STUDIES ON METHOD WITH INFLUENCE FUNCTION FOR MODEL SELECTION IN *B*-SPLINE SURFACE APPROXIMATION

MARCH, 2014

Bao Hongmei

Graduate School of Environmental Science
(Doctor Course)
OKAYAMA UNIVERSITY

Contents

Abstract

At the model selection the effective method requires much cost of time. The analysis of bivariate B-spline model with penalized term has many difficulties. It has many factors and parameters such that the number of the knots, the locations of those knots, number of B-spline functions and the value of the smoothing parameter of the penalized term. For the determination of the model we have to compare a large amount of the combinations of those parameters. Various information criteria are considered and the cross validation (CV) criterion is excellent but it requires a large amount of computational costs.

Various alternative schemes are considered to reduce its computation. Modified generalized information criterion (mGIC) is one of those alternative schemes. In this criterion an influence function is used to estimate the parameters of the models. The influence function is related to the first term of a Taylor expansion. By the numerical simulation we studied the effect of an influence function. But this information criterion could not determine the optimal model.

A new method generalized cross validation with influence function (GCV_{IF})is proposed. For the verification of GCV_{IF} theoretical proof and the computational results are shown.

We have to decide the value of the optimal smoother parameter which minimize the value of the evaluation function. Among the models which have suitable parameters respectively we have to choose the best model by using the information criteria such as CV or generalized CV with influence function (GCV_{IF}). However the method of

GCV_{IF} is not practical necessarily, because it requires the calculation of the inverse matrix of hat matrix and the influence function. Those calculations take a large cost of time when n increases. The efficient scheme which will take small amount of time is required.

On the other hand the parameters which we have to decide are the coefficients of the spline functions and the total number of knots and those positions and the smoother parameter of penalized term. The range of the total number of knots is decided by the total number of sample points. The range of the positions of the knots are decided by the area of the surface. But the estimation of the range of the value of the smoother parameter is difficult. So we have to estimate it quite roughly. We propose an effective method to estimate the range of the smoother parameter and consequently we can obtain more accurate value of its parameter. We can reduce the calculation time which does not contribute the selection of the optimal model. We can decide more accurate smoother parameter by a small amount of time.

The application of B-spline surface to the estimation of lake bottom topography is described. Using the analysis of bivariate B-spline the shape of the lake bottom is approximated. According to the validity of the estimation by bivariate B-spline function the method is applied to the actual data of the lake depth. A survey over the water area has more difficulties than that on the land, and the measurement data are distributed quite irregularly. The locations of the measured data don't exist regularly over the lake. Those locations were distributed along with the wake of the boat on which the sample data were collected. The density of the data is quite high in some small region and quite low in another wide region. Based on such irregular data we tried

the statistical estimation. The regularized term with a penalty coefficient make the proper approximation of the parameters of B-spline functions. There are many factors such that the number of the knots, the locations of those knots, the number of B-spline functions and the coefficient of penalized term. Appropriate information criterion which has sufficient accuracy and small amount of computation is applied to the determination of the optimal model.

Chapter 1

Introduction

Both parametric statistics and nonparametric statistics, in the establishment of the relationship between the response variables and the covariate variables, there is the problem of model selection, it is an important part for statistical modeling. One main purpose of model selection is to choose the true distribution.

In this thesis, the refined cross-validation value GCV_{IF} for the model selection is proposed, for the approximation of experimental data, spline function is smooth and useful because of its less oscillation. Its dominance becomes larger according to the appropriate locations of knots. In this thesis the approximation of the two dimensional surface by bivariate B-splines is described. It has some additional difficulties than univariate spline function.

Determination of smooth coefficients of B-splines, we use the maximum penalized likelihood estimator (MPLE; Good and Gaskins 1971; Green and Silverman 1994). Among some methods for the penalized term we choose the method of integration as the most favorable one.

An AIC-type criterion which is the approximation of Kullback-Leibler divergence for the MPLE is a generalized information criterion (GIC[16]) which forms the empirical log-likelihood with the correction term for the bias, derived analytically with the influence function. The GIC can evaluate the models not only with MPLE but also with a robust estimator, maximum weighted likelihood estimator, etc. Cross-validation (CV [25]) is applicable to choose the value of an optimal parameter in the maximum penalized likelihood method. The CV requires less analytic calculations than the GIC, although the computational cost for the CV is much higher than the GIC.

To overcome computational costs, the mGIC [28] is considered

which utilize influence function. The first order influence function is useful for its small costs of time. On the other hand the second order influence function has too much costs of time compared with its small profits. But mGIC is not sufficient to determine the optimal parameters. For more accuracy we use the generalized CV (GCV[17]) and we proposed GCV_{IF}, which is an improved GCV with the influence function. It is better for the model selection than CV, AIC, GIC and mGIC. GCV_{IF} is the criterion that includes the residual sum of squares, the number of sample and the number of parameters in the model. It is more stable and distinguishable than CV, GIC and mGIC. The computational result shows the excellence of our improved scheme GCV_{IF}.

The smoothing spline surfaces are often used to estimate a three-dimens-ional shape of the surface. When we use the regularization method with the penalized term the smoothing parameter is most important. Too large parameter gives a too flat surface and too small one causes overfitting and gives a surface without fluency. The value of the appropriate parameter varies according to the shape of the surface to be estimated. We have to search the appropriate range of the value of the smoothing parameter. When we have decided the set of the parameter and coefficients we evaluate them by the information criterion. But it will take a long time to calculate, when we have chosen CV. For the many sets of knots and many values of parameters we have to calculate the values of CV. However most of them don't contribute to the determination of the optimal model. We introduce a scheme which use as small sets as possible. Moreover we can obtain more accurate value of the smoother parameter.

At first we start with small sets of knots and a set of rough values of the parameters. After calculating the values of CV for them we obtain a value of the smoother parameter by using spline interpolation. This value is only estimated by interpolation, so it is not accurate necessarily. Based on this parameter we estimate the values of CV for the many sets of knots. The model which has the smallest value of CV is the optimal set of knots. We can obtain the same optimal set of knots from various estimated values of the smoother parameter. It is verified by the numerical calculation. Finally we determine the optimal value of the smoother parameter for the optimal set of knots. In this study we approximate the topography of the lake bottom with our statistical method. The lake is Kojima Lake which is located Okayama prefecture in Japan. Kojima lake is separated by the bank from Kojima bay and turned into a freshwater lake. The water quality of the closed water area like this lake tends to worsen because of the sedimentation or the pollutant from the upper stream. For the improvement of the water quality the dredging is tried and that requires the detailed depth data of the lake. Compared with land, the detailed survey of the lake depth is difficult and we applied statistic method. Based on the data measured from September 2010 to January 2011, we made the estimation by using B-splines. To make optimal model selection, various information criteria are devised. When there is a large number of models, CV is difficult to use for its computational cost. The GCV_{IF} is adopted because we can obtain the almost same information as CV and it has a small amount of computational cost. Furthermore, in order to high accuracy, the technique of using the influence function which we have proposed recently [2] is applied. We are able to obtain

by this method the optimal model which can approximate the smooth topography of the lake bottom. At first we applied CV and GCV_{IF} for the estimation of the selected two subdomains of the lake. After the evaluation of the methods we approximate the topography of the whole domain by GCV_{IF} and selected values of β.

Chapter 2

Statistical Models

2.1 Modeling

A statistical model is a formalization of relationships between variables in the form of mathematical equations. A statistical model describes how one or more random variables are related to one or more other variables. The model is statistical as the variables are not deterministically but stochastically related. In mathematical terms, a statistical model is frequently thought of as a pair where is the set of possible observations and the set of possible probability distributions on. It is assumed that there is a distinct element of which generates the observed data. Statistical inference enables us to make statements about which element(s) of this set are likely to be the true one.

If we assume that the observations $\boldsymbol{x} = \{x_1, x_2, \ldots, x_n\}$ are generated from the distribution function $G(x)$, then $G(x)$ is referred to as the true distribution, or the true model. On the other hand, the distribution function $F(x)$ used to approximate the true distribution is referred to as a model and is assumed to have either a density function or a probability function $f(x)$. If a model is specified by p-dimension parameters $\boldsymbol{\theta}$, then the model can be written as $f(x|\boldsymbol{\theta})$. If the parameters are represented as a point in the set $\Theta \subset \mathrm{R}^p$, then $\{f(x|\boldsymbol{\theta}); \boldsymbol{\theta} \in \Theta\}$ is called a parametric family of probability distributions or models.

An estimated model $f(x|\hat{\boldsymbol{\theta}})$ obtained by replacing an unknown parameter $\boldsymbol{\theta}$ with an estimator $\hat{\boldsymbol{\theta}}$ is referred to as a statistical model. The process of constructing a model that appropriately represents some phenomenon is referred to as modeling. Hoverer, setting up an appropriate family of probability models prior to estimating the parameters is of greater importance.

2.2 Regression Models

The regression model is used to model the relationship between a response variable y and several explanatory variables $\boldsymbol{x} = \{x_1, x_2, \ldots, x_p\}^T$. This is equivalent to assuming that the probability distribution of the response variable y varies depending on the explanatory varibales $\boldsymbol{x}$ and that a conditional distribution is given in the form of $f(y|\boldsymbol{x})$.

Let $\{(y_\alpha, \boldsymbol{x}_\alpha); \alpha = 1, 2, \ldots, n\}$ be n sets of data obtained in terms of the response variable y and p explanatory variables $\boldsymbol{x}$. Then the model

$$y_\alpha = u(\boldsymbol{x}_\alpha) + \varepsilon_\alpha, \qquad \alpha = 1, 2, \ldots, n, \tag{2.1}$$

of the observed data is called a regression model, where $u(\boldsymbol{x})$ is a function of the explanatory variables $\boldsymbol{x}$, and the error terms or noise ε_α are assumed to be independently distributed with mean $E[\varepsilon_\alpha] = 0$ and variance $V(\varepsilon_\alpha) = \sigma^2$. We often assume that the noise ε_α follows the normal distribution $N(0, \sigma^2)$. In such a case, y_α has the normal distribution $N(u(\boldsymbol{x}_\alpha), \sigma^2)$ with mean $u(\boldsymbol{x}_\alpha)$ and variance σ^2, and its density function is given by

$$f(y_\alpha|\boldsymbol{x}_\alpha) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_\alpha - u(\boldsymbol{x}_\alpha))^2}{2\sigma^2}\right\}, \qquad \alpha = 1, 2, \ldots, n, \tag{2.2}$$

The following models are used as regression functions that approximate the mean structure: (1) linear regression, (ii) polynomial regression, (iii) natural cubic splines given by piecewise polynomials Green and Silverman(1994, p.12)], (iv) B-splines[de Boor (1978), Imoto(2001), Imoto and Konishi(2003)], (v) kernel functions [Simonoff (1996)], adn (vi) neural networksBishop(1995), Ripley(1996)].

Chapter 3

Generalized CV with Influence Function

3.1 Influence Function

The general definition of the influence function is as follows. Its suitably normed limiting influence on the value of an estimate or test statistic $T(\hat{G})$ can be expressed as

$$T^{(1)}(x, G) = \lim_{\varepsilon \to 0} \frac{T((1-\varepsilon)G + \varepsilon\delta_x) - T(G)}{\varepsilon}, \tag{3.1}$$

where δ_x denotes the pointmass 1 at x. The above quantity, considered as a function of x, was introduced [12],[13] under the name *influence function ($T^{(1)}$)*, and is arguably the most useful heuristic tool of robust statistics. If T is sufficiently regular, it can be linearized near G in terms of the influence function: if H is near G , then the leading terms of a Taylor expansion are

$$T(H) = T(G) + \int T^{(1)}(x, G)[dH(x) - dG(x)] + \cdots, \tag{3.2}$$

We have

$$\int T^{(1)}(x, G)dG(x) = 0, \tag{3.3}$$

and, if we substitute the empirical distribution $\hat{G}$, for H in the equation (3.2), we obtain

$$\begin{aligned} \sqrt{n}(T(\hat{G}) - T(G)) &= \sqrt{n} \int T^{(1)}(x, G)d\hat{G}(x) + \cdots \\ &= \frac{1}{\sqrt{n}} \sum T^{(1)}(x_i, G) + \cdots . \end{aligned} \tag{3.4}$$

By the central limit theorem, the leading term on the right-hand side is asymptotically normal with mean 0, if the x_i are independent with common distribution G. The remaining terms are asymptotically negligible and $\sqrt{n}[T(\hat{G}) - T(G)]$ is asymptotically normal with mean 0

and variance

$$A(G,T) = \int T^{(1)}(x,G)^2 dG(x). \tag{3.5}$$

About distributions G, H, when we let

$$h(\varepsilon) = T((1-\varepsilon)G + \varepsilon H), \quad 0 \le \varepsilon \le 1, \tag{3.6}$$

The ith order derivative of functional $T(\cdot)$ at $(z_1, z_2, \cdots, z_i, G)$ as the symmetric function $T^{(i)}(z_1, z_2, \cdots, z_i, G)$ will be defined to satisfy the following equation with respect to any distribution function H [31],[33]

$$h^{(i)}(0) = \int \cdots \int T^{(i)}(z_1, z_2, \cdots, z_i, G) \prod_{j=1}^{i} d\{H(z_j) - G(z_j)\}. \tag{3.7}$$

To guarantee the uniqueness of the derivatives the next condition is imposed

$$\int T^{(i)}(z_1, z_2, \cdots, z_i, G) dG(z_k) = 0, \quad 1 \le k \le i. \tag{3.8}$$

Then $d\{H(z_j) - G(z_j)\}$ in the equation (3.7) can be replaced with $dH(z_j)$.

