Regularized Functional Regression Modeling for Functional Response and Predictors

Matsui, Hidetoshi
Faculty of Mathematics, Kyushu University

Konishi, Sadanori
Faculty of Mathematics, Kyushu University

出版情報：MI Preprint Series. 2009-2, 2009-01-19. 九州大学大学院数理学研究院
バージョン：
権利関係：Published in Journal of Math-for-Industry, 1, p17-25, published by Faculty of Mathematics, Kyushu University

MI Preprint Series

Kyushu University
The Grobal COE Program
Math-for-Industry Education & Research Hub

Regularized functional regression modeling for functional response and predictors

H. Matsui & S. Konishi

MI 2009-2

(Received January 19, 2009)

Faculty of Mathematics
Kyushu University
Fukuoka, JAPAN

Regularized Functional Regression Modeling for Functional Response and Predictors

Hidetoshi Matsui and Sadanori Konishi
Kyushu University

Abstract

We consider the problem of constructing a functional regression modeling procedure with functional predictors and a functional response. Discretely observed data set are expressed by Gaussian basis expansions for individuals, using smoothing methods. Parameters involved in the functional regression model are estimated by the regularized maximum likelihood method, assuming that coefficient functions are expressed by basis expansions. For the selection of regularization parameters involved in the regularization method, we extend information theoretic and Bayesian model selection criteria for evaluating the estimated model. The proposed modeling strategy is applied to the analysis of real data, predicting functions rather than scalars.

Key words: Basis expansion, Functional data, Model selection criteria, Regularization.

1 Introduction

Functional data analysis provides a useful tool for analyzing a data set observed at possibly differing time points for each individual, and its effectiveness has been reported in various fields of applications such as ergonomics, meteorology and chemometrics (see e.g., Ramsay and Silverman, 2002; 2005; Ferraty and Vieu, 2006). We consider the problem of constructing a functional regression model which is the functional version of the ordinary regression model.

For functional regression models for functional predictors and scalar responses, various kinds of estimation method are considered. For example, Cardot *et al.* (2003) considered estimating the model by the principal component regression. Rossi *et al.* (2005) described a neural network approach and James (2002) extended the model to the generalized linear model. Furthermore, Araki *et al.* (2008) proposed the use of Gaussian basis functions along with the technique of regularization and information criteria and Matsui *et al.* (2008) extended the model to the functional version of the multivariate regression model.

While on the other hand, Ramsay and Dalzell (1991) considered a functional regression model which both predictor and response are given as functions, and thereafter Ramsay and Silverman (2005) considered the modeling strategy of them. They estimated the model by the least squares method, and then evaluated it by R^2 in the framework of the functional regression model. Yao *et al.* (2005) also applied the modeling strategy to sparse longitudinal data. Furthermore, Malfait and Ramsay (2003) and Harezlak *et al.* (2007) considered historical functional linear models which are used to model such dependencies of the response on the history of the predictor values. Although they mainly estimated the model by the least squares method, it may give unstable or unfavorable estimates. Moreover, R^2 is considered to be the goodness of fit, and is not appropriate to the prediction of newly observed data. Yamanishi and Tanaka (2003) estimated it by the weighted least squares method and evaluated it by the cross-validation. Although cross-validation is commonly used for selecting smoothing parameters, it can be computationally expensive.

We develop estimation and evaluation methods for functional regression models where both multiple predictors and the response are functions. Discretized observations are converted into functions using a Gaussian basis expansion along with the technique of regularization. Advantages of Gaussian basis functions are that it can provide a useful instrument for transforming discrete observations into functional form and also be applied to analyze a set of surface fitting data. Then a functional regression model is estimated by the maximum penalized likelihood method. Our modeling strategy yields more flexible results for prediction ability.

A crucial issue in functional regression modeling is the choice of smoothing parameters involved in the maximum penalized likelihood procedure. We derive model selection criteria from an information-theoretic and Bayesian approach in order to select regularization parameters effectively. The proposed modeling strategy is applied to the analysis of meteorology data. We predict the fluctuation of annual precipitation using the information of weather data, and also predict the velocity of the wind of a typhoon, observed from the generation to disappearance of it, from the position and the center atmospheric pressure.

This paper is organized as follows. In Section 2 we introduce a functional regression model with functional predictors and a functional response. In Section 3 we discuss how to estimate the model. Firstly we introduce the ordinal estimation method and secondly propose an new estimation method. In Section 4 we derive some model selection criteria for evaluating the model estimated by the proposed method. The proposed modeling strategy is applied to the analysis of real data in Section 5. Concluding remarks are discussed in Section 6.

2 Functional regression model with functional predictors and a response

Suppose we have n sets of M functional predictors and a functional response $\{(x_{\alpha m}(s), y_\alpha(t));\ s \in \mathcal{S}_m,\ t \in \mathcal{T},\ \alpha = 1, \ldots, n,\ m = 1, \ldots, M\}$. We assume that both functional predictors $x_{\alpha m}(s)$ and functional responses $y_\alpha(t)$ are expressed by basis expansions, that is,

$$x_{\alpha m}(s) = \sum_{j=1}^{J_m} \tilde{c}_{\alpha mj}\phi_{mj}(s) = \tilde{\boldsymbol{c}}'_{\alpha m}\boldsymbol{\phi}_m(s), \quad y_\alpha(t) = \sum_{k=1}^{K} \tilde{d}_{\alpha k}\psi_k(t) = \tilde{\boldsymbol{d}}'_\alpha \boldsymbol{\psi}(t) \tag{1}$$

respectively, where $\tilde{\boldsymbol{c}}_{\alpha m} = (\tilde{c}_{\alpha m1}, \ldots, \tilde{c}_{\alpha mJ_m})'$ and $\tilde{\boldsymbol{d}}_\alpha = (\tilde{d}_{\alpha 1}, \ldots, \tilde{d}_{\alpha K})'$ are coefficient vectors, $\boldsymbol{\phi}_m(s) = (\phi_{m1}(s), \ldots, \phi_{mJ_m}(s))'$ and $\boldsymbol{\psi}(t) = (\psi_1(t), \ldots, \psi_K(t))'$ are vectors of Gaussian basis functions (Kawano and Konishi, 2007) in the form

$$\phi_{mj}(s) = \exp\left\{-\frac{(s - \tau_{j+2}^{(m)})^2}{2h_m^2}\right\}, \quad \psi_k(t) = \exp\left\{-\frac{(t - \tau_{k+2})^2}{2h^2}\right\}, \tag{2}$$

