an effective degrees of freedom (Hastie and Tibshirani, 1990) for the FMM defined by

$$\widehat{edf} = \text{tr}\mathbf{H} = \sum_{i=1}^n \text{tr} \left\{ \mathbf{A}_i + \mathbf{Z}_i \hat{\boldsymbol{\Delta}} \mathbf{Z}_i' \hat{\mathbf{V}}_i^{-1} (\mathbf{I}_{J_i} - \mathbf{A}_i) \right\},$$

where  $\mathbf{A}_i = \mathbf{X}_i \left( \sum_{i=1}^n \mathbf{X}_i' \hat{\mathbf{V}}_i^{-1} \mathbf{X}_i + n\lambda_\alpha \mathbf{G}_\alpha \right)^{-1} \mathbf{X}_i' \hat{\mathbf{V}}_i^{-1}$  and the matrix  $\mathbf{H}$  is commonly called as a hat matrix or a smoother matrix, and it is obtained from the following calculations:

$$\begin{aligned} \hat{\mathbf{y}} &= \mathbf{X}\hat{\boldsymbol{\alpha}} + \mathbf{Z}\hat{\mathbf{b}} \\ &= \mathbf{A}\mathbf{y} + \mathbf{Z}\tilde{\boldsymbol{\Delta}}\mathbf{Z}'\hat{\mathbf{V}}^{-1}(\mathbf{I}_J - \mathbf{A})\mathbf{y} \\ &= \mathbf{H}\mathbf{y}, \end{aligned}$$

where  $\mathbf{A} = \mathbf{X} \left( \mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X} + n\lambda_\alpha \tilde{\mathbf{G}}_\alpha \right)^{-1} \mathbf{X}'\hat{\mathbf{V}}^{-1}$ ,  $\tilde{\mathbf{G}}_\alpha = \text{diag}(\mathbf{G}_\alpha, \dots, \mathbf{G}_\alpha)$ , and  $\tilde{\boldsymbol{\Delta}} = \text{diag}(\hat{\boldsymbol{\Delta}}, \dots, \hat{\boldsymbol{\Delta}})$ . Furthermore,  $\hat{\boldsymbol{\alpha}}$  and  $\hat{\mathbf{b}} = (\hat{\mathbf{b}}_1', \dots, \hat{\mathbf{b}}_n')'$  are here assumed to be known parameters using the result from Step 1 of EM-algorithm in the Appendix, that is, they are written as  $\hat{\boldsymbol{\alpha}} = \left( \mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X} + n\lambda_\alpha \tilde{\mathbf{G}}_\alpha \right)^{-1} \mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y}$ , and  $\hat{\mathbf{b}} = \tilde{\boldsymbol{\Delta}}\mathbf{Z}'\hat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}}) = \tilde{\boldsymbol{\Delta}}\mathbf{Z}'\hat{\mathbf{V}}^{-1}(\mathbf{I}_J - \mathbf{A})\mathbf{y}$ .

#### 3.2. Marginal generalized information criterion

Konishi and Kitagawa (1996) derived a generalized information criterion (GIC) as an extension of AIC (Akaike, 1974) for evaluating models estimated by various proce-

dures including the maximum penalized likelihood method. The GIC has been successfully used in the evaluation of a variety of models (e.g., Konishi and Kitagawa, 2008; Kawano *et al.*, 2012; Matsui *et al.*, 2013).

Using this result, the mGIC for evaluating the FMM estimated by the maximum penalized marginal likelihood is given by

$$\text{mGIC} = -2\ell_m(\hat{\boldsymbol{\theta}}|\mathbf{y}) + 2\text{tr} \left\{ \mathbf{R}(\hat{\boldsymbol{\theta}})^{-1} \mathbf{Q}(\hat{\boldsymbol{\theta}}) \right\},$$

where  $\mathbf{R}(\hat{\boldsymbol{\theta}})$  and  $\mathbf{Q}(\hat{\boldsymbol{\theta}})$  are, respectively, given by