The Taylor series of $h(\varepsilon)$ around $\varepsilon = 0$ is

$$h(\varepsilon) = h(0) + \varepsilon h'(0) + \frac{1}{2}\varepsilon^2 h''(0) + \cdots. \tag{3.9}$$

If we set $\varepsilon = 1$ in the equation(3.9) formally then $h(0) = T(G)$, $h(1) = T(H)$,

$$\begin{aligned} T(H) = T(G) \;&+\; \int T^{(1)}(z_1; G)) dG(z_1) \\ &+\; \frac{1}{2} \iint T^{(2)}(z_1, z_2; G)) dG(z_1) dG(z_2) + \cdots \end{aligned} \tag{3.10}$$

3.2 Improved Generalized CV Criterion GCV_{IF}

The Kullbak-Leibler divergence KL measures the distance between the true probability density function $p(x)$ and estimated probability density function $q(x)$ as follows

$$\begin{aligned} \mathrm{KL}(p;q) &= \int p(x) \log \frac{p(x)}{q(x)} dx \\ &= \int p(x) \log p(x) dx - \int p(x) \log q(x) dx. \end{aligned} \tag{3.11}$$

This divergence is nonnegative and is equal to zero if and only if $p(x) = q(x)$. But this value includes the unknown function $p(x)$ we can only estimate its value from the observed samples. The first term $\int p(x) \log p(x) dx$ is constant and we only have to estimate the sencond term $-\int p(x) \log q(x) dx$. The negative log-likelyhood is an approximation of KL divergence and it is asymptoticaly equivalent to KL divergence according to the law of large numbers as follows

$$-\frac{1}{n} \sum_{\alpha=1}^{n} \log q(x_\alpha) \to -\int p(x) \log q(x) dx. \tag{3.12}$$

The property of the leave-one-out cross-validation (LOOCV) is as foloows

$$E[\mathrm{LOOCV}] = E[-\int p(x) \log q^{(-\alpha)}(x) dx] \tag{3.13}$$

where $q^{(-\alpha)}(x)$ is the probablity density function of the distribution without the α-th data point.

We considered an information criterion GCV_{IF} which is a generalized cross-validation with the influence function. From n observations the α-th data point $(z_\alpha, \boldsymbol{x}_\alpha)$ is removed and the parameter vector

$\boldsymbol{\theta} = (\boldsymbol{w}', \sigma^2)'$ is estimated based on the remaining $n-1$ observations. We denote the parameter as $\hat{\boldsymbol{\theta}}^{(-\alpha)} = (\hat{\boldsymbol{w}}^{(-\alpha)\prime}, \hat{\sigma}^{2(-\alpha)})'$. The corresponding estimated regression function is denoted as $\hat{u}^{(-\alpha)}(\boldsymbol{x})$. We use the log-likelihood for Cross-Validation(IC_{CV}) as

$$\begin{aligned} \mathrm{IC}_{CV} &= -2\sum_{\alpha=1}^{n} \log(f(\boldsymbol{x}_\alpha, \boldsymbol{\theta}^{(-\alpha)})) \\ &= \sum_{\alpha=1}^{n} \left\{ \log(2\pi\hat{\sigma}^{2(-\alpha)}) + \frac{(z_\alpha - \hat{u}^{(-\alpha)})^2}{\hat{\sigma}^{2(-\alpha)}} \right\}. \end{aligned} \tag{3.14}$$

This is asymptotically equivalent to AIC (Akaike Information criterion)-type criteria such as AIC or BIC (Bayesian Information criterion) and so on [1][24][34]. Minimizing the equation(3.14) is the method of selecting optimal model. Various alternative schemes are considered for the reduction of its computational costs.

And another scheme is called generalized CV (GCV)[17] which estimate the value of $u^{(-\alpha)}(\boldsymbol{x}_\alpha)$ directly as follows

$$z_\alpha - \hat{u}^{(-\alpha)}(\boldsymbol{x}_\alpha) = \frac{z_\alpha - \hat{u}(\boldsymbol{x}_\alpha)}{1 - h_{\alpha\alpha}}, \tag{3.15}$$

where the $h_{\alpha\alpha}$ is the $(\alpha, \alpha)^{\text{th}}$ component of the smoother matrix H. The matrix H transforms observed data $\boldsymbol{z}$ to predicted values $\hat{\boldsymbol{z}}$ where H does not depend on the data z, and it is referred to as a hat matrix or is called as smoother matrix. Then in cross-validation, the estimation process performed n times by removing observations one by one is not needed, and thus the amount of computation required can be reduced substantially. Then the generalized cross validation with influence function GCV_{IF} is calculated by

$$\mathrm{GCV}_{IF} = \sum_{\alpha=1}^{n} \left\{ \log(2\pi\hat{\sigma}^{2(-\alpha)}) + \left[\frac{z_\alpha - \hat{u}(\boldsymbol{x}_\alpha)}{\hat{\sigma}^{(-\alpha)}(1 - \frac{1}{n}\mathrm{tr}H)} \right]^2 \right\}, \tag{3.16}$$

where $h_{\alpha\alpha}$ is replaced with $\frac{1}{n}\mathrm{tr}H$ which is the average of it and $\mathrm{tr}(H)$ is called the effective number of parameters. The estimation $\hat{\sigma}^{2(-\alpha)}$ is approximated by the influence function $\boldsymbol{T}^{(1)}(z_\alpha; \hat{G})$ as follows [17]

$$\hat{\boldsymbol{\theta}}^{(-\alpha)} \approx \hat{\boldsymbol{\theta}} - \frac{1}{n}\boldsymbol{T}^{(1)}(z_\alpha; \hat{G}). \tag{3.17}$$

3.3 Method of Regularization

For the nonlinear statistical modeling the maximum penalized likelihood methods are often used [9],[10],[11]. Suppose that we have n observations $\{(z_\alpha, \boldsymbol{x}_\alpha);\ \alpha = 1, \cdots, n\}$, where z_α are the response variables generated from unknown true distribution $G(z|\boldsymbol{x})$ having a probability density $g(z|\boldsymbol{x})$ and $\boldsymbol{x}_\alpha$ are the vectors of explanatory variables. We estimate $\boldsymbol{w}$ which is a vector consisting of the unknown parameters and determines the model $z = u(\boldsymbol{x}|\boldsymbol{w})$. Let $f(z_\alpha|\boldsymbol{x}_\alpha; \boldsymbol{\theta})$ be a specified parametric model, where $\boldsymbol{\theta}$ is a vector of unknown parameters included in the model. The regression model with Gaussian noise is denoted as

$$z_\alpha = u(\boldsymbol{x}_\alpha|\boldsymbol{w}) + \varepsilon_\alpha, \quad \varepsilon_\alpha \sim N(0, \sigma^2), \quad \alpha = 1, \cdots, n \tag{3.18}$$

$$f(z_\alpha|\boldsymbol{x}_\alpha; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\{z_\alpha - u(\boldsymbol{x}_\alpha; \boldsymbol{w})\}^2}{2\sigma^2}\right], \tag{3.19}$$

where $\boldsymbol{\theta} = (\boldsymbol{w}', \sigma^2)'$. The parameter will be determined by the maximization of the penalized log-likelihood function expressed as

$$\ell_\lambda(\boldsymbol{\theta}) = \sum_{\alpha=1}^{n} \log f(z_\alpha|\boldsymbol{x}_\alpha; \boldsymbol{\theta}) - \frac{n}{2}\lambda H(\boldsymbol{w}). \tag{3.20}$$

As the regularized term or penalized terms $H(\boldsymbol{w})$ with an m-dimensional parameter vector $\boldsymbol{w}$, various types are used depending on the dimen-

sion of explanatory variables or the purpose of the analysis. For example, the discrete approximation of the integration of a second-order derivative, finite differences of the unknown parameters and sum of squares of w_i are used and those are $H_1(\boldsymbol{w}) = \frac{1}{n}\sum_{\alpha=1}^{n}\sum_{i=1}^{d}\left\{\frac{\partial^2 u(\boldsymbol{x}_\alpha|\boldsymbol{w})}{\partial x_i^2}\right\}^2$, $H_2(\boldsymbol{w}) = \sum_{i=k+1}^{m}(\Delta^k w_i)^2$ and $H_3(\boldsymbol{w}) = \sum_{i=1}^{m} w_i^2$. For the three dimensional approximation we use [30]

$$H(\boldsymbol{w}) = \iint \left\{ \left(\frac{\partial^2 u}{\partial x^2}\right)^2 + \left(\frac{\partial^2 u}{\partial y^2}\right)^2 \right\} dxdy, \tag{3.21}$$

and it is represented in the quadratic form

$$H(\boldsymbol{w}) = \boldsymbol{w}' K \boldsymbol{w}. \tag{3.22}$$

Therefore the equation(3.20) will be

$$\ell_\lambda(\boldsymbol{\theta}) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(\boldsymbol{z} - B\boldsymbol{w})'(\boldsymbol{z} - B\boldsymbol{w}) - \frac{n}{2}\lambda \boldsymbol{w}' K \boldsymbol{w}, \tag{3.23}$$

where $\boldsymbol{z} = (z_1, \cdots, z_n)'$, $u(\boldsymbol{x}_\alpha|\boldsymbol{w}) = \boldsymbol{w}'\boldsymbol{b}(\boldsymbol{x}_\alpha)$ and B is an $n \times m$ matrix composed of the basis functions as

$$B = (\boldsymbol{b}(\boldsymbol{x}_1)', \cdots, \boldsymbol{b}(\boldsymbol{x}_n)')'. \tag{3.24}$$

Differentiating $\ell_\lambda(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ and setting the result equal to zero and we obtain the solution of them. Then the estimation of parameters are

$$\begin{aligned} \hat{\boldsymbol{w}} &= (B'B + n\lambda\hat{\sigma}^2 K)^{-1} B'\boldsymbol{z}, \\ \hat{\sigma}^2 &= \frac{1}{n}(\boldsymbol{z} - B\hat{\boldsymbol{w}})'(\boldsymbol{z} - B\hat{\boldsymbol{w}}). \end{aligned} \tag{3.25}$$

At first we set the constant value of $\beta = \lambda\hat{\sigma}^2$ and determine $\hat{\boldsymbol{w}}$ for a given value of β. After we obtain the variance estimator $\hat{\sigma}^2$ then we can obtain the smoothing parameter $\lambda = \beta/\hat{\sigma}^2$.

3.3.1 B-Splines

We consist B-spline function $M_{m,i}(x)$ of required degree $r-1$ (order r) by the algorithm of de Boor-Cox[6],[7],[8]. This calculation can be started by the first step

$$M_{1,j}(x) = \begin{cases} (\xi_j - \xi_{j-1})^{-1} & (\xi_{j-1} \leq x < \xi_j) \\ 0 & (\text{otherwise}) \end{cases}, \tag{3.26}$$

and the successive recurrence formula is below

$$M_{r,j}(x) = \frac{(x - \xi_{j-r})M_{r-1,j-1}(x) + (\xi_j - x)M_{r-1,j}(x)}{\xi_j - \xi_{j-r}}, \tag{3.27}$$

where $\{\xi_k\}, k = 1-r, \cdots, n+r$ are the knots and n is the total number of intervals for the approximation. The univariate spline functions are shown in Figure 3.3.1, where $\xi_{-3}=\xi_{-2}=\xi_{-1}=\xi_0=0$, $\xi_1=1$, $\xi_2=2$, $\xi_3=3$, $\xi_4=\xi_5=\xi_6=\xi_7=4$. For the adequate approximation the selection of the knots is quite important. The division by equal intervals cannot provide the best approximation always.