where $\tau_j^{(m)}$ and τ_k are equally spaced knots so that the $\tau_j^{(m)}$ satisfy $\tau_1^{(m)} < \ldots < \tau_4^{(m)} = \min(s) < \ldots < \tau_{J+2}^{(m)} = \max(s) < \ldots < \tau_{J+4}^{(m)}$ and τ_k similarly, $h_m = (\tau_{j+2}^{(m)} - \tau_j^{(m)})/3$ and $h = (\tau_{k+2} - \tau_k)/3$. Coefficients $\tilde{\boldsymbol{c}}_{\alpha m}$ and $\tilde{\boldsymbol{d}}_\alpha$ are obtained by smoothing techniques described in Appendix A. In order to model the relationship between predictors and a response, we consider the following functional regression model (Ramsay and Silverman, 2005; Shimokawa *et al.*, 2000):

$$y_\alpha(t) = \beta_0(t) + \sum_{m=1}^{M} \int_{\mathcal{S}_m} x_{\alpha m}(s)\beta_m(s,t)ds + \varepsilon_\alpha(t), \tag{3}$$

where $\beta_0(t)$ is a parameter function, $\beta_m(s,t)$ are bivariate coefficient functions which impose varying weights on $x_{\alpha m}(s)$ at arbitrary time $t \in \mathcal{T}$, and $\varepsilon_\alpha(t)$ are error functions. The coefficient functions $\beta_m(s,t)$ are assumed to be expressed using the same basis functions as those used for the predictor and response functions as follows:

$$\beta_m(s,t) = \sum_{j,k} \phi_{mj}(s)b_{mjk}\psi_k(t) = \boldsymbol{\phi}'_m(s)B_m\boldsymbol{\psi}(t), \tag{4}$$

where $B_m = (b_{mjk})_{j,k}$ are $J_m \times K$ coefficient matrices.

The function $\beta_0(t)$ plays the role of a constant term in the standard regression model. Here, we eliminate it by centering the functional regression model (3) for the subsequent estimation procedure. Centered predictors $x_{\alpha m}(s)$ and responses $y_\alpha(t)$ are obtained by

$$\begin{aligned} x^*_{\alpha m}(s) &= x_{\alpha m}(s) - \bar{x}_m(s) & y^*_\alpha(t) &= y_\alpha(t) - \bar{y}(t) \\ &= \tilde{\boldsymbol{c}}'_{\alpha m}\boldsymbol{\phi}_m(s) - \bar{\boldsymbol{c}}'_m\boldsymbol{\phi}_m(s) & &= \tilde{\boldsymbol{d}}'_\alpha\boldsymbol{\psi}(t) - \bar{\boldsymbol{d}}'\boldsymbol{\psi}(t) \\ &= \boldsymbol{c}'_{\alpha m}\boldsymbol{\phi}_m(s), & &= \boldsymbol{d}'_\alpha\boldsymbol{\psi}(t) \end{aligned} \tag{5}$$

respectively, where $\boldsymbol{c}_{\alpha m} = \tilde{\boldsymbol{c}}_{\alpha m} - \bar{\boldsymbol{c}}_m$ and $\boldsymbol{d}_\alpha = \tilde{\boldsymbol{d}}_\alpha - \bar{\boldsymbol{d}}$ with $\bar{\boldsymbol{c}}_m = \sum_\alpha \tilde{\boldsymbol{c}}_{\alpha m}/n$ and $\bar{\boldsymbol{d}} = \sum_\alpha \tilde{\boldsymbol{d}}_\alpha/n$. Then (3) can be rewritten in the form

$$y^*_\alpha(t) = \sum_{m=1}^{M} \int_{\mathcal{S}_m} x^*_{\alpha m}(s)\beta_m(s,t)ds + \varepsilon^*_\alpha(t), \tag{6}$$

where $\varepsilon^*_\alpha(t) = \varepsilon_\alpha(t) - \bar{\varepsilon}(t)$. Using assumptions (1) and (4), the functional regression model (6) can be expressed using matrix and vector notation as

$$\begin{aligned} \boldsymbol{d}'_\alpha\boldsymbol{\psi}(t) &= \sum_{m=1}^{M} \boldsymbol{c}'_{\alpha m} W_{\phi_m} B_m \boldsymbol{\psi}(t) ds + \varepsilon^*_\alpha(t) \\ &= \boldsymbol{z}'_\alpha B\boldsymbol{\psi}(t) + \varepsilon^*_\alpha(t), \end{aligned} \tag{7}$$

where $\boldsymbol{z}_\alpha = (\boldsymbol{c}'_{\alpha 1}W_{\phi_1}, \ldots, \boldsymbol{c}'_{\alpha M}W_{\phi_M})'$ with $W_{\phi_m} = \int \boldsymbol{\phi}_m(s)\boldsymbol{\phi}'_m(s)ds$ and $B = (B'_1, \ldots, B'_M)'$. When we use the Gaussian basis functions given in (2), (j,k)-th elements of W_{ϕ_m} are given by

$$W^{(j,k)}_{\phi_m} = \sqrt{\pi h_m^2} \exp\left\{ -\frac{(\tau^{(m)}_{j+2} - \tau^{(m)}_{k+2})^2}{4h_m^2} \right\},$$

and therefore W_{ϕ_m} is positive definite. From equation (7), the problem of estimating the coefficient functions $\beta_m(s,t)$ in (3) is replaced by the problem of estimating the parameter matrix B.

3 Estimation

We consider the problem of estimating the parameter matrix B. First we describe the least squares method, given by Ramsay and Silverman (2005), and then describe the proposed method, based on the maximum likelihood procedure.

3.1 Least squares method

Ramsay and Silverman (2005) and Shimokawa *et al.* (2000) estimated B in the model (7) by minimizing the integrated residual sum of squares given by

$$\begin{aligned}\mathrm{LMSSE}(B) &= \sum_{\alpha=1}^{n} \int_{\mathcal{T}} \left[y_\alpha^*(t) - \sum_{m=1}^{M} \int_{\mathcal{S}_m} x_{\alpha m}^*(s)\beta_m(s,t)ds \right]^2 dt \\ &= \int_{\mathcal{T}} \mathrm{tr}\left\{ (D\boldsymbol{\psi}(t) - ZB\boldsymbol{\psi}(t)) (D\boldsymbol{\psi}(t) - ZB\boldsymbol{\psi}(t))' \right\} dt \\ &= \mathrm{tr}\left\{ (D - ZB) W_\psi (D - ZB)' \right\}, \qquad (8)\end{aligned}$$

where $D = (\boldsymbol{d}_1^*, \dots, \boldsymbol{d}_n^*)'$, $Z = (\boldsymbol{z}_1, \dots, \boldsymbol{z}_n)'$ and $W_\psi = \int_{\mathcal{T}} \boldsymbol{\psi}(t)\boldsymbol{\psi}'(t)dt$. Therefore the least squares estimator $\hat{B}$ is given by

$$\mathrm{vec}\hat{B} = (W_\psi \otimes Z'Z)^{-1}\mathrm{vec}(Z'DW_\psi). \qquad (9)$$

When we use Gaussian basis functions (2) W_ψ is nonsingular, so then $\hat{B}$ can be expressed as

$$\hat{B} = (Z'Z)^{-1}Z'D. \qquad (10)$$

This has the same form as a least squares estimator for ordinary multivariate regression models with a design matrix Z and a response matrix D.