$$\mathbf{R}(\hat{\boldsymbol{\theta}}) = -\frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \{\ell_m^{\text{PEN}(i)}(\boldsymbol{\theta})\}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \quad \mathbf{Q}(\hat{\boldsymbol{\theta}}) = \frac{1}{n} \sum_{i=1}^n \frac{\partial \{\ell_m^{\text{PEN}(i)}(\boldsymbol{\theta})\}}{\partial \boldsymbol{\theta}} \frac{\partial \{\ell_m^{(i)}(\boldsymbol{\theta})\}}{\partial \boldsymbol{\theta}'} \bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},$$

where  $\ell_m^{\text{PEN}(i)}(\boldsymbol{\theta})$  and  $\ell_m^{(i)}(\boldsymbol{\theta})$  represents the penalized marginal log-likelihood function and marginal log-likelihood function of the  $i$  th subject, respectively. The matrix derivations are referred to Matsui (2014).

### 3.3. Marginal generalized Bayesian information criterion

Konishi *et al.* (2004) proposed a generalized Bayesian information criterion (GBIC) which is an extension of Schwarz's BIC (1978), for evaluating models estimated by penalized maximum likelihood method.

In order to derive the mGBIC for evaluating the FMM, we consider the prior density for  $\boldsymbol{\theta}$  as

$$\pi(\boldsymbol{\theta}|\lambda_\alpha) = (2\pi)^{-(K-r_\alpha)/2} (n\lambda_\alpha)^{(K-r_\alpha)/2} |\mathbf{G}_\alpha|^{1/2} \exp \left( -\frac{n\lambda_\alpha}{2} \hat{\boldsymbol{\alpha}}' \mathbf{G}_\alpha \hat{\boldsymbol{\alpha}} \right),$$

where  $r_\alpha = K - \text{rank}(\mathbf{G}_\alpha)$ . Then the marginal likelihood function of  $\mathbf{y}$  can be expressed as

$$\begin{aligned} P(\mathbf{y}|\lambda_\alpha) &= \int L_m(\mathbf{y}|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}|\lambda_\alpha) d\boldsymbol{\theta} \\ &= \int \exp \left[ n \times \frac{1}{n} \log \left\{ L_m(\mathbf{y}|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}|\lambda_\alpha) \right\} \right] d\boldsymbol{\theta} \\ &\approx \frac{(2\pi)^{d/2}}{n^{d/2} |\mathbf{R}(\hat{\boldsymbol{\theta}})|} \exp \left[ n \times \frac{1}{n} \log \left\{ L_m(\mathbf{y}|\hat{\boldsymbol{\theta}}) \pi(\hat{\boldsymbol{\theta}}|\lambda_\alpha) \right\} \right], \end{aligned}$$

where  $d = K + 1$ . Multiplying minus twice of the marginal log-likelihood function, we have

$$\begin{aligned} \text{mGBIC} &= -2\ell_m(\mathbf{y}|\hat{\boldsymbol{\theta}}) + n\lambda_\alpha \hat{\boldsymbol{\alpha}}' \mathbf{G}_\alpha \hat{\boldsymbol{\alpha}} + (r_\alpha + 1) \log \left( \frac{n}{2\pi} \right) \\ &\quad - (K - r_\alpha) \log \lambda_\alpha - \log |\mathbf{G}_\alpha| + \log |\mathbf{R}(\hat{\boldsymbol{\theta}})|, \end{aligned}$$

where  $r_\alpha = K - \text{rank}(\mathbf{G}_\alpha)$ . The matrix  $\mathbf{R}(\hat{\boldsymbol{\theta}})$  is the same as that of the mGIC.

## 4. Numerical example

The effectiveness of our proposed modeling procedures is investigated through Monte Carlo simulations and a real data analysis.



#### 4.1. Monte Carlo simulations

In this section, the following two types of Monte Carlo simulation studies were conducted in order to examine the effectiveness of our proposed model selection criteria, by comparing with the conventional model selection criteria in terms of its fitting performance. The simulation data were generated as independent observations  $\{y_{ij}; i = 1, \dots, n, j = 1, \dots, J\}$  with sample sizes of  $n = 25, 50, 100$  and  $J$  time-points for 3000 repetitions using the models described below. We applied our functional mixed modeling to the generated data, and the models were fitted by the maximum penalized marginal likelihood method along with the restricted maximum likelihood (REML) based EM-algorithm. The number of basis functions  $K$  for the population mean function and  $L$  for the subject specific functional random effects, and the smoothing parameter  $\lambda_\alpha$  were selected by using the mGIC, mGBIC, mAIC, mBIC and mGCV introduced in Section 3.