We set the approximation for the three dimensional surface as

$$u(x,y) = \sum_{i=1}^{p_1}\sum_{j=1}^{p_2} w_{ij}M_i(x)N_j(y), \tag{3.28}$$

where p_1, p_2 is the total number of basis B-splines $\{M_i(x)\}, \{N_j(y)\}$ respectively. And these functions have the support $[\xi_{i-r}, \xi_i), [\eta_{j-r}, \eta_j)$ for x, y direction respectively. The shape of the three dimensional B-splines are shown in Figure 3.2. The left figure shows a one function with $\xi_i = (i-1)\times 10, \eta_i = (i-1)\times 10, i = 1, 2, \cdots, 5$. The right figure shows four functions with $p_1 = p_2 = 2, \xi_i = (i-1)\times 10, \eta_i =$

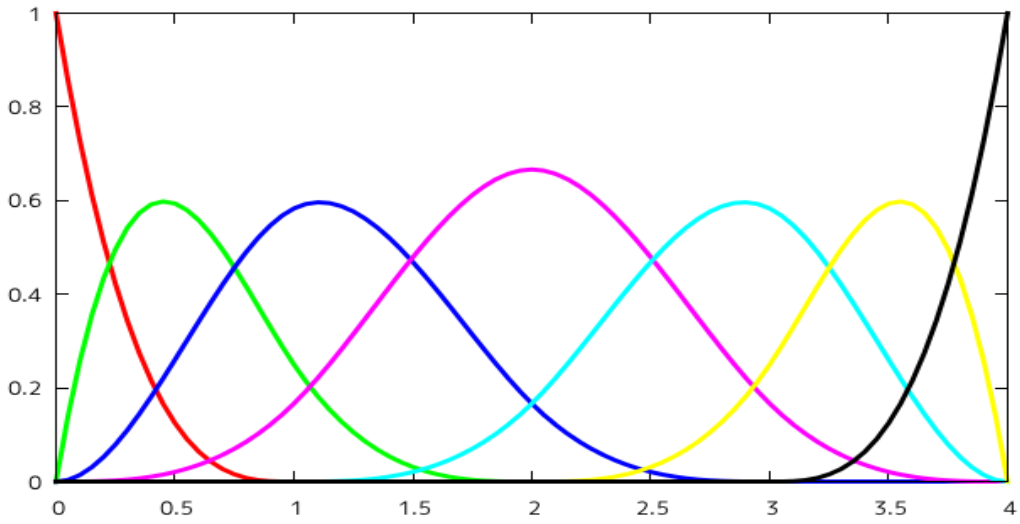


Figure 3.1: Spline functions (order four)

$(i-1)\times 10, i = 1, 2, \cdots, 6$. We have to satisfy the Schoenberg-Whitney condition[23]. Because if there is no sample point in the domain $\{(x,y)|\,\xi_{i-r} \le x < \xi_i, \eta_{j-r} \le y < \eta_j\}$, then we cannot determine the parameter w_{ij}.

In the equation of integration (3.21)

$$\left(\frac{\partial^2 u}{\partial x^2}\right)^2 = \left(\sum_{i=1}^{p_1}\sum_{j=1}^{p_2} w_{ij}\frac{d^2 M_i(x)}{dx^2}N_j(y)\right)^2, \tag{3.29}$$

$$\left(\frac{\partial^2 u}{\partial y^2}\right)^2 = \left(\sum_{i=1}^{p_1}\sum_{j=1}^{p_2} w_{ij}M_i(x)\frac{d^2 N_j(y)}{dy^2}\right)^2. \tag{3.30}$$

We set $w_{ij} = \tilde{w}_k, i = i_k, j = j_k, p_1 p_2 = m, \frac{d^2 M_i(x)}{dx^2}N_j(y) = \tilde{B}_{1,k}$, $M_i(x)\frac{d^2 N_j(y)}{dy^2} = \tilde{B}_{2,k}$.

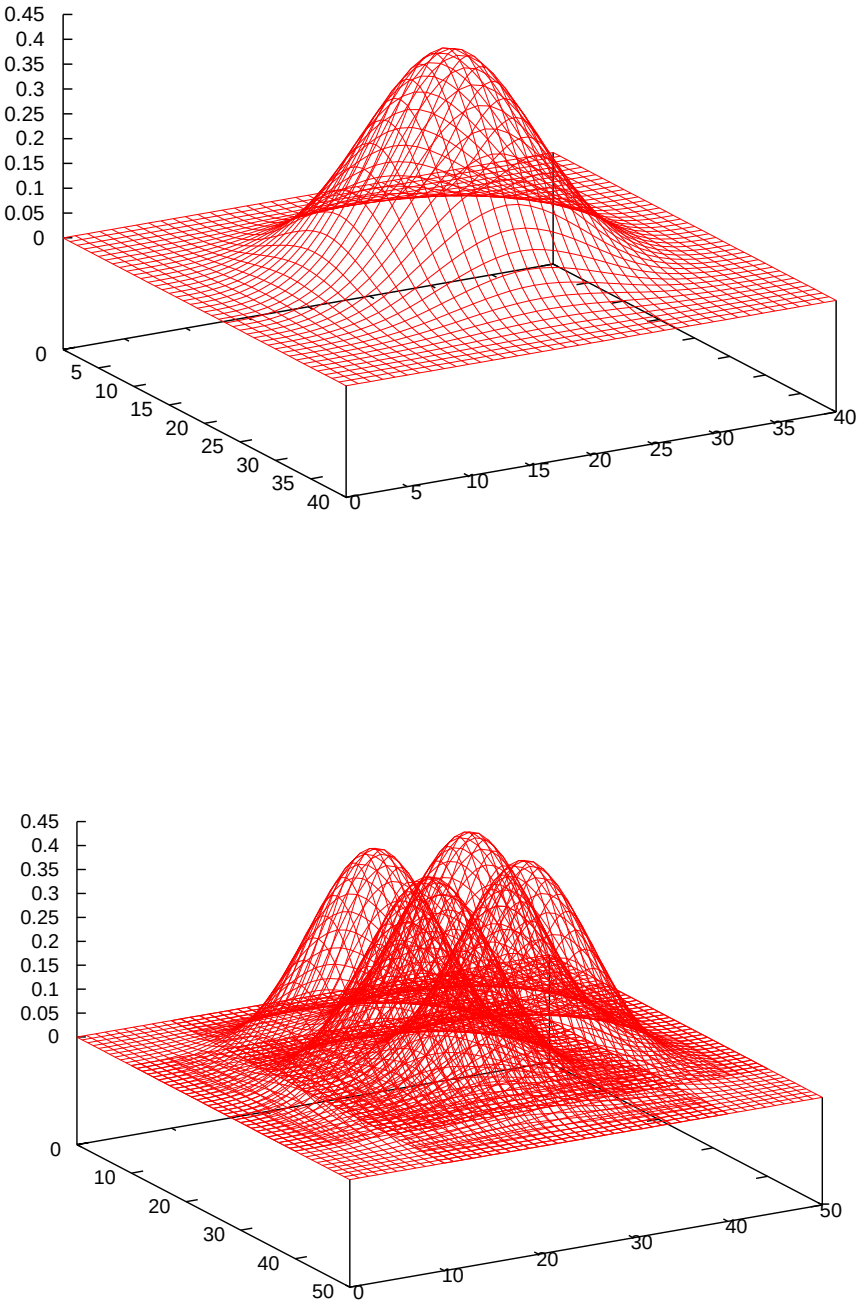


Figure 3.2: Three dimensional spline functions (order four)

Then the equation(3.29) (3.30) can be rewritten as

$$\left(\frac{\partial^2 u}{\partial x^2}\right)^2 = \left(\sum_{k=1}^{m} \tilde{w}_k \tilde{B}_{1,k}\right)^2, \tag{3.31}$$

$$\left(\frac{\partial^2 u}{\partial y^2}\right)^2 = \left(\sum_{k=1}^{m} \tilde{w}_k \tilde{B}_{2,k}\right)^2. \tag{3.32}$$

And the integration becomes below

$$\begin{aligned} &\iint \tilde{B}_{1,p}\tilde{B}_{1,q}dxdy = \\ &\int \frac{d^2 M_{i_p}(x)}{dx^2}\frac{d^2 M_{i_q}(x)}{dx^2}dx \int N_{j_p}(y)N_{j_q}(y)dy, \end{aligned} \tag{3.33}$$

$$\begin{aligned} &\iint \tilde{B}_{2,p}\tilde{B}_{2,q}dxdy = \\ &\int M_{i_p}(x)M_{i_q}(x)dx \int \frac{d^2 N_{j_p}(y)}{dy^2}\frac{d^2 N_{j_q}(y)}{dy^2}dy. \end{aligned} \tag{3.34}$$

The sum of equation(3.33) and (3.34) will be K_{pq} which is the component of $m \times m$ nonnegative matrix K that is represented in the equation (3.22). Therefore when we set $b_{ij}(x_\alpha, y_\alpha) = M_i(x_\alpha)N_j(y_\alpha)$, the equation(3.20) will be

$$\begin{aligned} \ell_\lambda(\boldsymbol{\theta}) &= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{\alpha=1}^{n}\{z_\alpha - \boldsymbol{w'b}(x_\alpha, y_\alpha)\}^2 - \frac{n}{2}\lambda \boldsymbol{w}' K \boldsymbol{w} \\ &= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(\boldsymbol{z} - B\boldsymbol{w})'(\boldsymbol{z} - B\boldsymbol{w}) - \frac{n}{2}\lambda \boldsymbol{w}' K \boldsymbol{w}, \end{aligned} \tag{3.35}$$

3.3.2 The Empirical Influence Function

When we differentiate this equation with respect to ε, and set $\varepsilon = 0$ then

$$\int \boldsymbol{\psi}(z, \boldsymbol{T}(G)) d\{\delta_x(z) - G(z)\} + \int \frac{\partial \boldsymbol{\psi}(z, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{T}(G)} dG \cdot \frac{\partial}{\partial \varepsilon}\{\boldsymbol{T}(H_\varepsilon)\}\bigg|_{\varepsilon=0} = \mathbf{0}, \tag{3.36}$$

where,$H_\varepsilon(z) = (1-\varepsilon)G(z) + \varepsilon\delta_x(z)$. Therefore the vector of the influence function is given by

$$\frac{\partial}{\partial \varepsilon}\{\boldsymbol{T}(H_\varepsilon)\}\bigg|_{\varepsilon=0} \equiv \boldsymbol{T}^{(1)}(x; \hat{G}). \tag{3.37}$$

Here when we set

$$R(\boldsymbol{\psi}, \hat{G}) = -\int \frac{\partial \boldsymbol{\psi}(z, \boldsymbol{\theta})'}{\partial \boldsymbol{\theta}} dG\bigg|_{\boldsymbol{\theta}=\boldsymbol{T}(\hat{G})}, \tag{3.38}$$

then

$$\boldsymbol{T}^{(1)}(z, \hat{G}) = R(\boldsymbol{\psi}, G)^{-1}\boldsymbol{\psi}(z, \boldsymbol{T}(\hat{G})). \tag{3.39}$$

3.3.3 Higher Order Empirical Influence Function

We can denote equations(3.25) as follows

$$\sum_{\alpha=1}^{n} \psi_i(\boldsymbol{x}_\alpha; \boldsymbol{\theta}) = 0 \quad (i = 1, 2, \cdots, p), p = p_1 p_2 + 1, \tag{3.40}$$

where $\boldsymbol{\theta} = (\boldsymbol{w}', \sigma^2)'$. When we denote $\boldsymbol{\psi} = (\psi_1, \cdots, \psi_p)'$ the solution $\hat{\boldsymbol{\theta}}$ of the equation(3.40) is given by $\hat{\boldsymbol{\theta}} = \boldsymbol{T}(\hat{G})$ which is the vector of functional with degree p defined with distribution G as follows

$$\int \boldsymbol{\psi}(x, \boldsymbol{T}(G)) dG(x) = \mathbf{0}. \tag{3.41}$$

Replacing the distribution G with $(1-\varepsilon)G + \varepsilon\delta_x$ we obtain

$$\int \boldsymbol{\psi}(z, \boldsymbol{T}((1-\varepsilon)G + \varepsilon\delta_x))d\{(1-\varepsilon)G(z) + \varepsilon\delta_x(z)\} = \mathbf{0}. \quad (3.42)$$

For the higher order influence function we differentiate the equation (3.42) with respect to ε twice and let $\varepsilon = 0$ then we obtain below

$$\begin{aligned} 2\int \frac{\partial \boldsymbol{\psi}(z, \boldsymbol{T}(G))'}{\partial \boldsymbol{\theta}} d\{\delta_x(z) - G(z)\} \cdot \frac{\partial}{\partial \varepsilon}\{\boldsymbol{T}(H_\varepsilon)\}\Big|_{\varepsilon=0} \\ + \int \frac{\partial^2 \boldsymbol{\psi}}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}} \frac{\partial \boldsymbol{T}(H_\varepsilon)\}}{\partial \varepsilon}\Big|_{\varepsilon=0} dG \cdot \frac{\partial \boldsymbol{T}(H_\varepsilon)}{\partial \varepsilon}\Big|_{\varepsilon=0} \\ + \int \frac{\partial \boldsymbol{\psi}}{\partial \boldsymbol{\theta}} dG \cdot \frac{\partial^2 \boldsymbol{T}(H_\varepsilon)}{\partial \varepsilon^2}\Big|_{\varepsilon=0} = \mathbf{0}. \end{aligned} \quad (3.43)$$

Therefore the influence function with second order as

$$\frac{\partial^2}{\partial \varepsilon^2}\{\boldsymbol{T}(H_\varepsilon)\}\Big|_{\varepsilon=0} \equiv \boldsymbol{T}^{(2)}(z, z; G). \quad (3.44)$$