3.2 Maximum likelihood method

Here we consider estimating the functional regression model (7) in the framework of the maximum likelihood method. Suppose error functions $\varepsilon_\alpha^*(t)$ are represented by a linear combination of basis functions $\psi_k(t)$, which are the same as those for the response functions $y_\alpha^*(t)$, that is,

$$\varepsilon_\alpha^*(t) = \sum_{k=1}^{K} e_{\alpha k}\psi_k(t) = \boldsymbol{e}_\alpha'\boldsymbol{\psi}(t), \qquad (11)$$

where $\boldsymbol{e}_\alpha = (e_{\alpha 1}, \dots, e_{\alpha K})'$ are K-dimensional vectors which are independent and identically normally distributed with mean vector $\mathbf{0}$ and variance-covariance matrix Σ. Therefore, (7) can be represented as

$$\boldsymbol{d}_\alpha'\boldsymbol{\psi}(t) = \boldsymbol{z}_\alpha' B\boldsymbol{\psi}(t) + \boldsymbol{e}_\alpha'\boldsymbol{\psi}(t), \quad \boldsymbol{e}_\alpha \overset{i.i.d}{\sim} N_K(\mathbf{0}, \Sigma). \qquad (12)$$

By multiplying the equation (12) by $\boldsymbol{\psi}'(t)$ and then integrating with respect to $\mathcal{T}$, (12) can be rewritten as

$$\boldsymbol{d}'_\alpha W_\psi = \boldsymbol{z}'_\alpha B W_\psi + \boldsymbol{e}'_\alpha W_\psi. \tag{13}$$

Then if W_ψ is nonsingular we obtain

$$\boldsymbol{d}_\alpha = B'\boldsymbol{z}_\alpha + \boldsymbol{e}_\alpha, \quad \boldsymbol{e}_\alpha \overset{i.i.d}{\sim} N_K(\boldsymbol{0}, \Sigma), \tag{14}$$

which has the same form as a multivariate regression model with predictors $\boldsymbol{z}_\alpha$ and responses $\boldsymbol{d}_\alpha$.

From (14) we can obtain a statistical model for a functional response y_α given a functional predictor $\boldsymbol{x}_\alpha$ as follows:

$$f(y_\alpha|\boldsymbol{x}_\alpha;\boldsymbol{\theta}) = \frac{1}{(2\pi)^{K/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{d}_\alpha - B'\boldsymbol{z}_\alpha)'\Sigma^{-1}(\boldsymbol{d}_\alpha - B'\boldsymbol{z}_\alpha)\right\}, \tag{15}$$

where $\boldsymbol{\theta} = \{B, \Sigma\}$ is a parameter vector. Therefore, maximum likelihood estimators of B and Σ are given by

$$\hat{B} = (Z'Z)^{-1}Z'D, \qquad \hat{\Sigma} = \frac{1}{n}(D - Z\hat{B})'(D - Z\hat{B}) \tag{16}$$

respectively. Comparing this result with (10), we find that the maximum likelihood estimator of B coincides with the least squares estimator.

3.3 Maximum penalized likelihood method

Since least squares or maximum likelihood method sometimes results in unstable estimators, we consider estimating the functional regression model using the maximum penalized likelihood method. From the statistical model (15) the penalized log-likelihood function is given by

$$l_\lambda(\boldsymbol{\theta}) = \sum_{\alpha=1}^{n} f(y_\alpha|\boldsymbol{x}_\alpha;\boldsymbol{\theta}) - \frac{n}{2}\mathrm{tr}\left\{B'(\Lambda_M \odot \Omega)B\right\}, \tag{17}$$

where Λ_M is a $(\sum_m J_m) \times (\sum_m J_m)$ matrix of regularization parameters $\lambda_1, \ldots, \lambda_M$ that adjust a fluctuation of B, that is, $\Lambda_M = \boldsymbol{\lambda}_M\boldsymbol{\lambda}'_M$ with $\boldsymbol{\lambda}_M = (\sqrt{\lambda_1}\mathbf{1}'_{J_1}, \ldots, \sqrt{\lambda_M}\mathbf{1}'_{J_M})'$. Furthermore, $\odot$ represents the Hadamard product and Ω is a $(\sum_m J_m) \times (\sum_m J_m)$ positive semi-definite matrix. Maximizing the function (17), maximum penalized likelihood estimators $\hat{B}$, $\hat{\Sigma}$ are given by

$$\begin{aligned} \mathrm{vec}\hat{B} &= \left(\hat{\Sigma}^{-1} \otimes Z'Z + nI_K \otimes (\Lambda_M \odot \Omega)\right)^{-1} (\hat{\Sigma}^{-1} \otimes Z')\mathrm{vec}D, \\ \hat{\Sigma} &= \frac{1}{n}(D - Z\hat{B})'(D - Z\hat{B}) \end{aligned} \tag{18}$$

respectively, where $\otimes$ represents a Kronecker product. Since $\hat{B}$ and $\hat{\Sigma}$ depend on each other, we provide an initial value for the variance covariance matrix; then they are updated until convergence. Therefore, the maximum penalized likelihood estimator of D is given by

$$\begin{aligned}\mathrm{vec}\hat{D} &= \mathrm{vec}(Z\hat{B}) \\ &= S_\lambda \mathrm{vec} D,\end{aligned} \tag{19}$$

where $S_\lambda = (I_K \otimes Z)(\hat{\Sigma}^{-1} \otimes Z'Z + nI_K \otimes (\Lambda_M \odot \Omega))^{-1}(\hat{\Sigma}^{-1} \otimes Z')$ is a hat matrix for vec$D$. Substituting the maximum penalized likelihood estimator $\hat{\boldsymbol{\theta}} = \{\hat{B}, \hat{\Sigma}\}$ into (15) we obtain the statistical model

$$f(y_\alpha|\boldsymbol{x}_\alpha;\hat{\boldsymbol{\theta}}) = \frac{1}{(2\pi)^{K/2}|\hat{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{d}_\alpha - \hat{B}'\boldsymbol{z}_\alpha)'\hat{\Sigma}^{-1}(\boldsymbol{d}_\alpha - \hat{B}'\boldsymbol{z}_\alpha)\right\}. \tag{20}$$

4 Model selection criteria

Since the statistical model (20) estimated by the regularization method depends on the regularization parameters $\lambda_1, \ldots, \lambda_M$, selection of these values is an important issue. Although cross-validation is widely used for the regularization parameter selection, it can be computationally expensive. We propose the use of certain model selection criteria for evaluating the functional regression model. We select the model that minimizes these values and then consider the corresponding model to be the optimal model.