##### Simulation study 1

In our first simulation study, the data were generated from the following functional mixed model, and had 20 time-points at regular intervals on  $[0, 1]$ . The model was constructed by the population mean function  $f(t)$  and the subject-specific functional random effects  $\gamma_i(t)$ :

$$y_{ij} = g(t_{ij}) + \varepsilon_{ij}, \quad g(t_{ij}) = f(t_{ij}) + \gamma_i(t_{ij}) = \boldsymbol{\varphi}(t_{ij})' \boldsymbol{\alpha} + \boldsymbol{\psi}(t_{ij})' \mathbf{b}_i,$$

$$\varepsilon_{ij} \sim N(0, \sigma^2), \quad \sigma = 0.05 \left\{ \max_{t \in [0, 1]} g(t) - \min_{t \in [0, 1]} g(t) \right\},$$

where  $\boldsymbol{\varphi}(t_{ij})$  and  $\boldsymbol{\psi}(t_{ij})$  represent  $m_\varphi$  and  $m_\psi$  dimensional vectors of the  $B$ -spline basis functions of degree three, respectively. We assume that  $m_\varphi = 6$ ,  $m_\psi = 8$  and  $\boldsymbol{\alpha} = (-8, -2, 6, 5, 7, 1)'$ , and  $\mathbf{b}_i$  are generated from  $N_8(\mathbf{0}, \Sigma)$  with  $\Sigma = (0.5^{|j-k|})_{j,k}$ . In order to assess the goodness of fit, we calculated the following two types of average mean squared errors,

$$\text{AMSPE}_P = \frac{1}{3000} \sum_{k=1}^{3000} \left\{ \frac{1}{n} \sum_{i=1}^n \frac{1}{J} \sum_{j=1}^J (f^{(k)}(t_{ij}) - \hat{f}^{(k)}(t_{ij}))^2 \right\},$$

$$\text{AMSPE}_Y = \frac{1}{3000} \sum_{k=1}^{3000} \left\{ \frac{1}{n} \sum_{i=1}^n \frac{1}{J} \sum_{j=1}^J (g^{(k)}(t_{ij}) - \hat{g}^{(k)}(t_{ij}))^2 \right\},$$

where  $\text{AMSPE}_P$  provides the average mean squared prediction error for the population functions,  $\text{AMSPE}_Y$  provides the average mean squared prediction error for all measurements,  $k$  represents an iteration number, and  $\hat{g}(t_{ij}) = \hat{f}(t_{ij}) + \hat{\gamma}_i(t_{ij})$  is a predicted value for observed data with estimated functions  $\hat{f}(t_{ij})$  and  $\hat{\gamma}_i(t_{ij})$ .

Table 1 shows the results of the simulation study 1, in which the smallest values are indicated in bold. It may be seen from the table that the models evaluated by the mGIC or the mGBIC are competitive or superior to those by the mAIC, mBIC and mGCV with respect to minimizing the  $\text{AMSPE}_Y$ . The mGIC and mGBIC tend to select better models in a balanced manner regarding the  $\text{AMSPE}_P$  and  $\text{AMSPE}_Y$  as compared to the mAIC, mBIC and mGCV.

Table 1: Comparison of average mean squared prediction errors (standard deviation) in the simulation study 1.