Recall

$$\int \frac{\partial \boldsymbol{\psi}(z, \boldsymbol{T}(G))'}{\partial \boldsymbol{\theta}} dG(z) = \mathbf{0}, \quad (3.45)$$

and the equation(3.43) can be rewritten as below

$$\begin{aligned} 2\left(\frac{\partial \boldsymbol{\psi}(z, \boldsymbol{T}(G))'}{\partial \boldsymbol{\theta}}\Big|_{z=x} + R(\boldsymbol{\psi}, G)\right)\boldsymbol{T}^{(1)} \\ + \left(\int \frac{\partial^2 \boldsymbol{\psi}}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}} \boldsymbol{T}^{(1)}(x; G) dG\right)\boldsymbol{T}^{(1)} - R(\boldsymbol{\psi}, G)\boldsymbol{T}^{(2)} = \mathbf{0}. \end{aligned} \quad (3.46)$$

Thus we obtained

$$\begin{aligned} &\boldsymbol{T}^{(2)}(z, z; \hat{G}) = R(\boldsymbol{\psi}, \hat{G})^{-1} \\ &\left(2\,\frac{\partial \boldsymbol{\psi}(z, \boldsymbol{T}(G))'}{\partial \boldsymbol{\theta}}\Big|_{z=x} + \int \frac{\partial^2 \boldsymbol{\psi}}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}} \boldsymbol{T}^{(1)}(x; \hat{G}) d\hat{G}\right)\boldsymbol{T}^{(1)} + 2\,\boldsymbol{T}^{(1)}. \end{aligned} \quad (3.47)$$

3.4 Other Information Criteria

An information criterion for the model $f(z|\boldsymbol{x};\hat{\boldsymbol{\theta}})$ with obtained by maximizing the penalized log-likelihood function(3.20) is given by

$$\mathrm{GIC}_P = -2\sum_{\alpha=1}^{n}\log f(z_\alpha|\boldsymbol{x};\hat{\boldsymbol{\theta}}) + 2\mathrm{tr}\left\{R(\boldsymbol{\psi},\hat{G})^{-1}Q(\boldsymbol{\psi},\hat{G})\right\}, \quad (3.48)$$

where $R(\boldsymbol{\psi},\hat{G})$ and $Q(\boldsymbol{\psi},\hat{G})$ are $(m+1)\times(m+1)$ matrices [17]. respectively given by

$$\begin{aligned} R(\boldsymbol{\psi},\hat{G}) &= -\frac{1}{n}\sum_{\alpha}^{n}\left.\frac{\partial\boldsymbol{\psi}(z_\alpha,\boldsymbol{\theta})'}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \\ &= \frac{1}{n\hat{\sigma}^2}\begin{bmatrix} B'B+n\lambda\hat{\sigma}^2K & \frac{1}{\hat{\sigma}^2}B'\Lambda\mathbf{1}_n \\ \frac{1}{\hat{\sigma}^2}\mathbf{1}_n'\Lambda B & \frac{n}{2\hat{\sigma}^2}\end{bmatrix}, \end{aligned} \quad (3.49)$$

$$\begin{aligned} Q(\boldsymbol{\psi},\hat{G}) &= \frac{1}{n}\sum_{\alpha}^{n}\boldsymbol{\psi}(z_\alpha,\boldsymbol{\theta})\left.\frac{\partial\log f(z_\alpha|\boldsymbol{x}_\alpha;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \\ &= \frac{1}{n\hat{\sigma}^2}\begin{bmatrix} \frac{1}{\hat{\sigma}^2}B'\Lambda^2B-\lambda K\hat{\boldsymbol{\omega}}\mathbf{1}_n'\Lambda B & \frac{1}{2\hat{\sigma}^4}B'\Lambda^3\mathbf{1}_n-\frac{1}{2\hat{\sigma}^2}B'\Lambda\mathbf{1}_n \\ \frac{1}{2\hat{\sigma}^4}\mathbf{1}_n'\Lambda^3B-\frac{1}{2\hat{\sigma}^2}\mathbf{1}_n'\Lambda B & \frac{1}{4\hat{\sigma}^6}\mathbf{1}_n'\Lambda^4\mathbf{1}_n-\frac{n}{4\hat{\sigma}^2}\end{bmatrix}. \end{aligned} \quad (3.50)$$

When we denote $H=\hat{G}^{(-\alpha)}$ in equation(3.10), then the Taylor series of the functional $\hat{\boldsymbol{\theta}}^{(-\alpha)}=\boldsymbol{T}(\hat{G}^{(-\alpha)})$ becomes below

$$\begin{aligned} \boldsymbol{T}(\hat{G}^{(-\alpha)}) = \boldsymbol{T}(G) &+ \frac{1}{n-1}\sum_{i\neq\alpha}^{n}\boldsymbol{T}^{(1)}(y_i;G)) \\ &+ \frac{1}{2(n-1)^2}\sum_{i\neq\alpha}^{n}\sum_{j\neq\alpha}^{n}\boldsymbol{T}^{(2)}(y_i,y_j;G)+o_p(n^{-1}) \\ = \boldsymbol{T}(\hat{G}) &- \frac{1}{n-1}\boldsymbol{T}^{(1)}(y_\alpha;\hat{G})+\frac{1}{2(n-1)^2}\boldsymbol{T}^{(2)}(y_\alpha,y_\alpha;\hat{G}) \\ &+ o_p(n^{-1}), \end{aligned} \quad (3.51)$$

where $\hat{G}^{(-\alpha)}$ is the empirical distribution that consists of $n-1$ samples without α-th sample. The first order of influence function $\boldsymbol{T}^{(1)}$ is effective and it does not take much calculation. But the second order of influence function $\boldsymbol{T}^{(2)}$ has small profits and large calculation as much as CV. in our experiment.

Therefore We adopt the next approximation for the alternative CV [17]

$$\begin{aligned} \boldsymbol{T}(\hat{G}^{(-\alpha)}) &\approx \boldsymbol{T}(G) + \frac{1}{n-1}\sum_{i\neq\alpha}^{n} \boldsymbol{T}^{(1)}(z_i; G) \\ &\approx \boldsymbol{T}(\hat{G}) - \frac{1}{n}\boldsymbol{T}^{(1)}(z_\alpha; \hat{G}). \end{aligned} \tag{3.52}$$

In the equation(3.14) of IC_{CV} we replace the $\hat{\boldsymbol{\theta}}^{(-\alpha)}$ with $\hat{\boldsymbol{\theta}} - \frac{1}{n}\boldsymbol{T}^{(1)}(z_\alpha; \hat{G})$ and its scheme is called as modified GIC (mGIC)[29],

$$\mathrm{mGIC} = -2\sum_{\alpha=1}^{n} \log f\left(\boldsymbol{x}_\alpha; \hat{\boldsymbol{\theta}} - \frac{1}{n}\boldsymbol{T}^{(1)}(x_\alpha; \hat{G})\right). \tag{3.53}$$

3.5 Numerical Simulation for Surface Estimation

3.5.1 Surfaces and Samples

We assume two models of the equations of surface I and surface II as follows

$$\mathrm{I}: \quad z = \sin(2\pi x) + 2\cos(2\pi(x+y)), \tag{3.54}$$

$$\mathrm{II}: \quad z = (1-x)\exp(-x^2) + xy\exp(-y^2). \tag{3.55}$$

Those topographies are shown in Figure 3.3. For the estimation we generated a set of 300 sample coordinates data with the Gaussian noise

according to the normal distribution $N(0, \sigma^2)$.

3.5.2 Estimation of Parameters

Usually B-splines with order four (degree three) are used in the calculation. Along x direction we set the knots $x_1, x_2, \cdots, x_p$ and the knots at the both ends are four-folded.
So the total number of basis B-splines will be $p-4$. At every interval $[x_{n-1}, x_n), n = 5, 6, \cdots, p-3$ there exist four basis B-splines and those are below.

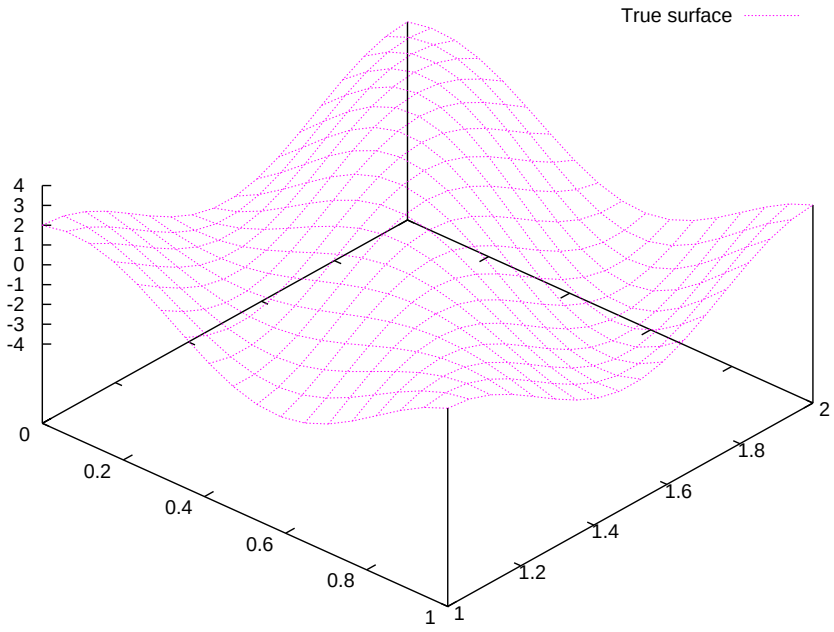
$$B_{n,1}(x) = \frac{-(x-x_n)^3}{(x_n - x_{n-3})(x_n - x_{n-2})(x_n - x_{n-1})}, \tag{3.56}$$

$$\begin{aligned} B_{n,2}(x) = & \frac{(x-x_{n-3})(x-x_n)(x-x_n)}{(x_n - x_{n-3})(x_n - x_{n-2})(x_n - x_{n-1})} \\ & + \frac{(x-x_{n-2})(x-x_n)(x-x_{n+1})}{(x_{n+1} - x_{n-2})(x_n - x_{n-2})(x_n - x_{n-1})} \\ & + \frac{(x-x_{n-1})(x-x_{n+1})(x-x_{n+1})}{(x_{n+1} - x_{n-1})(x_{n+1} - x_{n-2})(x_n - x_{n-1})}, \end{aligned} \tag{3.57}$$

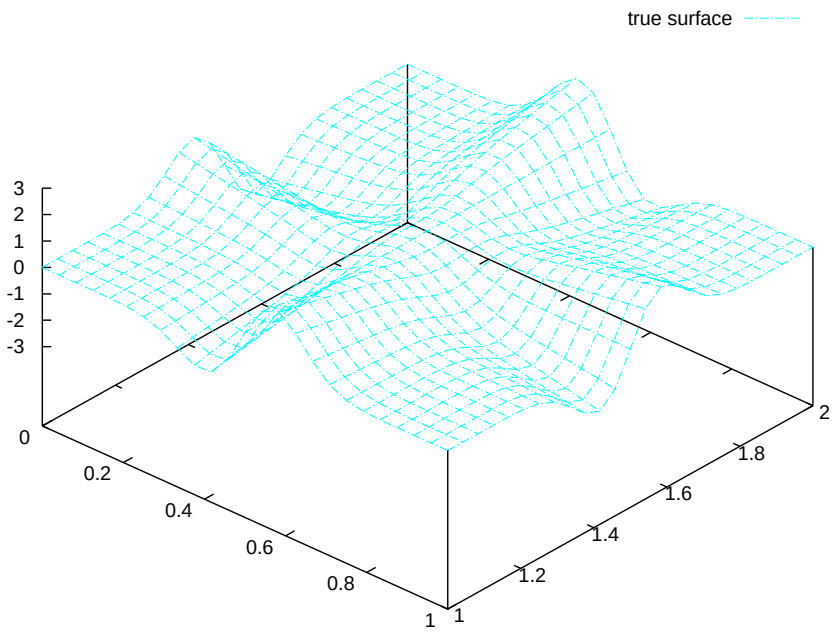
$$\begin{aligned} B_{n,3}(x) = & -\frac{(x-x_{n-2})(x-x_{n-2})(x-x_n)}{(x_{n+1} - x_{n-2})(x_n - x_{n-2})(x_n - x_{n-1})} \\ & - \frac{(x-x_{n-2})(x-x_{n-1})(x-x_{n+1})}{(x_{n+1} - x_{n-2})(x_n - x_{n-1})(x_{n+1} - x_{n-1})} \\ & - \frac{(x-x_{n-1})(x-x_{n-1})(x-x_{n+2})}{(x_{n+1} - x_{n-1})(x_n - x_{n-1})(x_{n+2} - x_{n-1})}, \end{aligned} \tag{3.58}$$

$$B_{n,4}(x) = \frac{(x-x_{n-1})^3}{(x_{n+2} - x_{n-1})(x_{n+1} - x_{n-1})(x_n - x_{n-1})}. \tag{3.59}$$

According to the total number of sample data we set 10 - 20 knots along the every axis. We denote the total number of knots (n_1, n_2)



(a) Surface I



(b) Surface II

Figure 3.3: The topographies of two surfaces

where n_1 and n_2 are the total number of knots along x and y directions respectively. And for every (n_1, n_2), we tested 100 sets of randomized knots generated uniformly. But some of them don't satisfy the Schoenberg-Whitney condition then we generate other set of knots again. Furthermore the equations of matrices made from ill conditioned sets cannot be solved properly then we also generate another sets of knots again so we tested 100 solvable sets for every (n_1, n_2). The total number of basis will be $(n_1-4)(n_2-4)$ and the total number of the parameters will be $(n_1-4)(n_2-4)+1$ which is consist of the coefficients of the basis and the variance. For the regularization term we use the equation(3.21) and at the numerical calculation we use the equations (3.33) and (3.34). We tested the estimation with various β's which are from 10^{-1} to 10^{-10} in principle.