(1) Generalized cross validation

Generalized cross validation (GCV; Craven and Wahba, 1979) for evaluating the functional regression model (20) is obtained by applying the hat matrix S_λ given in (19), that is,

$$\mathrm{GCV} = \frac{\mathrm{tr}\left\{(D - ZB)'(D - ZB)\right\}}{nK\left(1 - \mathrm{tr}S_\lambda/(nK)\right)^2}. \tag{21}$$

(2) Modified AIC

Hastie and Tibshirani (1990) modified the AIC (Akaike, 1973) for evaluating the model estimated by the regularization method by substituting a trace of the hat matrix for the number of degrees of freedom, since the hat matrix can be viewed as a measure of the complexity of the model estimated by the regularization method. Using this result, the modified AIC evaluating the model (20) is given by

$$\mathrm{mAIC} = -2\sum_{\alpha=1}^{n} f(y_\alpha|\boldsymbol{x}_\alpha;\hat{\boldsymbol{\theta}}) + 2\mathrm{tr}S_\lambda. \tag{22}$$

A problem may arise in the theoretical justification for the use of the bias-correction terms in MAIC, since AIC covers only models estimated by the maximum likelihood method.

(3) Generalized information criterion

Imoto and Konishi (2003) derived an information criterion GIC (Konishi and Kitagawa (1996)) for evaluating a statistical model estimated by the maximum penalized likelihood method. Using this result, the GIC for evaluating the model (20) is given by

$$\mathrm{GIC} = -2\sum_{\alpha=1}^{n} f(y_\alpha|\boldsymbol{x}_\alpha;\hat{\boldsymbol{\theta}}) + 2\mathrm{tr}\{R_\lambda(\hat{\boldsymbol{\theta}})^{-1}Q_\lambda(\hat{\boldsymbol{\theta}})\}, \tag{23}$$

where $R_\lambda(\boldsymbol{\theta})$, $Q_\lambda(\boldsymbol{\theta})$ are given by

$$R_\lambda(\boldsymbol{\theta}) = -\frac{1}{n}\sum_{\alpha=1}^{n} \frac{\partial^2 \left\{\log f(y_\alpha|\boldsymbol{x}_\alpha;\boldsymbol{\theta}) - \mathrm{tr}\left\{B'(\Lambda_M \odot \Omega)B\right\}/2\right\}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}, \tag{24}$$

$$Q_\lambda(\boldsymbol{\theta}) = \frac{1}{n}\sum_{\alpha=1}^{n} \frac{\partial \left\{\log f(y_\alpha|\boldsymbol{x}_\alpha;\boldsymbol{\theta}) - \mathrm{tr}\left\{B'(\Lambda_M \odot \Omega)B\right\}/2\right\}}{\partial\boldsymbol{\theta}} \cdot \frac{\partial \log f(y_\alpha|\boldsymbol{x}_\alpha;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} \tag{25}$$

respectively.

(4) Generalized Bayesian information criterion

The Bayesian information criterion (BIC) has been proposed by Schwarz (1978), from the viewpoint of Bayesian inference, based on the idea of maximizing the posterior probability of candidate models. However, the BIC only covers models estimated by the maximum likelihood method. Konishi *et al.* (2004) extended the BIC so that it could be used for evaluating models fitted by the maximum penalized likelihood method, thus deriving GBIC. We derive the GBIC for evaluating the model (20) fitted by the maximum penalized likelihood method, which is given by

$$\begin{aligned}\mathrm{GBIC} = &-2\sum_{\alpha=1}^{n} f(y_\alpha|\boldsymbol{x}_\alpha;\hat{\boldsymbol{\theta}}) + n\mathrm{tr}\left\{B'(\Lambda_M \odot \Omega)B\right\} + (r+Kq)\log n \\ &- (r+Kq)\log(2\pi) - K\log|\Lambda_M \odot \Omega|_+ + \log|R_\lambda(\hat{\boldsymbol{\theta}})|,\end{aligned} \tag{26}$$

where $q = p - \mathrm{rank}(\Omega)$, $p = \sum_m J_m$, $r = K(K+1)/2$ and $|\cdot|_+$ denotes the product of the non-zero eigenvalues of a matrix. The derivation of (26) is given in Appendix B.

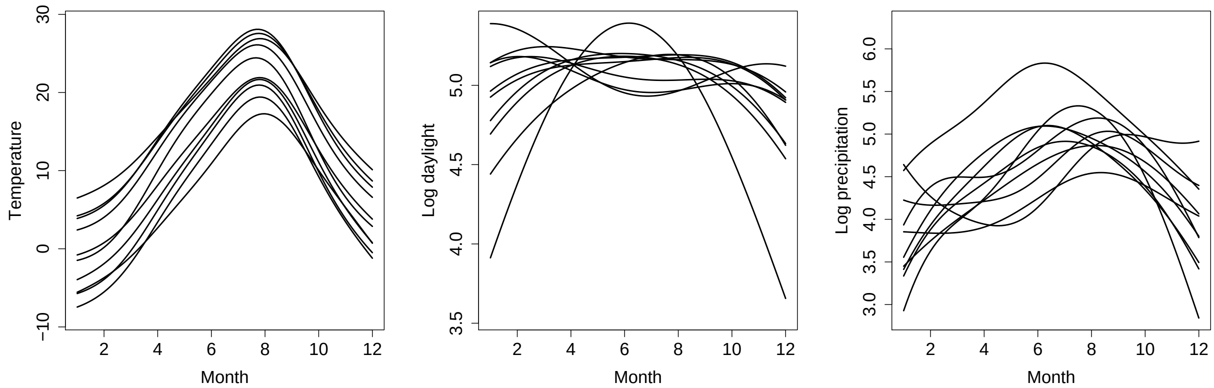

Figure 1: Examples of weather data converted into functions.

5 Real data example

In this section we apply the proposed functional regression modeling strategy to the analysis of real data, examining the effectiveness of the modeling strategy. Firstly we applied it to the analysis of Japanese weather data, predicting the variation of monthly precipitation. We then applied it to the analysis of typhoon data, and predicted the velocity of the wind of the typhoon using information about its position and pressure, observed from the typhoon's generation to its disappearance.

5.1 Japanese weather data

Weather data, available on Chronological Scientific Tables 2005, are recorded from January to December at 79 weather stations in Japan. These data include the annual monthly average temperature, monthly total times of daylight and monthly total precipitation. These data are averaged over the values obtained from 1971 to 2000. We consider predicting monthly total precipitation using the temperature and times of daylight. For daylight and precipitation data we used the logarithms of observed data.