$n = 25$	mGIC	mGBIC	mAIC	mBIC	mGCV
AMSPE <sub>P</sub> $[\times 10^{-2}(\times 10^{-3})]$	1.04(7.70)	1.05(7.80)	1.04(7.60)	<b>1.03</b> (7.50)	1.04( <b>7.04</b> )
AMSPE <sub>Y</sub> $[\times 10^{-2}(\times 10^{-3})]$	<b>1.68</b> (3.06)	1.71( <b>3.01</b> )	1.77(3.28)	1.85(3.07)	1.71(3.55)
$n = 50$					
AMSPE <sub>P</sub> $[\times 10^{-3}(\times 10^{-3})]$	5.26(3.91)	5.26(3.91)	5.26( <b>3.90</b> )	5.26(3.91)	<b>5.25</b> ( <b>3.90</b> )
AMSPE <sub>Y</sub> $[\times 10^{-2}(\times 10^{-3})]$	<b>1.73</b> (2.59)	1.79( <b>2.58</b> )	1.97(2.71)	1.99(2.75)	1.95(2.77)
$n = 100$					
AMSPE <sub>P</sub> $[\times 10^{-3}(\times 10^{-3})]$	2.92(2.08)	2.91( <b>2.06</b> )	2.90(2.08)	<b>2.89</b> (2.08)	2.92(2.07)
AMSPE <sub>Y</sub> $[\times 10^{-2}(\times 10^{-3})]$	<b>1.81</b> ( <b>1.89</b> )	1.83(2.07)	2.14(2.03)	2.14(2.03)	2.14(2.19)

## Simulation study 2

In our second simulation study, the data were generated using the following true functional mixed model, and had 15 time-points generated from a uniform distribution on  $[0, 1]$ . The population mean function  $f(t)$  and the subject-specific functional random effects  $\gamma_i(t)$  were given as follows:

$$y_{ij} = g(t_{ij}) + \varepsilon_{ij}, \quad g(t_{ij}) = f(t_{ij}) + \gamma_i(t_{ij}),$$

$$g(t_{ij}) = 2 \cos(2\pi t_{ij}), \quad \gamma_i(t_{ij}) = b_{0i} + b_{1i} \sin(2\pi t_{ij}),$$

$$b_{0i} \sim N(0, 1), \quad b_{1i} \sim N(0, 2), \quad \varepsilon_{ij} \sim N(0, \sigma^2(t_{ij})), \quad \sigma^2(t_{ij}) = \exp \{-5(t_{ij} - 0.1)^2\}.$$

Table 2: Comparison of average mean squared prediction errors (standard deviation) in the simulation study 2.

$n = 25$	mGIC	mGBIC	mAIC	mBIC	mGCV
AMSPE <sub>P</sub> $[\times 10^{-1}(\times 10^{-1})]$	2.08(1.50)	<b>2.05</b> ( <b>1.49</b> )	2.09(1.55)	2.08(1.52)	2.08(1.51)
AMSPE <sub>Y</sub> $[\times 10^{-1}(\times 10^{-2})]$	<b>1.06</b> ( <b>2.10</b> )	1.09(2.16)	1.09(2.18)	1.07(2.11)	1.07( <b>2.10</b> )
$n = 50$					
AMSPE <sub>P</sub> $[\times 10^{-1}(\times 10^{-2})]$	<b>1.51</b> (9.26)	<b>1.51</b> ( <b>8.94</b> )	1.53(9.32)	1.53(9.32)	1.55(9.23)
AMSPE <sub>Y</sub> $[\times 10^{-1}(\times 10^{-2})]$	1.06( <b>1.53</b> )	1.06(1.59)	1.06(1.57)	1.06(1.58)	<b>1.05</b> ( <b>1.53</b> )
$n = 100$					
AMSPE <sub>P</sub> $[\times 10^{-1}(\times 10^{-2})]$	<b>1.37</b> (6.35)	<b>1.37</b> ( <b>6.33</b> )	<b>1.37</b> (6.36)	<b>1.37</b> (6.34)	<b>1.37</b> (6.34)
AMSPE <sub>Y</sub> $[\times 10^{-1}(\times 10^{-2})]$	<b>1.03</b> ( <b>9.64</b> )	1.04(9.77)	1.04(9.73)	1.04(9.73)	1.04(9.74)

Table 2 shows the results of the simulation study 2, using the AMSPE<sub>P</sub> and AMSPE<sub>Y</sub> in common with the simulation study 1. The mGIC and mGBIC also present good fitting performances as with the simulation study 1. In addition, models calculated by the mGIC and mGBIC provide smaller AMSPE<sub>P</sub>s relative to that of other criteria.