3.5.3 Evaluation of Models

For the evaluation of the obtained parameters we test some criteria such as GIC_P, mGIC, CV and GCV_{IF}. Those results are shown in the Tables and Figures below.

Tables 3.1,3.2 summarize the results of GIC_P over the various values of β. The GIC_P values are monotone decreasing so we cannot determine the optimal parameters in this case.

Tables 3.3,3.4 summarize the results of CV over the various values of β. Optimal value which minimizes the information criterion Cross-Validation (3.14) determine the number of knots and the value of β. We can determine the optimal parameters which minimize CV. But the repetition is 12100 and it takes about 800 minutes in our simulation for every β and every surface.

Table 3.1: GIC_P results for the surface I

total number of knots					
x-axis	y-axis	β	σ^2	λ	GIC_P
19	13	1.00E-01	1.850992	5.4025E-02	1049.27
19	13	1.00E-02	1.672307	5.9798E-03	1020.16
20	17	1.00E-03	0.937645	1.0665E-03	856.57
20	17	1.00E-04	0.316010	3.1645E-04	550.51
13	11	1.00E-05	0.224000	4.4643E-05	454.96
19	16	1.00E-06	0.181253	5.5171E-06	448.37
19	19	1.00E-07	0.106759	9.3669E-07	390.52
19	19	1.00E-08	0.082473	1.2125E-07	337.60
19	19	1.00E-09	0.068697	1.4557E-08	293.71
20	20	1.00E-10	0.060981	1.6399E-09	257.63

Table 3.2: GIC_P results for the surface II

total number of knots					
x-axis	y-axis	β	σ^2	λ	GIC_P
18	11	1.000E-01	0.407434	2.4544E-01	596.15
20	19	1.000E-02	0.397532	2.5155E-02	590.43
20	15	1.000E-03	0.322460	3.1012E-03	535.90
17	20	1.000E-04	0.165861	6.0291E-04	359.24
17	20	1.000E-05	0.047773	2.0932E-04	33.99
17	20	1.000E-06	0.005674	1.7623E-04	-604.09
20	20	1.000E-07	0.001115	8.9720E-05	-947.39
20	20	1.000E-08	0.000771	1.2969E-05	-1040.50
20	20	1.000E-09	0.000631	1.5844E-06	-1100.57
20	20	1.000E-10	0.000551	1.8140E-07	-1138.09

Table 3.3: CV results for the surface I

total number of knots					
x-axis	y-axis	β	σ^2	λ	CV
19	13	1.00E-01	1.85E+00	5.4025E-02	1050.12
19	13	1.00E-02	1.67E+00	5.9798E-03	1021.12
20	17	1.00E-03	9.38E-01	1.0665E-03	859.33
20	17	1.00E-04	3.16E-01	3.1645E-04	560.37
13	**11**	**1.00E-05**	**2.24E-01**	**4.4643E-05**	**468.14**
10	11	1.00E-06	2.29E-01	4.3673E-06	475.16
10	11	1.00E-07	2.28E-01	4.3939E-07	478.01
10	11	1.00E-08	2.24E-01	4.4721E-08	478.89
10	11	1.00E-09	2.20E-01	4.5520E-09	479.82
10	10	1.00E-10	2.20E-01	4.5354E-10	486.15

Table 3.4: CV results for the surface II

total number of knots					
x-axis	y-axis	β	σ^2	λ	CV
13	13	1.00E-01	0.003851	2.59673E+01	596.86
18	18	1.00E-02	0.053463	1.87045E-01	591.29
20	15	1.00E-03	0.322460	3.10116E-03	537.93
17	20	1.00E-04	0.165861	6.02913E-04	367.40
20	15	1.00E-05	0.048688	2.05391E-04	74.36
16	14	1.00E-06	0.006367	1.57050E-04	-346.11
16	**14**	**1.00E-07**	**0.002303**	**4.34181E-05**	**-441.16**
19	13	1.00E-08	0.002769	3.61129E-06	-244.25
13	13	1.00E-09	0.003851	2.59673E-07	25.99
13	13	1.00E-10	0.003851	2.59673E-08	31.07

Table 3.5: mGIC results for the surface I

total number of knots					
x-axis	y-axis	β	σ^2	λ	mGIC
19	13	1.000E-01	1.850992	5.4025E-02	1049.57
19	13	1.000E-02	1.672307	5.9798E-03	1020.50
20	17	1.000E-03	0.937645	1.0665E-03	857.52
20	17	1.000E-04	0.316010	3.1645E-04	553.14
13	11	1.000E-05	0.224000	4.4643E-05	458.32
12	10	1.000E-06	0.214460	4.6629E-06	456.67
19	19	1.000E-07	0.106759	9.3669E-07	434.30
19	19	1.000E-08	0.082473	1.2125E-07	392.92
20	19	1.000E-09	0.081656	1.2247E-08	352.85
19	19	1.000E-10	0.061331	1.6305E-09	323.84

Table 3.6: mGIC results for the surface II

total number of knots					
x-axis	y-axis	β	σ^2	λ	mGIC
12	14	1.00E-01	0.200834	0.497923	556.37
18	18	1.00E-02	0.053463	0.187047	547.02
20	15	1.00E-03	0.322460	0.003101	536.67
17	20	1.00E-04	0.165861	0.000603	361.89
17	20	1.00E-05	0.047773	0.000209	44.76
19	15	1.00E-06	0.005595	0.000179	-500.40
20	20	1.00E-07	0.001115	0.000090	-889.99
20	20	1.00E-08	0.000771	0.000013	-973.54
20	20	1.00E-09	0.000509	0.000002	-1095.15
20	20	1.00E-10	0.000483	0.000000	-1103.50

Table 3.7: GCV_{IF} results for the surface I

total number of knots					
x-axis	y-axis	β	σ^2	λ	GCV_{IF}
19	13	1.000E-01	1.850993	5.40E-02	1050.13
19	13	1.000E-02	1.672307	5.98E-03	1021.47
20	17	1.000E-03	0.937645	1.07E-03	857.20
20	17	1.000E-04	0.316010	3.16E-04	556.99
13	**11**	**1.000E-05**	**0.224000**	**4.46E-05**	**470.02**
12	10	1.000E-06	0.214460	4.66E-06	478.49
10	11	1.000E-07	0.227586	4.39E-07	481.99
10	11	1.000E-08	0.223611	4.47E-08	484.37
10	10	1.000E-09	0.220486	4.54E-09	487.35
10	10	1.000E-10	0.220486	4.54E-10	487.36

Table 3.8: GCV_{IF} results for the surface II

total number of knots					
x-axis	y-axis	β	σ^2	λ	GCV_{IF}
18	11	1.000E-01	0.407434	2.45E-01	596.93
20	19	1.000E-02	0.397532	2.52E-02	591.12
20	15	1.000E-03	0.355688	2.81E-03	536.83
17	20	1.000E-04	0.165862	6.03E-04	363.96
10	10	1.000E-05	0.227627	4.39E-05	60.32
14	17	1.000E-06	0.005844	1.71E-04	-463.05
14	**17**	**1.000E-07**	**0.001890**	**5.29E-05**	**-659.31**
14	17	1.000E-08	0.001712	5.84E-06	-582.75
13	13	1.000E-09	0.003851	2.60E-07	-549.87
13	13	1.000E-10	0.003851	2.60E-08	-549.52

Tables 3.5,3.6 summarize the results of mGIC over the various values of β. We use the influence function to estimate the value of parameters. This method can obtain the almost same result in parameters as CV and it takes very small time. It is almost 35 minutes for every β and it is 1/23 of CV. In this approximation of parameters the difference between the mGIC and CV are quite small. The correlation coefficient is almost 1.0. For example, when $\beta = 10^{-5}, n_1 = n_2 = 20$ the average of the correlation coefficients over 100 sets of knots is 0.99999611. But the values of mGIC are monotone decreasing so we cannot determine the optimal parameters in this case.

The results of an alternative method GCV_{IF} with influence function (3.16) for model selection are shown in Table 3.7 ,3.8. In the estimation of variance we use the influence function. For the evaluation of the obtained parameters we test some criteria such as GICp, mGIC, CV and GCV_{IF}. Those results are shown in Figure 3.4 and 3.5. Figure 3.4 summarize the results of GICp, mGIC, CV and GCV_{IF} over the various values of for surface I. And Figure 3.5 shows the results of four criteria for surface II. The GICp values are monotone decreasing so we cannot determine the optimal parameters in this case. The selected models with optimal parameters determined by the GCV_{IF} are shown in Figure 3.6 and Figure 3.7 for the surface I and II respectively. In Figure 3.6 the estimated surface is based on the 95-th set of (13,11) knots. In Figure 3.7 the estimated surface is based on the 99-th set of (14,17) knots. In those figures the locations of knots and samples and the residuals are also shown.