We performed some pre-processings before applying functional regression modeling. First, we obtained functional data sets by smoothing the data via regularized Gaussian basis function expansion. The resulting functional data sets are shown in Figure 1. Next, the 79 observed data sets were randomly divided into 45 training data sets and 34 test data sets. The training data were centered by subtracting the sample average. We treated temperature and daylight functions as predictors and the precipitation function as a response, thereby constructing a functional regression model.

The model was estimated by the maximum likelihood and maximum penalized like-

Table 1: Results on the analysis of weather data. λ_1 and λ_2 are regularization parameters selected by each model selection criteria.

	MLE	GCV	mAIC	GIC	GBIC
λ_1	—	2.51×10^{-2}	3.98×10^{-2}	5.01×10^{-2}	8.91×10^{-1}
λ_2	—	1.26×10^{-1}	1.78×10^{-1}	1.78×10^{-1}	8.91×10^{-1}
Test error	7.25×10^{-2}	5.98×10^{-2}	5.90×10^{-1}	5.83×10^{-2}	5.37×10^{-2}

lihood method; four model selection criteria were then used to evaluate the model for maximum penalized likelihood estimates. We used the average squared errors between the smoothed test data and the predicted functional data at 100 time points as the test error.

Table 1 shows regularization parameters for temperature (λ_1) and daylight (λ_2) selected by each model selection criteria and test errors of corresponding models. From these results we find that the maximum penalized likelihood method is superior to the maximum likelihood method in prediction accuracy. In particular, for the four model selection criteria, GBIC minimized the test error. Figure 3 shows the results of fitting eight weather stations with the test set. These figures reveal that the predicting functions captured the original data well. The estimated coefficient functions of each predictor are shown in Figure 2. This figure shows that while the temperature around January and the times of daylight around October have negative weights, the temperature at the end of the year and the times of daylight around March have a positive weight for predicting the precipitation. Therefore, if the former values increase the precipitation decreases, and if the latter values increase the precipitation increases.

5.2 Typhoon data

Next, we applied functional regression modeling to the analysis of typhoon data. Typhoon data are available on the website "Digital Typhoon: Typhoon Images and Information" [1]. The data set contains the position (longitude and latitude), the central atmospheric pressure and the wind velocity near the center of typhoons generated from 1951 to 2006. Each of them was observed every six hours from the generation to the disappearance of the typhoon. However, for the wind velocity only those generated after 1977 are observed. Thus, we predicted the velocity of the wind of the typhoons before 1976 using information

[1] URL: http://agora.ex.nii.ac.jp/digital-typhoon/

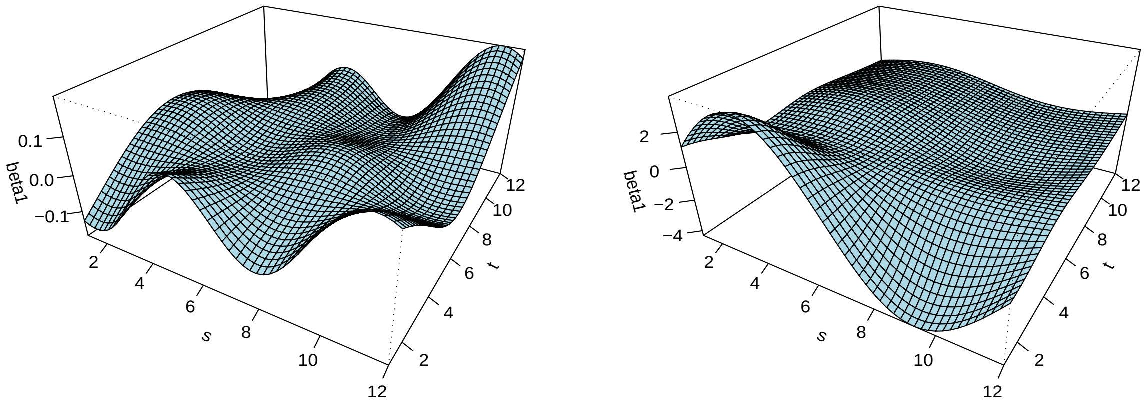

Figure 2: Estimated coefficient functions.