#### 4.2. Application to the yeast cell cycle data

We apply our proposed modeling procedures for the FMM to longitudinal gene expression data (Spellman *et al.*, 1998). They identified 800 genes as cell cycle related genes from all 6,178 genes in the yeast genome were measured by cDNA microarrays, and also grouped these genes into five classes, G1, G2, M, M/G1 and S, by considering peaks in the expression patterns. We focused on the repeatedly measurement “ $\alpha$  factor-based synchronization data” at 7 min intervals for 119 mins with a maximum total of 18 time-points in our analysis. Furthermore, we selected 791 genes which have two and more time-points, and in consequence, the number of genes in each class were  $N^{G1} = 297$ ,  $N^{G2} = 119$ ,  $N^M = 193$ ,  $N^{M/G1} = 113$  and  $N^S = 69$ , and the raw data are shown in the left side of Figure 1. The objective of this analysis is to explore not only the population mean functions but also gene-specific functions in each classes by using the FMM. The model was fitted by the maximum penalized likelihood method along with the REML based EM-algorithm. The number of basis functions  $K$  and  $L$ , and the smoothing parameter  $\lambda_\alpha$  were selected by model selection criteria described in Section 3.

The predicted curves with functional random effects for all measurements are shown in the center of Figure 1, and the estimated population mean functions and its 95% confidence intervals for each gene expression class are shown in the right side of Figure 1, where the tuning parameters selected by the mGIC are used. These results show that the predicted curves are well fitted for any measurements, and suggest that unknown population mean functions are also well estimated and there are obvious trends among gene classes. Table 3 shows the comparison of mean squared errors for repeated observations between above described model selection criteria, which is defined as

$$\text{MSE}_c = \frac{1}{N^c} \sum_{i=1}^{N^c} \frac{1}{J_i^c} \sum_{j=1}^{J_i^c} (y_{ij}^c - \hat{y}_{ij}^c)^2,$$

where the superscript  $c$  represents the class of gene. This result shows that the mGIC and mGBIC provide better fitting performance than conventional marginal model selection criteria.

Table 3: Comparisons of mean squared errors (standard deviation) for repeated observations among model selection criteria.

Class	mGIC	mGBIC	mAIC	mBIC	mGCV
G1[ $\times 10^{-2}(\times 10^{-2})$ ]	<b>1.81(1.54)</b>	<b>1.81(1.55)</b>	2.69(3.12)	2.69(3.12)	2.69(3.12)
G2[ $\times 10^{-2}(\times 10^{-2})$ ]	<b>1.76(1.44)</b>	<b>1.76(1.57)</b>	2.56(2.66)	2.56(2.66)	2.52(2.63)
M[ $\times 10^{-2}(\times 10^{-2})$ ]	<b>1.93(1.48)</b>	<b>1.93(1.48)</b>	2.63(2.21)	2.63(2.21)	2.63(2.21)
M/G1[ $\times 10^{-2}(\times 10^{-2})$ ]	<b>2.04(1.44)</b>	2.11(1.46)	3.16(2.71)	3.19(2.70)	3.16(2.71)
S[ $\times 10^{-2}(\times 10^{-2})$ ]	<b>4.21(3.99)</b>	4.45( <b>3.93</b> )	5.12(4.47)	5.24(4.52)	4.66(4.01)