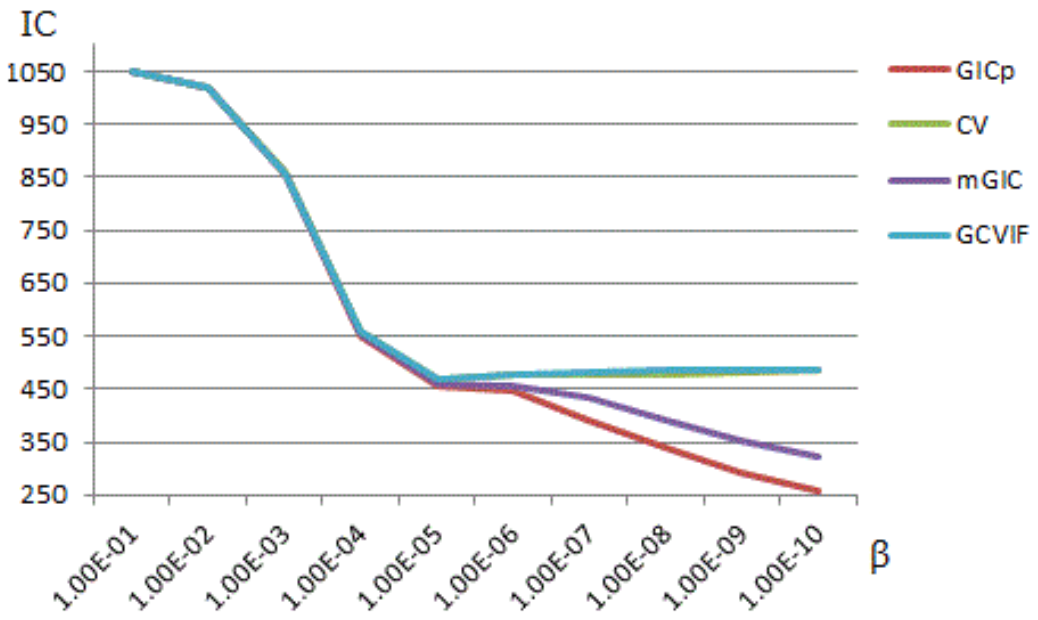


Figure 3.4: Four criteria of Surface I

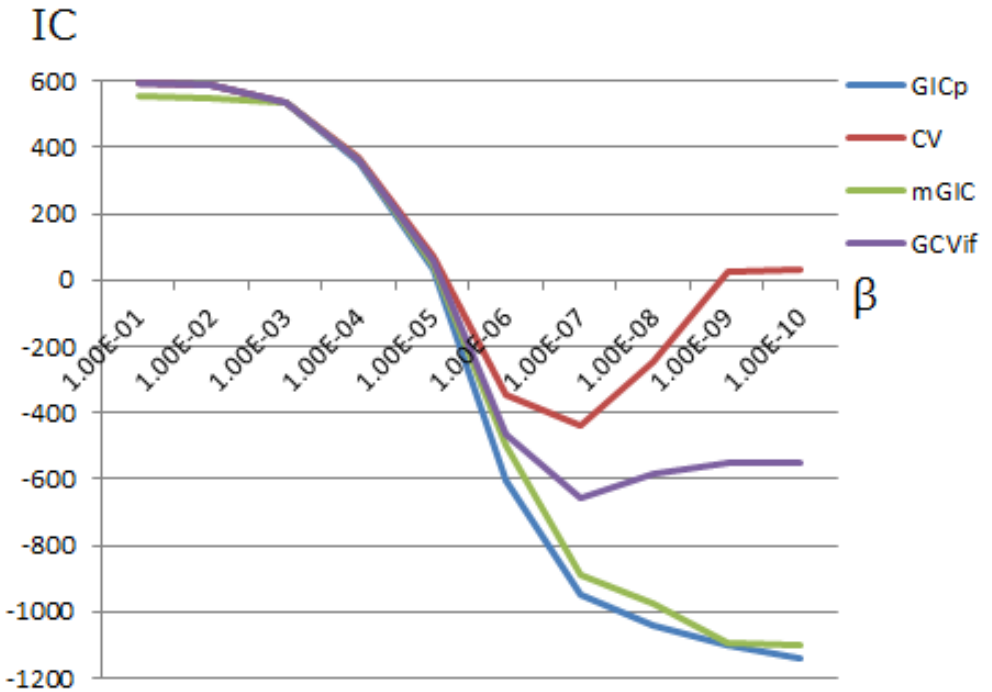
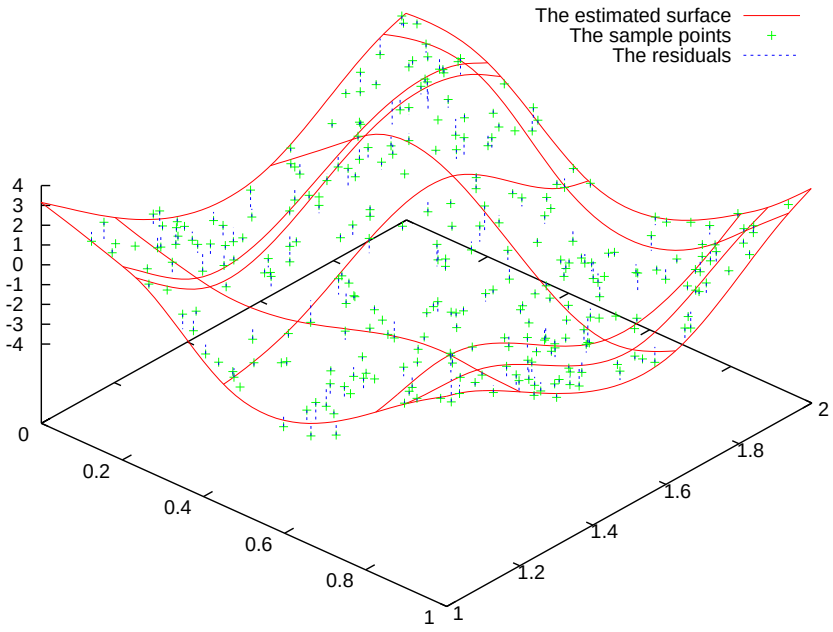
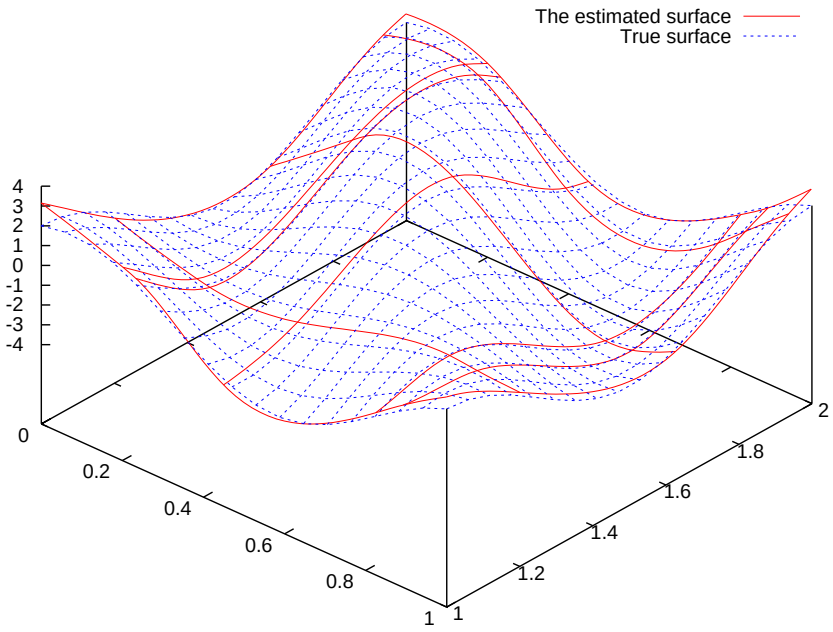


Figure 3.5: Four criteria of Surface II

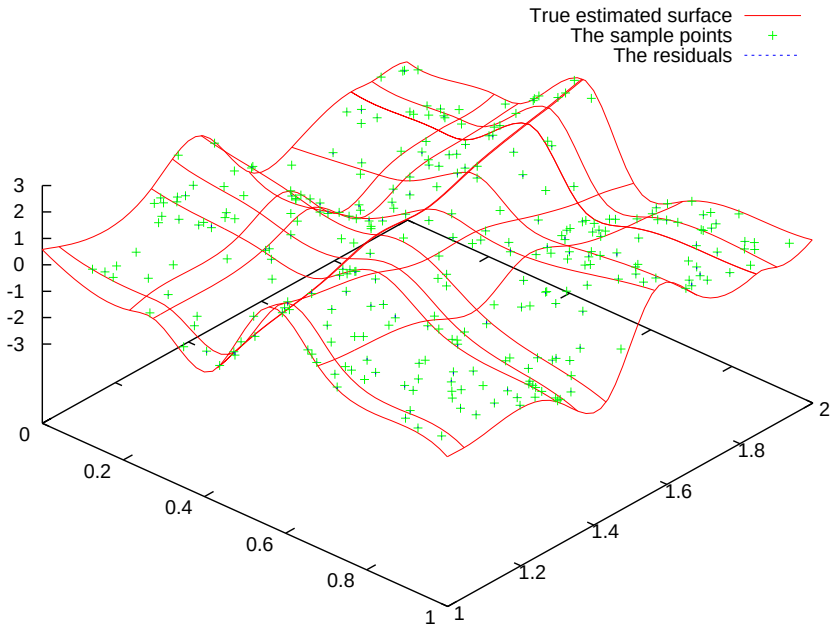


(a) Sample and estimated surface

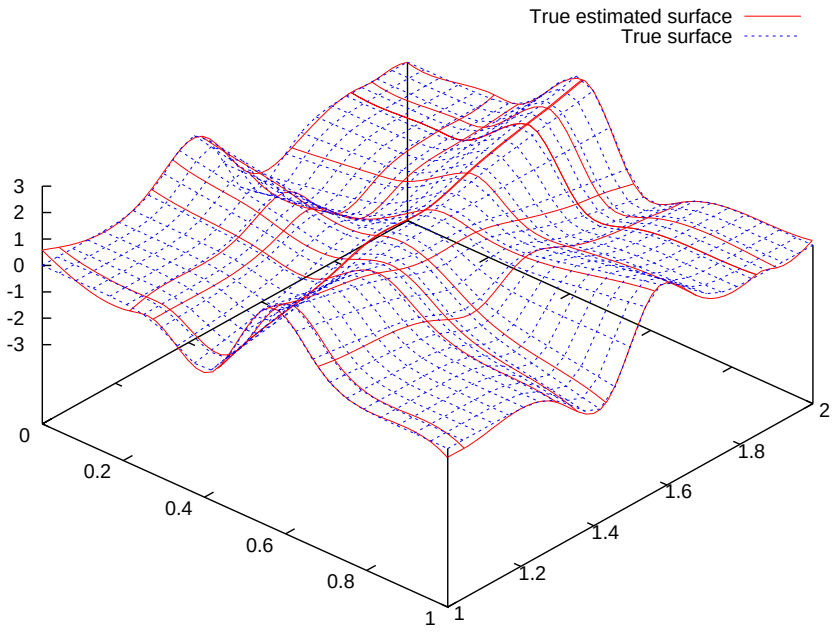


(b) Comparison of true surface and estimated surface

Figure 3.6: Surface I



(a) Sample and estimated surface



(b) Comparison of true surface and estimated surface

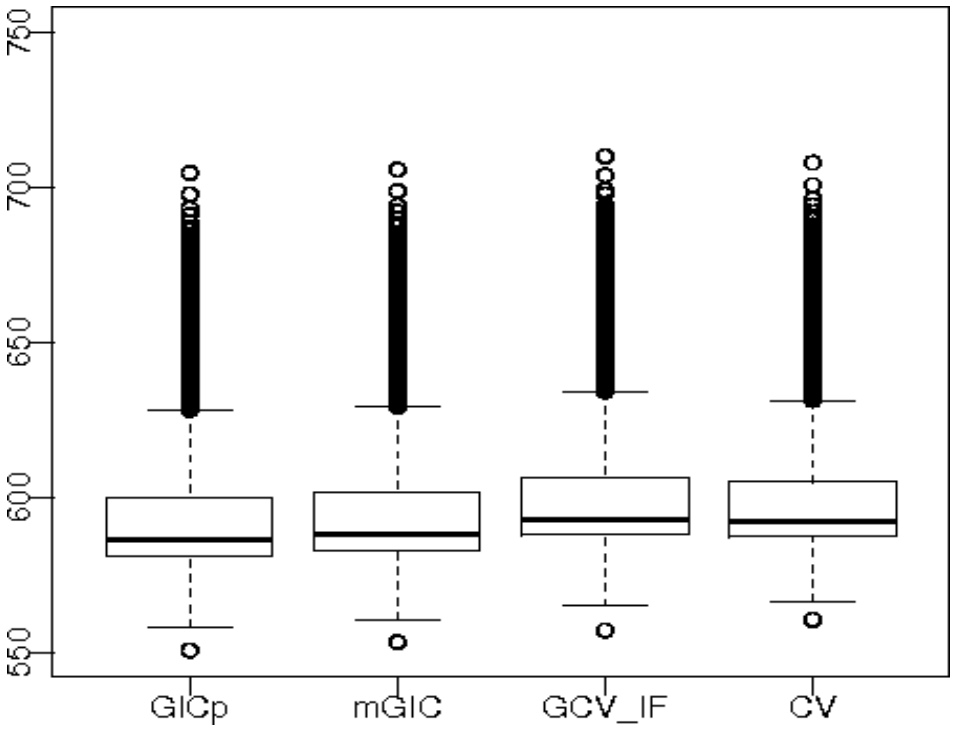
Figure 3.7: Surface II

3.5.4 Comparison between the Distributions of Criteria

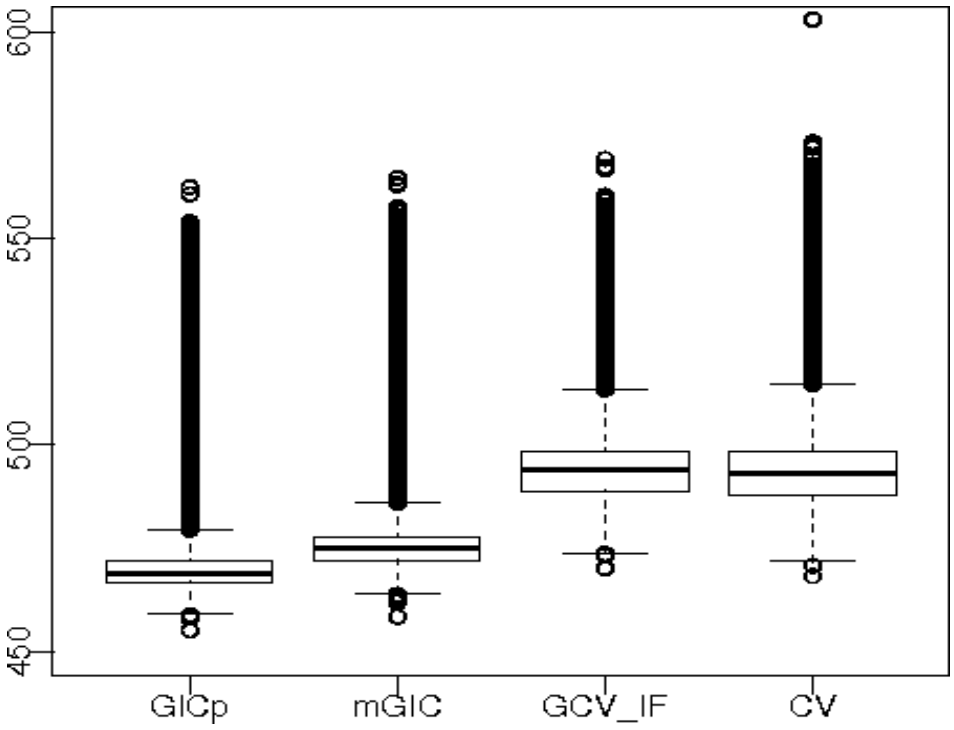
We compare the distributions of four information criteria for the two surfaces I,II. The criteria are GIC_P, CV, mGIC and GCV_{IF} we have improved. On the surface I, we show the *Boxplots* of four criteria over $\beta = 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}$ in Figures 3.8,3.10 and graphs of those in Figures 3.9, 3.11. Similarly on the surface II, we show the *Boxplots* of four criteria over $\beta = 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}$ in Figures 3.12,3.14 and graphs of those in Figures 3.13, 3.15. We can find from these Tables and Figures that mGIC is larger than GIC_P and GCV_{IF} is closer to CV than others.