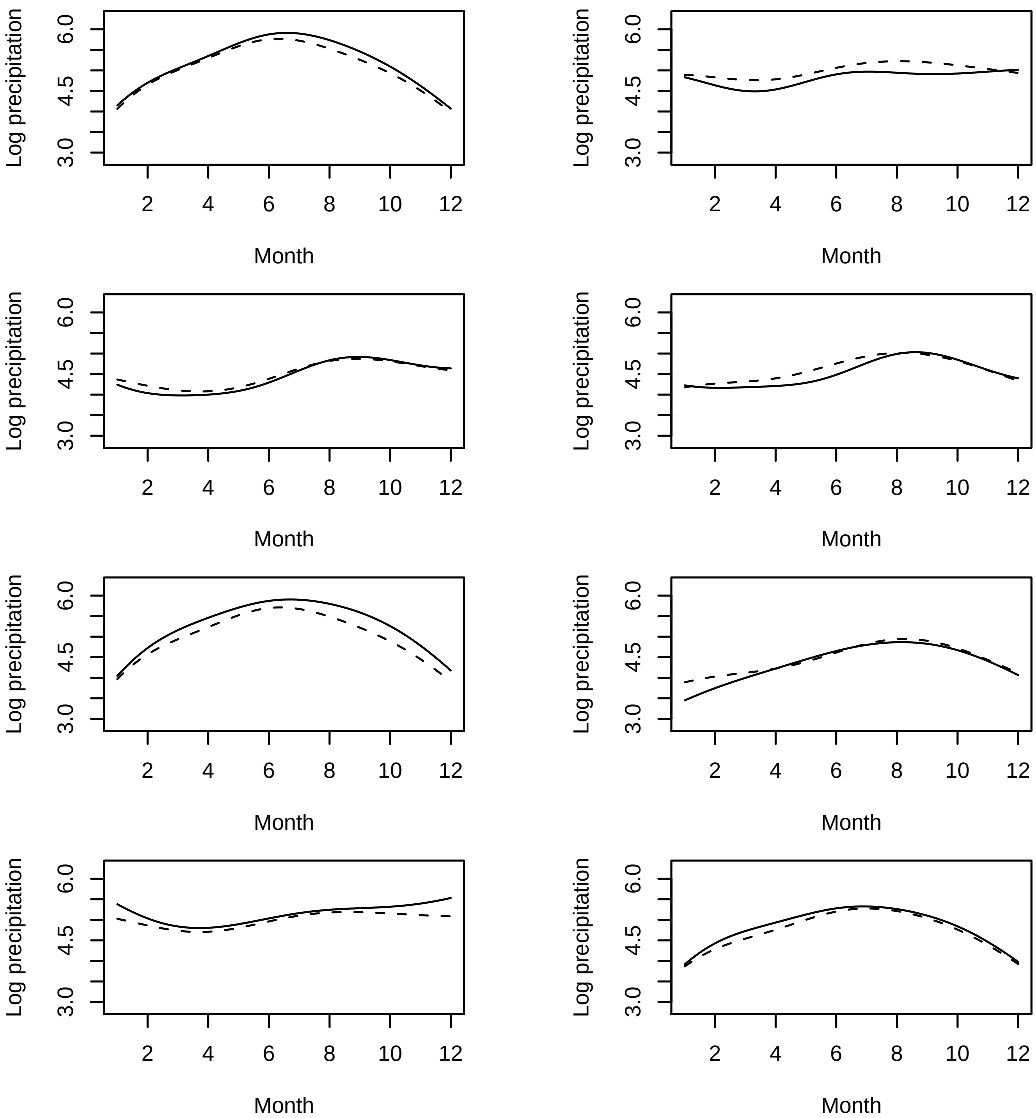

Figure 3: Results on fitting the test data for 8 stations. Solid lines show smoothed test data and dashed lines show the predicted functions by the functional regression model.

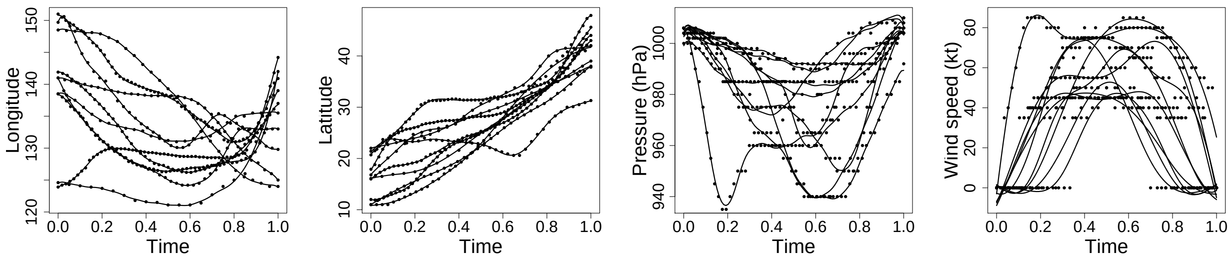

Figure 4: 10 examples of raw typhoon data and functional data. Top left: Longitude. Top right: Latitude. Bottom left: Central atmospheric pressure. Bottom right: Velocity of wind near centers.

about the position and the central atmospheric pressure. Moreover, we conducted variable selection to determine which pair of predictors are the most informative.

The data were standardized with respect to the generation and disappearance time of all typhoons by scaling time points at which the data were observed by scaling the time to [0,1]. Since survival times differ for each typhoon, the time points also differ. Therefore, it is difficult to apply the ordinary regression model directly. However, by treating these data as smooth functions, we can analyze them easily. Figure 4 shows 10 examples of typhoon data smoothed by the technique described in the Appendix A. We treated longitude, latitude and central atmospheric pressure as functional predictors and the wind speed as a response, constructing a functional regression model. We estimated the model by the maximum penalized likelihood method and then evaluated it by the model selection criterion GBIC. Furthermore, in order to examine which variable has the greatest effect on predicting the velocity of the wind, we also conducted variable selection using GBIC.

Table 2 shows the result of the model selection . From this table, the model with X_2 and X_3 as predictors minimize the GBIC. It indicates that latitude and center atmospheric pressure are the most informative for predicting the velocity of the wind. We fitted the data observed before 1976 to the selected model in order to predict the velocity of the wind. The results are shown in Figure 5. In particular, the two typhoons depicted with heavy lines (Typhoon No. 22 in 1955 and No. 12 in 1956) had strong velocities. Indeed, these typhoons caused damage to tens of thousands of properties and also damaged a great deal of agricultural land. The result shows the effectiveness of our modeling strategy.

Table 2: Results on the model selection for typhoon data. Predictors X_1, X_2 and X_3 indicate longitude, latitude and central atmospheric pressure respectively, and λ_m are the corresponding regularization parameters selected by GBIC.

Predictor	GBIC	λ_1	λ_2	λ_3
X_1	2053	3.16×10^1	—	—
X_2	2028	—	1.00×10^2	—
X_3	2021	—	—	3.16×10^1
X_1, X_2	2038	3.98×10^1	3.98×10^1	—
X_2, X_3	2013	—	3.16×10^1	1.00×10^2
X_1, X_3	2039	1.00×10^2	—	2.51×10^2
FULL	2074	1.00×10^1	1.00×10^1	5.62×10

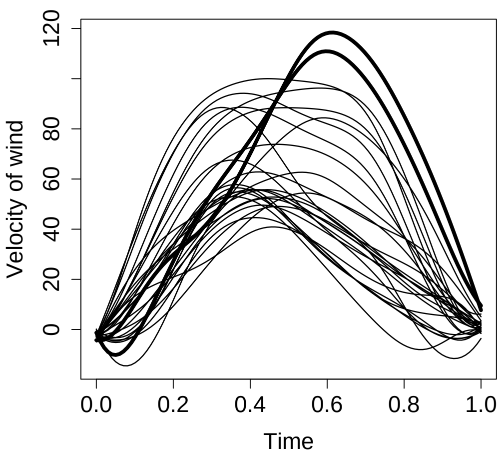

Figure 5: Result of the prediction of wind velocities.

6 Concluding remarks

We discussed the functional regression modeling procedure for functional predictors and a functional response. We proposed to estimate the model by the maximum penalized likelihood method to obtain more stable estimates. In order to select regularization parameters involved in the maximum penalized likelihood estimates, we derived model selection criteria by extending those of ordinary models. We applied the proposed modeling strategy to the analysis of weather and typhoon data, predicting response functions rather than scalars. Results show that our modeling strategy perform well in stability and prediction ability.

Future works reminds on constructing functional regression models that take the correlation among predictors into consideration. Furthermore, nonlinear functional regression model for functional predictors and functional responses are needed.

Appendix

A Converting discrete data to functional data

Since data are generally obtained discretely we need to explain these data as functions. We use a smoothing method via regularized basis expansions for converting raw data into functional data. In this section we only refer to the predictor, however, same is true of the response.