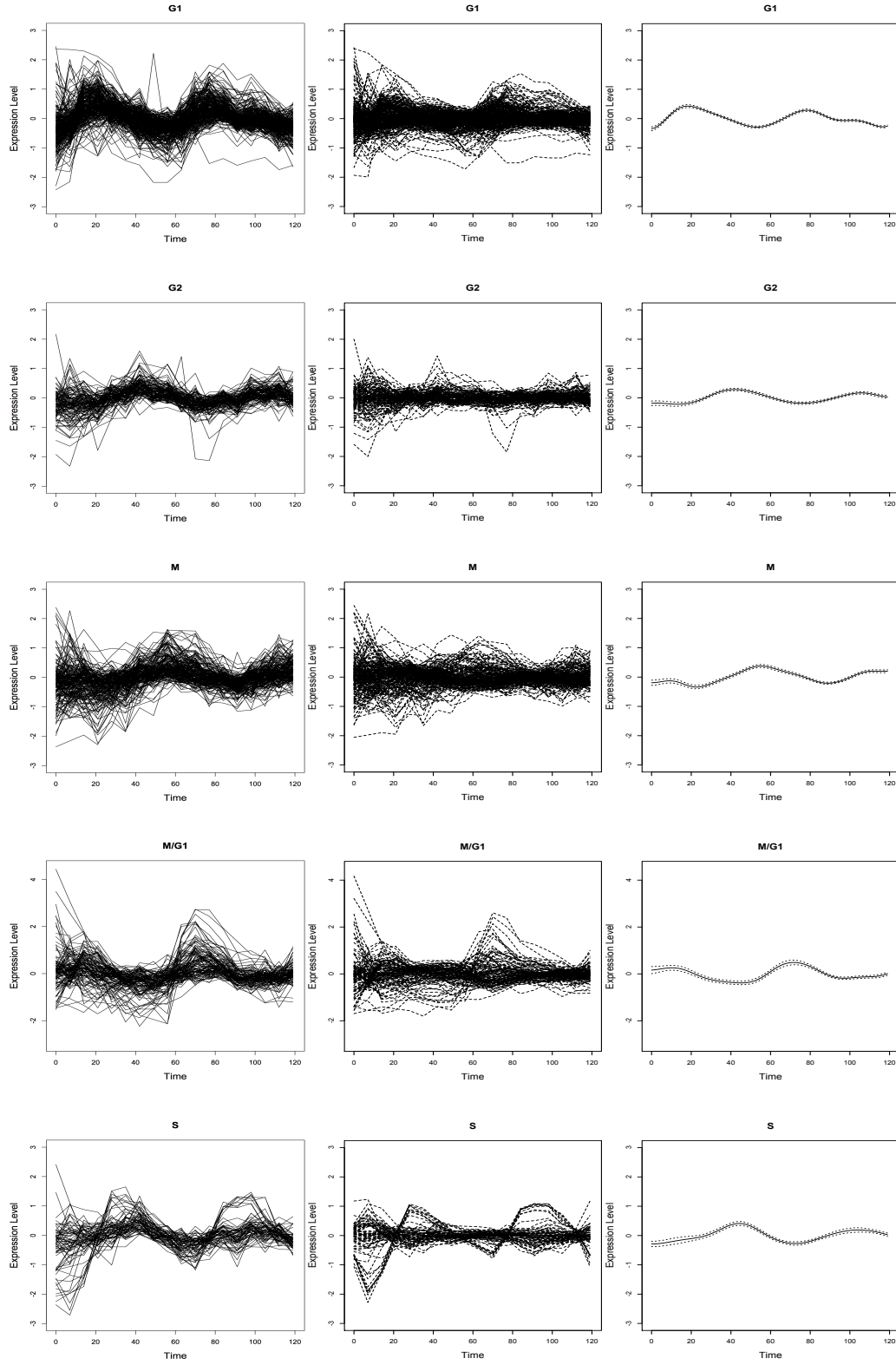


Figure 1: [Left panels] Raw data, [Center panels] Predicted curves with functional random effects, [Right panels] Estimated population mean functions (solid lines) with 95% confidence interval (dashed lines).

## 5. Summary and discussion

In this article we have introduced the functional mixed modeling via Gaussian process regression, and derived two model selection criteria for models estimated by the maximum penalized marginal likelihood method. Results from Monte Carlo simulation studies have revealed that our modeling procedures with the marginal GIC and marginal GBIC are superior or comparable to those with the conventional model selection criteria such as the marginal AIC, marginal BIC and marginal GCV. Real data analysis using the yeast cell cycle data has showed the good fitting performance and the obvious trends among the gene classes.