Table 3.9: Five number summary of information criteria for $\beta = 10^{-4}$ of surface I

	GICp		mGIC		GCV_{IF}		CV
Min. :	550.5	Min. :	553.1	Min. :	557	Min. :	560.4
1st Qu.:	580.8	1st Qu.:	582.8	1st Qu.:	587.8	1st Qu.:	587.1
Median :	586	Median :	587.8	Median :	592.7	Median :	591.8
Mean :	594	Mean :	595.8	Mean :	600.9	Mean :	599.8
3rd Qu.:	599.7	3rd Qu.:	601.3	3rd Qu.:	606.3	3rd Qu.:	604.7
Max. :	704.5	Max. :	705.5	Max. :	709.9	Max. :	1847.5

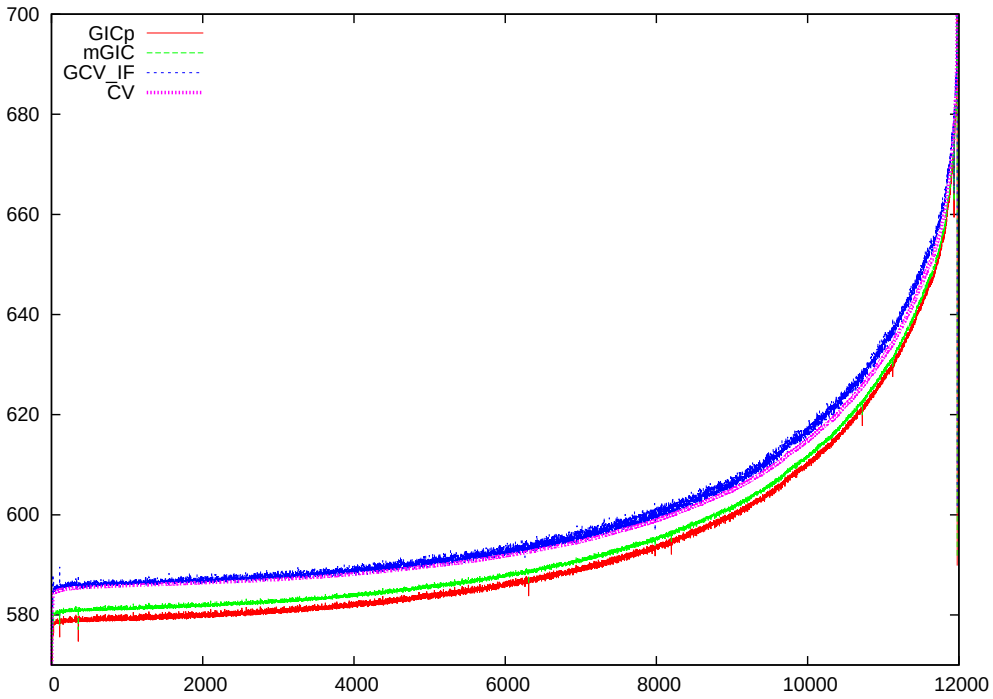


(a) $\beta = 10^{-4}$

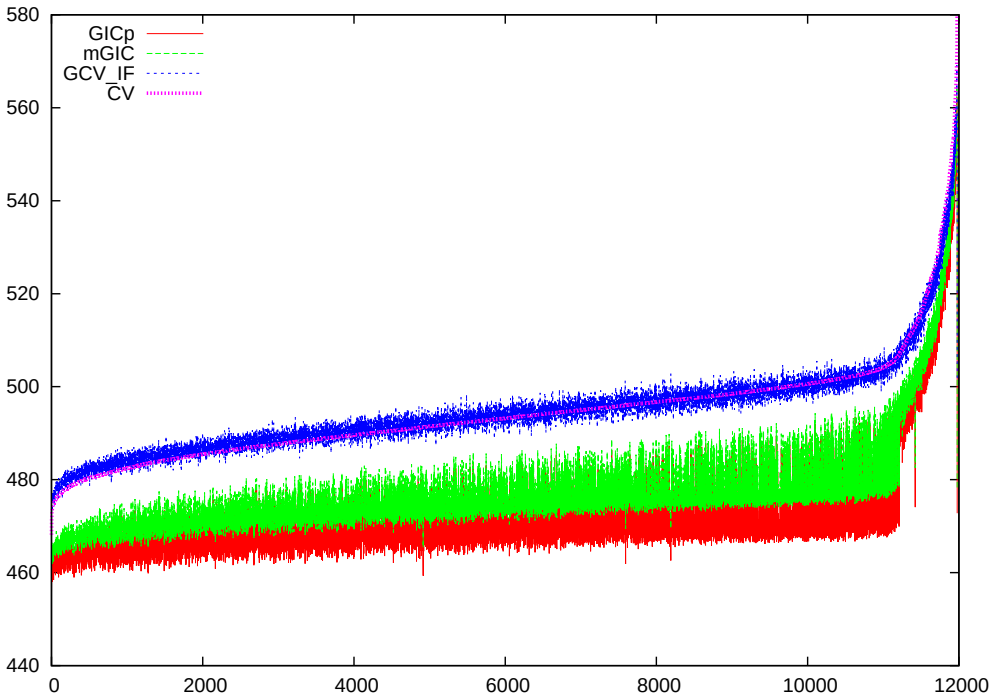


(b) $\beta = 10^{-5}$

Figure 3.8: *Boxplots* for four criteria of surface I



(a) $\beta = 10^{-4}$



(b) $\beta = 10^{-5}$

Figure 3.9: Comparison of four criteria (ascending order of CV)

Table 3.10: Five number summary of information criteria for $\beta = 10^{-5}$ of surface I

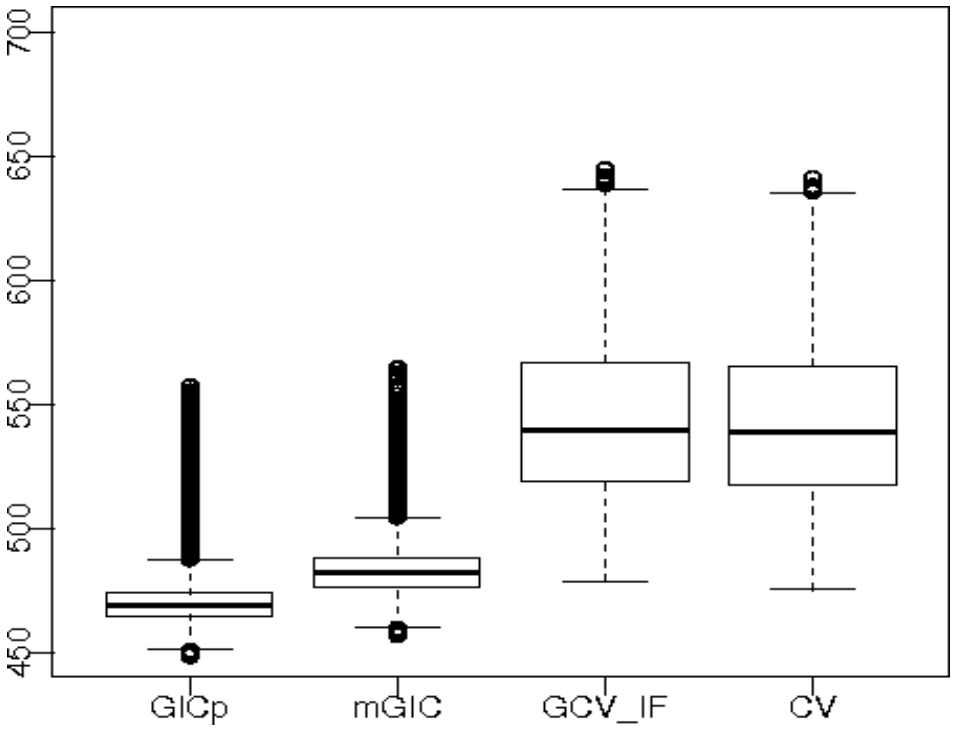
	GICp		mGIC		GCV_{IF}		CV
Min. :	455	Min. :	458.3	Min. :	470	Min. :	468.1
1st Qu.:	466.6	1st Qu.:	472	1st Qu.:	488.6	1st Qu.:	487.7
Median :	468.7	Median :	474.8	Median :	493.8	Median :	493.2
Mean :	472.1	Mean :	477.4	Mean :	494.7	Mean :	494.4
3rd Qu.:	471.7	3rd Qu.:	477.6	3rd Qu.:	498.5	3rd Qu.:	498.4
Max. :	562.3	Max. :	564.6	Max. :	569.2	Max. :	1366

Table 3.11: Five number summary of information criteria for $\beta = 10^{-6}$ of surface I

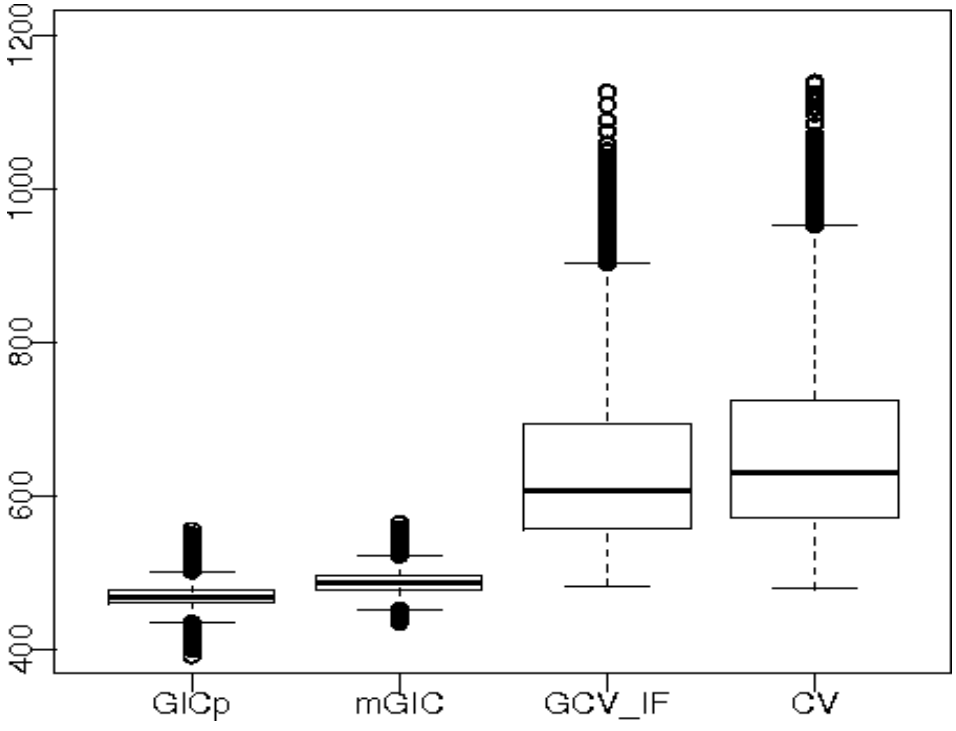
	GICp		mGIC		GCV_{IF}		CV
Min. :	448.4	Min. :	456.7	Min. :	478.5	Min. :	475.2
1st Qu.:	464.5	1st Qu.:	476.3	1st Qu.:	519	1st Qu.:	517.6
Median :	468.8	Median :	482.1	Median :	539.3	Median :	538.4
Mean :	470.5	Mean :	482.7	Mean :	543.9	Mean :	542.8
3rd Qu.:	473.7	3rd Qu.:	487.6	3rd Qu.:	566.5	3rd Qu.:	564.8
Max. :	557.2	Max. :	564.4	Max. :	644.6	Max. :	1357.8

Table 3.12: Five number summary of information criteria for $\beta = 10^{-7}$ of surface I

	GICp		mGIC		GCV_{IF}		CV
Min. :	390.5	Min. :	434.3	Min. :	482	Min. :	478
1st Qu.:	459	1st Qu.:	477.4	1st Qu.:	555.7	1st Qu.:	570.3
Median :	467.7	Median :	486.3	Median :	606.9	Median :	629.2
Mean :	467	Mean :	486.6	Mean :	636.1	Mean :	667.2
3rd Qu.:	475.7	3rd Qu.:	495	3rd Qu.:	694	3rd Qu.:	722.7
Max. :	555.2	Max. :	563.9	Max. :	1125.8	Max. :1	17387.5