Suppose we have n observations $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$, where each $\boldsymbol{x}_\alpha$ are vectors of N_α observations $\{x_{\alpha 1}, \dots, x_{\alpha N_\alpha};\ \alpha = 1, \dots, n\}$ at $\{s_{\alpha 1}, \dots, s_{\alpha N_\alpha};\ s_{\alpha i} \in \mathcal{S} \subset \mathbb{R},\ i = 1, \dots, N_\alpha\}$. We assume that $x_{\alpha i}$s are given by adding Gaussian noises $\varepsilon_{\alpha i}$ to unknown smooth functions $u_\alpha(s)$ at $s_{\alpha i}$, that is,

$$x_{\alpha i} = u_\alpha(s_{\alpha i}) + \varepsilon_{\alpha i}, \quad i = 1, \dots, N_\alpha, \tag{27}$$

where $\varepsilon_{\alpha i}$ are independently normally distributed with mean 0 and variance $\sigma^2_{x\alpha}$.

We assume that $u_\alpha(s)$ are represented by the basis function expansion such as

$$u_\alpha(s) = \sum_{j=1}^{J} c_{\alpha j}\phi_j(s) = \boldsymbol{c}'_\alpha \boldsymbol{\phi}(s), \tag{28}$$

where $\boldsymbol{c}_\alpha = (c_{\alpha 1}, \dots, c_{\alpha J})'$ are vectors of coefficient parameters and $\boldsymbol{\phi}(s) = (\phi_1(s), \dots, \phi_J(s))'$ are vectors of basis functions. We assume that basis functions $\phi_j(s)$ $(j = 1, \dots, J)$ are Gaussian basis functions defined in (2). From these results the regression model (27) has a probability density function

$$f(x_{\alpha i}|s_{\alpha i}; \boldsymbol{c}_\alpha, \sigma^2_{x\alpha}) = \frac{1}{\sqrt{2\pi\sigma^2_{x\alpha}}} \exp\left\{ -\frac{(x_{\alpha i} - \boldsymbol{c}'_\alpha \boldsymbol{\phi}(s_{\alpha i}))^2}{2\sigma^2_{x\alpha}} \right\}. \tag{29}$$

The parameters $\boldsymbol{c}_\alpha$ and $\sigma^2_{x\alpha}$ are estimated by using regularization method, which maximizes a penalized log-likelihood function

$$l_{\zeta_\alpha}(\boldsymbol{c}_\alpha, \sigma^2_{x\alpha}) = \sum_{i=1}^{N_\alpha} \log f(x_{\alpha i}|s_{\alpha i}; \boldsymbol{c}_\alpha, \sigma^2_{x\alpha}) - \frac{N_\alpha \zeta_\alpha}{2} \boldsymbol{c}'_\alpha \Omega \boldsymbol{c}_\alpha, \tag{30}$$

where ζ_α are smoothing parameters which adjust the smoothness of the estimated function, and Ω is a $J \times J$ positive semi-definite matrix. The maximum penalized likelihood estimators $\hat{\boldsymbol{c}}_\alpha$ and $\hat{\sigma}^2_{x\alpha}$ are given by

$$\hat{\boldsymbol{c}}_\alpha = (\Phi'_\alpha \Phi_\alpha + N_\alpha \zeta_\alpha \hat{\sigma}^2_{x\alpha}\Omega)^{-1}\Phi'_\alpha \boldsymbol{x}_{(\alpha)}, \quad \hat{\sigma}^2_{x\alpha} = \frac{1}{N_\alpha}(\boldsymbol{x}_{(\alpha)} - \Phi_\alpha \hat{\boldsymbol{c}}_\alpha)'(\boldsymbol{x}_{(\alpha)} - \Phi_\alpha \hat{\boldsymbol{c}}_\alpha), \tag{31}$$

respectively, where $\Phi_\alpha = (\boldsymbol{\phi}(s_{\alpha 1}), \ldots, \boldsymbol{\phi}(s_{\alpha N_\alpha}))'$ and $\boldsymbol{x}_{(\alpha)} = (x_{\alpha 1}, \ldots, x_{\alpha N_\alpha})'$.

The maximum penalized likelihood estimates based on Gaussian basis functions depend on the regularization parameters ζ_α and the number of basis functions J. For the choice of these parameters some model selection criteria are used. Details are referred to Konishi and Kitagawa (2008). Selecting appropriate values of ζ_α and J, leading to appropriate estimates $\hat{u}_\alpha(s)$. Therefore we obtain functional data

$$x_\alpha(s) \equiv \hat{u}_\alpha(s) = \hat{\boldsymbol{c}}'_\alpha \boldsymbol{\phi}(s). \tag{32}$$

We use a set of functions $\{x_\alpha(s);\ s \in \mathcal{S},\ \alpha = 1, \ldots, n\}$ as data instead of observed data set $\{(s_{\alpha i}, x_{\alpha i});\ i = 1, \ldots, N_\alpha,\ \alpha = 1, \ldots, n\}$.

B Derivation of GBIC

We show the derivation of the model selection criterion GBIC (26) for evaluating the functional regression model estimated by the regularization method.

The penalized log-likelihood function (17) is rewritten as

$$\begin{aligned} l_\Lambda(B, \Sigma) &= \log\left\{ f(\boldsymbol{y}|\boldsymbol{\theta}) \exp\left[-\frac{n}{2}\mathrm{tr}\{B'(\Lambda_M \odot \Omega)B\} \right] \right\} \\ &= \log\left\{ f(\boldsymbol{y}|\boldsymbol{\theta}) \prod_{k=1}^K \exp\left[-\frac{n}{2}\boldsymbol{b}'_{(k)}(\Lambda_M \odot \Omega)\boldsymbol{b}_{(k)} \right] \right\}, \end{aligned} \tag{33}$$

where $\log f(\boldsymbol{y}|\boldsymbol{\theta}) = \sum_\alpha f(y_\alpha|\boldsymbol{x}_\alpha; \boldsymbol{\theta})$. We set the prior density of $\boldsymbol{\theta}$ as a product of K multivariate normal distribution, that is,

$$\pi(\boldsymbol{\theta}|\Lambda_M) = \prod_{k=1}^K \frac{n^{(p-q)/2}|\Lambda_M \odot \Omega|_+^{1/2}}{(2\pi)^{(p-q)/2}} \exp\left[-\frac{n}{2}\boldsymbol{b}'_{(k)}(\Lambda_M \odot \Omega)\boldsymbol{b}_{(k)} \right]. \tag{34}$$

Then the marginal likelihood of $\boldsymbol{y}$ given $\boldsymbol{\theta}$ with prior distribution (34) can be expressed as

$$\begin{aligned} p(\boldsymbol{y}|\Lambda_M) &= \int f(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\Lambda_M)d\boldsymbol{\theta} \\ &= \int \exp\left[n \times \frac{1}{n}\log\{f(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\Lambda_M)\} \right] d\boldsymbol{\theta} \\ &= \int \exp\{nq(\boldsymbol{\theta}|\Lambda_M)\}\, d\boldsymbol{\theta}, \end{aligned} \tag{35}$$

where $q(\boldsymbol{\theta}|\Lambda_M) = \log\{f(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\Lambda_M)\}/n$.