Although we consider a nonlinear regression as population mean function, a varying coefficient model (Hastie and Tibishriani, 1993) is also applicable. The varying coefficient model defined as  $f(t) = \beta(t)x(t)$  with  $x(t)$  being a time-dependent covariate is called a functional mixed varying coefficient model. The essential idea behind the varying coefficient model is that the coefficients of the regression model are represented by time-dependent function, and it has been deeply developed on methodological, theoretical and applied sides (see, for example, Fan and Zhang, 2008). While Matsui *et al.* (2013) proposed the GIC and the GBIC for the varying coefficient model, the functional mixed varying coefficient model can also be derived by exchanging the population function  $f(x)$  in the FMM.

Recently, Vaida and Blanchard (2005) and Liang *et al.* (2008) proposed the conditional AIC from conditional model formulation with the effective degrees of freedom in a linear mixed model framework. Greven and Kneib (2010) showed that the marginal AIC is not an asymptotically unbiased estimator of the AIC. We would like to consider more appropriate model selection criteria for the FMM using the idea of the conditional AIC in the future work.

## Appendix: EM-algorithm for the FMM

We give the EM algorithm for estimating the functional mixed model proposed in equation (1).

**Step 0** Initialize  $\hat{\sigma}_{\varepsilon(\xi)}^2 = 1$  and  $\hat{\Delta}_{(\xi)} = \mathbf{I}_L$  for the iteration number  $\xi = 0$ .

**Step 1** Set  $\xi = \xi + 1$ , update  $\hat{\alpha}_{(\xi)}$  and  $\hat{\mathbf{b}}_{i(\xi)}$  using

$$\begin{aligned}\hat{\alpha}_{(\xi)} &= \left( \sum_{i=1}^n \mathbf{X}_i' \hat{\mathbf{V}}_{i(\xi-1)}^{-1} \mathbf{X}_i + n\lambda_{\alpha} \mathbf{G}_{\alpha} \right)^{-1} \left( \sum_{i=1}^n \mathbf{X}_i' \hat{\mathbf{V}}_{i(\xi-1)}^{-1} \mathbf{y}_i \right), \\ \hat{\mathbf{b}}_{i(\xi)} &= \hat{\Delta}_{(\xi-1)} \mathbf{Z}_i' \hat{\mathbf{V}}_{i(\xi-1)}^{-1} (\mathbf{y}_i - \mathbf{X}_i \hat{\alpha}_{(\xi)}) \quad (i = 1, 2, \dots, n).\end{aligned}$$

**Step 2** Update  $\hat{\sigma}_{\varepsilon(\xi)}^2$  and  $\hat{\Delta}_{(\xi)}$  using the following conditional expectations,

$$\begin{aligned}\hat{\sigma}_{\varepsilon(\xi)}^2 &= \frac{1}{J} \sum_{i=1}^n \left\{ \hat{\varepsilon}_{i(\xi)}' \hat{\varepsilon}_{i(\xi)} + \hat{\sigma}_{\varepsilon(\xi-1)}^2 [J_i - \hat{\sigma}_{\varepsilon(\xi-1)}^2 \text{tr}(\mathbf{\Gamma}_{i(\xi-1)})] \right\}, \\ \hat{\Delta}_{(\xi)} &= \frac{1}{n} \sum_{i=1}^n \left\{ \hat{\mathbf{b}}_{i(\xi)} \hat{\mathbf{b}}_{i(\xi)}' + [\hat{\Delta}_{(\xi-1)} - \hat{\Delta}_{(\xi-1)} \mathbf{Z}_i' \mathbf{\Gamma}_{i(\xi-1)} \mathbf{Z}_i \hat{\Delta}_{(\xi-1)}] \right\},\end{aligned}$$

where  $\hat{\varepsilon}_{i(\xi)} = \mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\alpha}}_{(\xi)} - \mathbf{Z}_i \hat{\mathbf{b}}_{i(\xi)}$  and  $\boldsymbol{\Gamma}_{i(\xi-1)}$  is a  $J_i \times J_i$  matrix. When  $\boldsymbol{\Gamma}_{i(\xi-1)} = \hat{\mathbf{V}}_{i(\xi-1)}^{-1}$  then the ML based,  $\boldsymbol{\Gamma}_{i(\xi-1)} = \hat{\mathbf{V}}_{i(\xi-1)}^{-1} - \hat{\mathbf{V}}_{i(\xi-1)}^{-1} \mathbf{X}_i (\sum_{i=1}^n \mathbf{X}_i' \hat{\mathbf{V}}_i^{-1} \mathbf{X}_i + n\lambda_\alpha \mathbf{G}_\alpha)^{-1} \mathbf{X}_i' \hat{\mathbf{V}}_{i(\xi-1)}^{-1}$  then the REML based EM-algorithm are provided.