A Taylor series expansion of $q(\boldsymbol{\theta}|\Lambda_M)$ around $\hat{\boldsymbol{\theta}}$, the maximum penalized likelihood estimator of $\boldsymbol{\theta}$, is given by

$$q(\boldsymbol{\theta}|\Lambda_M) = q(\hat{\boldsymbol{\theta}}|\Lambda_M) - \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' R_\lambda(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + \cdots \tag{36}$$

since $\partial q(\hat{\boldsymbol{\theta}}|\Lambda_M)/\partial\boldsymbol{\theta} = \mathbf{0}$. Substituting (36) into (35), we obtain the following Laplace approximation

$$\begin{aligned}\int \exp\{nq(\boldsymbol{\theta}|\Lambda_M)\}\, d\boldsymbol{\theta} &= \int \exp\left[n\left\{q(\hat{\boldsymbol{\theta}}|\Lambda_M) - \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' R_\lambda(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + \cdots\right\}\right] d\boldsymbol{\theta} \\ &\approx \exp\left\{nq(\hat{\boldsymbol{\theta}}|\Lambda_M)\right\} \int \exp\left\{-\frac{n}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' R_\lambda(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\right\} d\boldsymbol{\theta} \\ &= \frac{(2\pi)^{(r+Kp)/2}}{n^{(r+Kp)/2}|R_\lambda(\hat{\boldsymbol{\theta}})|^{1/2}} \exp\left\{nq(\hat{\boldsymbol{\theta}}|\Lambda_M)\right\}. \end{aligned} \tag{37}$$

Therefore, the GBIC evaluating the multivariate functional regression model estimated by the maximum penalized likelihood method is given by

$$\begin{aligned}-2\log p(\boldsymbol{y}|\Lambda_M) =& -2\log\left\{\int f(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\Lambda_M) d\boldsymbol{\theta}\right\} \\ \approx& -2\sum_{\alpha=1}^{n} f(y_\alpha|\boldsymbol{x}_\alpha;\hat{\boldsymbol{\theta}}) + n\mathrm{tr}\{\hat{B}'(\Lambda_M \odot \Omega)\hat{B}\} + (r+Kq)\log n \\ & - (r+Kq)\log(2\pi) - \sum_{k=1}^{K} \log|\Lambda_M \odot \Omega|_+ + \log|R_\lambda(\hat{\boldsymbol{\theta}})|. \end{aligned} \tag{38}$$

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *2nd International Symposium on Information Theory* (Petrov, B. N. and Csaki, F. eds.), Akademiai Kiado, Budapest, 267–281.

Araki, Y., Konishi, S., Kawano, S. and Matsui, H. (2008). Functional Regression Modeling via Regularized Gaussian Basis Expansions. To appear in *Ann. Inst. Statist. Math.*

Cardot, H., Ferraty, F. and Sarda, P. (2003). Spline estimators for the functional linear model. *Stat. Sinica* **13**, 571–591.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross validation. *Numer. Math.* **31**, 317–403.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis.* Springer, New York.

Harezlak, J., Coull, B., Laird, N., Magari, S. and Christiani, D. (2007). Penalized solutions to functional regression problems. *Comput. Statist. Data Anal.* **51**, 4911–4925.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models.* Chapman and Hall, London.

Imoto, S. and Konishi, S. (2003). Selection of smoothing parameters in B-spline nonparametric regression models using information criteria, *Ann. Inst. Statist. Math* **55**, 671–687.

James, G. M. (2002). Generalized linear models with functional predictor variables. *J. Roy. Statist. Soc. Ser. B* **64**, 411–432.

Kawano, S. and Konishi, S. (2007). Nonlinear regression modeling via radial basis functions. *Bull. Inform. Cybern.* **39**, 83–96.

Konishi, S., Ando, T. and Imoto, S. (2004). Bayesian information criteria and smoothing parameter selection in radial basis function network, *Biometrika* **91**, 27–43.

Konishi, S. and Kitagawa, G. (1996). Generalised information criteria in model selection. *Biometrika* **83**, 875–890.

Konishi, S. and Kitagawa, G. (2008). *Information Criteria and Statistical Modeling.* Springer, New York.

Malfait, N. and Ramsay, J. O. (2003). The historical functional linear model. *Canad. J. Statist.* **31**, 115–128.

Matsui, H., Araki, Y. and Konishi, S. (2008). Multivariate regression modeling for functional data. *J. Data Sci.* **6**, 313–331.

Ramsay, J. O. and Dalzell, C. J. (1991). Some tools for functional data analysis (with Discussion). *J. Roy. Stat. Soc. Ser. B*, Series B **58**, 495–508.

Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis.* Springer, New York.

Ramsay, J.O. and Silverman, B.W. (2005). *Functional Data Analysis Second Edition.* Springer, New York.

Rossi, F., Delannay, N., Conan-Guez, B. and Verleysen, M. (2005). Representation of functional data in neural networks, *Neurocomputing* **64**, 183–210.

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

Shimokawa, M., Mizuta, M. and Sato, Y. (2000). An expansion of functional regression analysis. *J. Jap. App. Statist. (in Japanese)* **29**, 27–39.

Yamanishi, Y. and Tanaka, Y. (2003). Geographically weighted functional multiple regression analysis: a numerical investigation, *J. Jap. Soc. Comp. Statist*, **15**, 307–317.

Yao, F., Müller, H. G. and Wang, J. L. (2005). Functional linear regression analysis for longitudinal data. *Ann. Statist.* **33**, 2873–2903.

List of MI Preprint Series, Kyushu University

The Grobal COE Program
Math-for-Industry Education & Research Hub

MI

MI2008-10 Sangyeol LEE & Hiroki MASUDA
Jarque-Bera Normality Test for the Driving Lévy Process of a Discretely Observed Univariate SDE

MI2008-11 Hiroyuki CHIHARA & Eiji ONODERA
A third order dispersive flow for closed curves into almost Hermitian manifolds

MI2008-12 Takehiko KINOSHITA, Kouji HASHIMOTO and Mitsuhiro T. NAKAO
On the L^2 a priori error estimates to the finite element solution of elliptic problems with singular adjoint operator

MI2008-13 Jacques FARAUT and Masato WAKAYAMA
Hermitian symmetric spaces of tube type and multivariate Meixner-Pollaczek polynomials

MI2008-14 Takashi NAKAMURA
Riemann zeta-values, Euler polynomials and the best constant of Sobolev inequality

MI2008-15 Takashi NAKAMURA
Some topics related to Hurwitz-Lerch zeta functions

MI2009-1 Yasuhide FUKUMOTO
Global time evolution of viscous vortex rings

MI2009-2 Hidetoshi MATSUI & Sadanori KONISHI
Regularized functional regression modeling for functional response and predictors