**Step 3** Repeat the above steps until some convergence conditions (e.g.  $|\hat{\sigma}_{\varepsilon(\xi)}^2 - \hat{\sigma}_{\varepsilon(\xi-1)}^2| < 10^{-5}$ ) are satisfied.

### Acknowledgement

The author would like to thank Professor Sadanori Konishi for his insightful advice, and the reviewer for his/her careful reading and valuable suggestions.

### References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transaction on Automatic Control*, **AC-19**, 716–723.
- Burnham, K. and Anderson, D. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed. Springer: New York.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*, Springer: New York.
- Chen, H. and Wang, Y. (2011). A penalized spline approach to functional mixed effects model analysis. *Biometrics*, **67**, 861–870.
- Davidian, M. and Giltinan, D. M. (1995). *Nonlinear Models for Repeated Measurement Data*. Chapman and Hall/CRC.
- Fan, J. and Zhang, J. T. (2008). Statistical methods with varying coefficient models. *Statistics and Its Interface*, **1**, 179–195.
- Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models*. Chapman and Hall/CRC.
- Greven, S. and Kneib, T. (2010). On the behaviour of marginal and conditional AIC in linear mixed models. *Biometrika*, **97**, 773–789.
- Guo, W. (2002). Functional mixed effects models. *Biometrics*, **58**, 121–128.
- Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Chapman and Hall/CRC.
- Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society: Series B*, **55**, 757–796.
- Kawano, S. and Konishi, S. (2007). Nonlinear regression modeling via regularized Gaussian basis functions. *Bulletin of Informatics and Cybernetics*, **39**, 83–96.
- Kawano, S., Misumi, T. and Konishi, S. (2012). Semi-supervised logistic discrimination via graph-based regularization. *Neural Processing Letters*, **36**, 203–216.
- Konishi, S. and Kitagawa, G. (1996). Generalized information criteria in model selection. *Biometrika*, **83**, 875–890.
- Konishi, S. and Kitagawa, G. (2008). *Information Criteria and Statistical Modeling*. Springer: New York.

- Konishi, S., Ando, T. and Imoto, S. (2004). Bayesian information criteria and smoothing parameter selection in radial basis function networks. *Biometrika*, **91**, 27–43.
- Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, **38**, 963–974.
- Liang, H., Wu H. and Zou, G. (2008). A note on conditional AIC for linear mixed-effects models. *Biometrika*, **95**, 773–778.
- Matsui, H., Misumi, T. and Kawano, S. (2013). Model selection criteria for the varying-coefficient modelling via regularized basis expansions. *Journal of Statistical Computation and Simulation*, (in press).
- Matsui, H. (2014). Model selection criteria for nonlinear mixed effects modeling. arXiv, 1402.5724.
- Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer: New York.
- Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. The MIT Press: Cambridge.
- Rice, J. A. and Wu, C. O. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics*, **57**, 253–259.
- Ruppert, D., Wand, M. P. and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge University Press.
- Schölkopf, B. and Smola, A. (2001). *Learning with Kernels*. MIT Press: Cambridge.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–464.
- Shi, J. Q. and Choi, T. (2011). *Gaussian Process Regression Analysis for Functional Data*. Chapman and Hall/CRC.
- Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Bostein, D. and Futcher, B. (1998). Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Molecular Biology of the Cell*, **9**, 3273–3297.
- Vaida, F. and Blanchard, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika*, **92**, 351–370.
- Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer: New York.
- Wu, H. and Zhang, J. (2006). *Nonparametric Regression Methods for Longitudinal Data Analysis Mixed-Effects Modeling Approaches*. Wiley: New Jersey.

Received March 18, 2014

Revised May 28, 2014