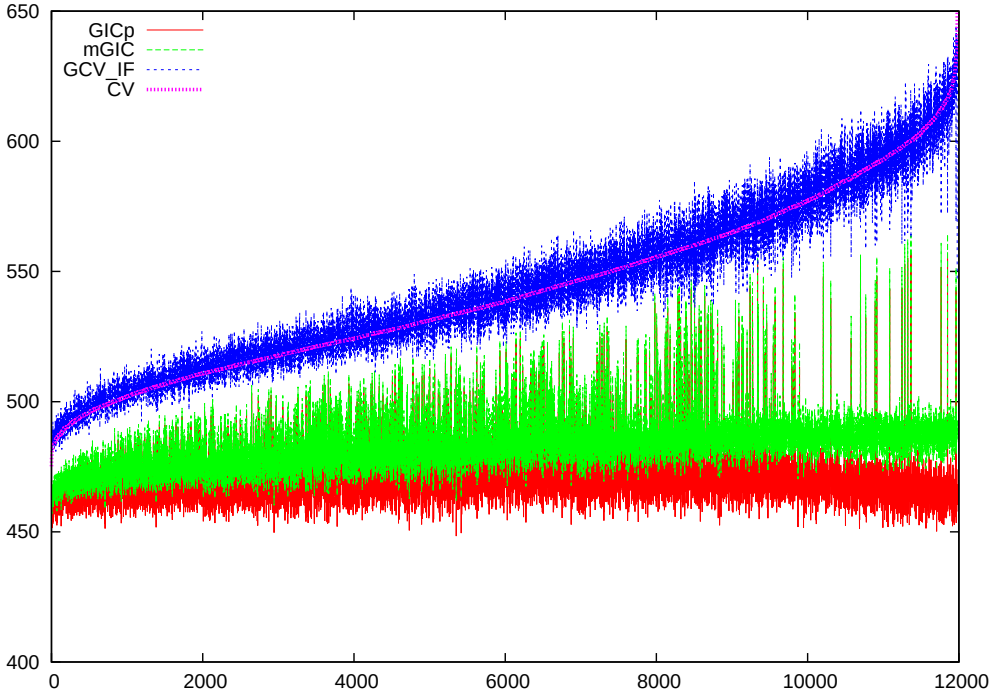


(a) $\beta = 10^{-6}$

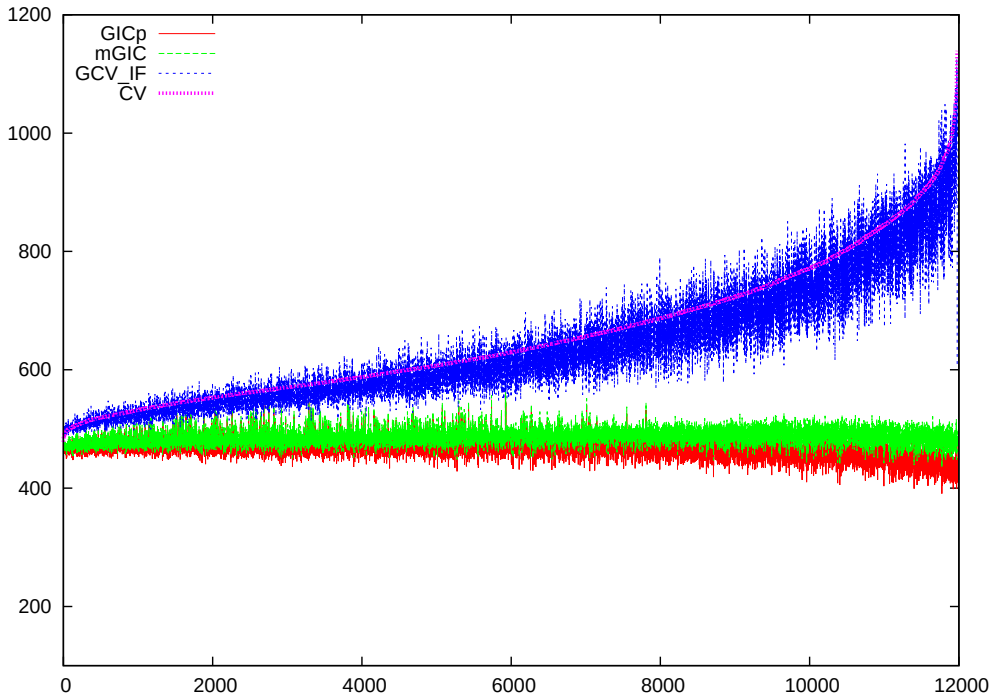


(b) $\beta = 10^{-7}$

Figure 3.10: *Boxplots* for four criteria of surface I

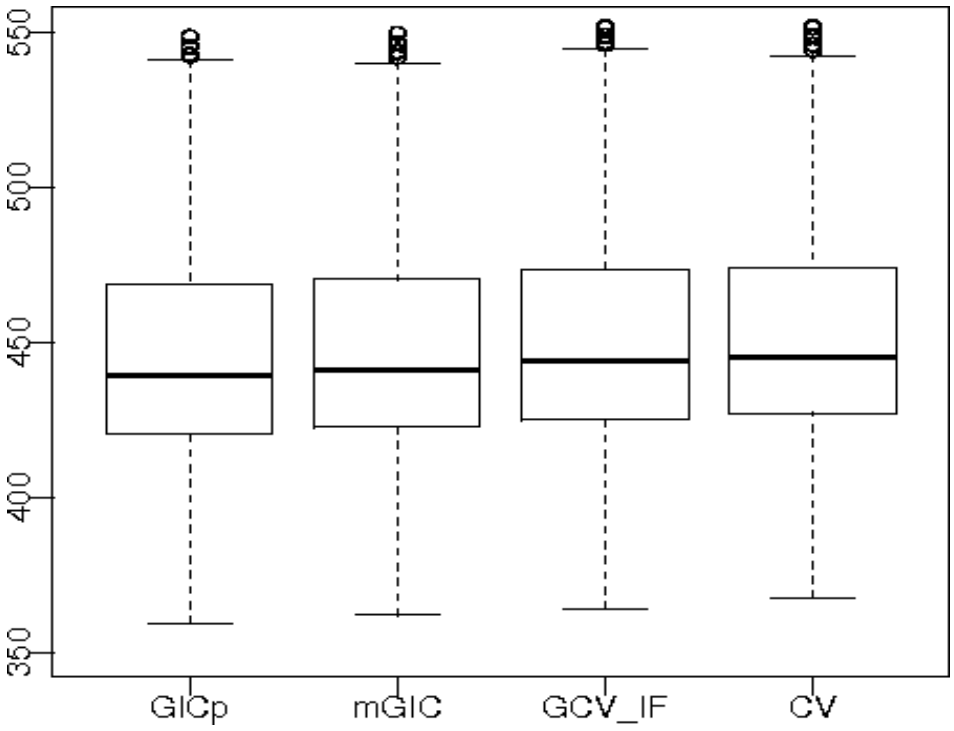


(a) $\beta = 10^{-6}$

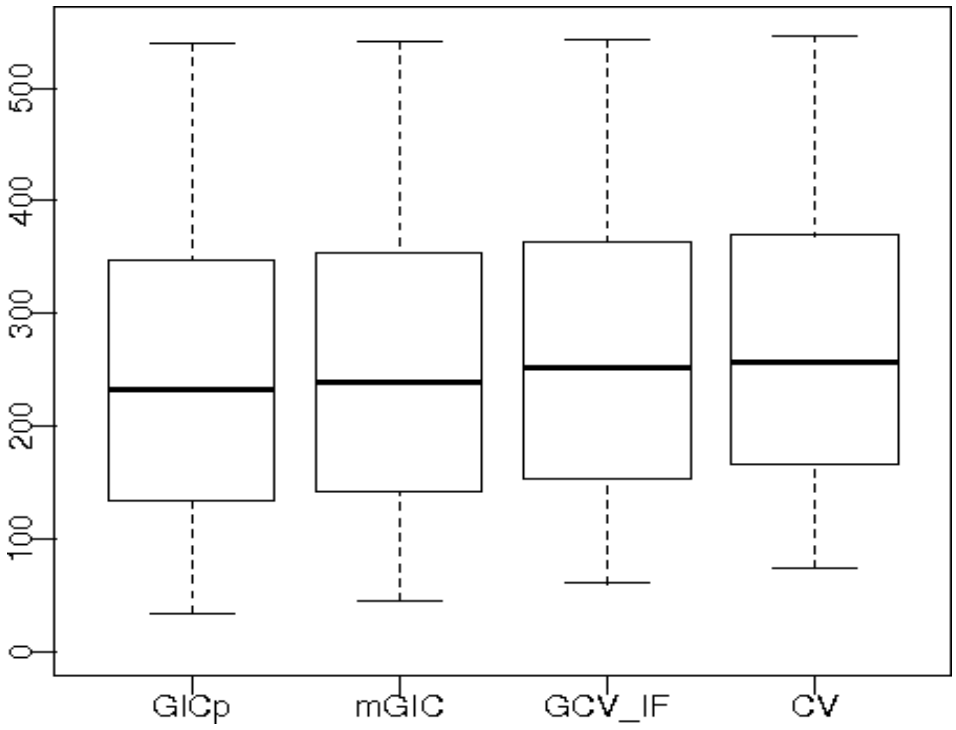


(b) $\beta = 10^{-7}$

Figure 3.11: Comparison of four criteria (ascending order of CV)

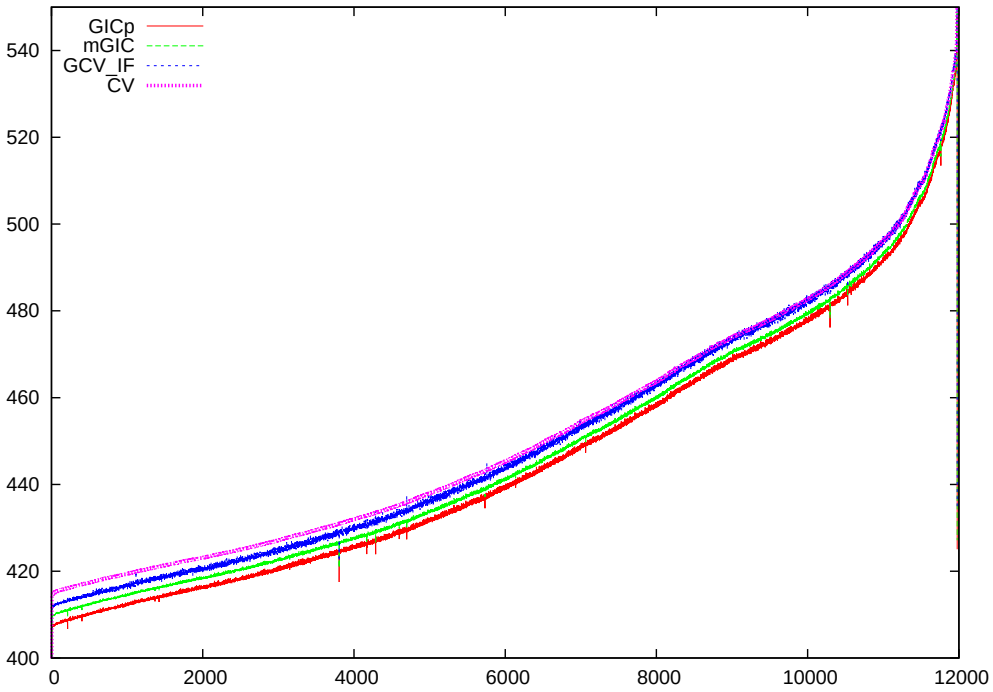


(a) $\beta = 10^{-4}$

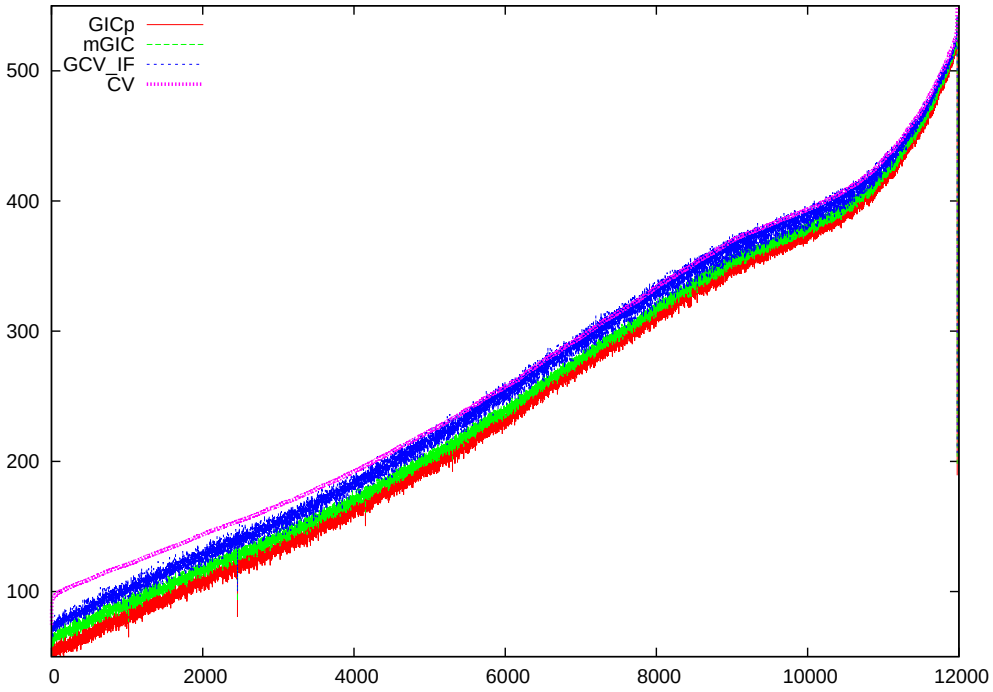


(b) $\beta = 10^{-5}$

Figure 3.12: *Boxplots* for four criteria of surface I



(a) $\beta = 10^{-4}$



(b) $\beta = 10^{-5}$

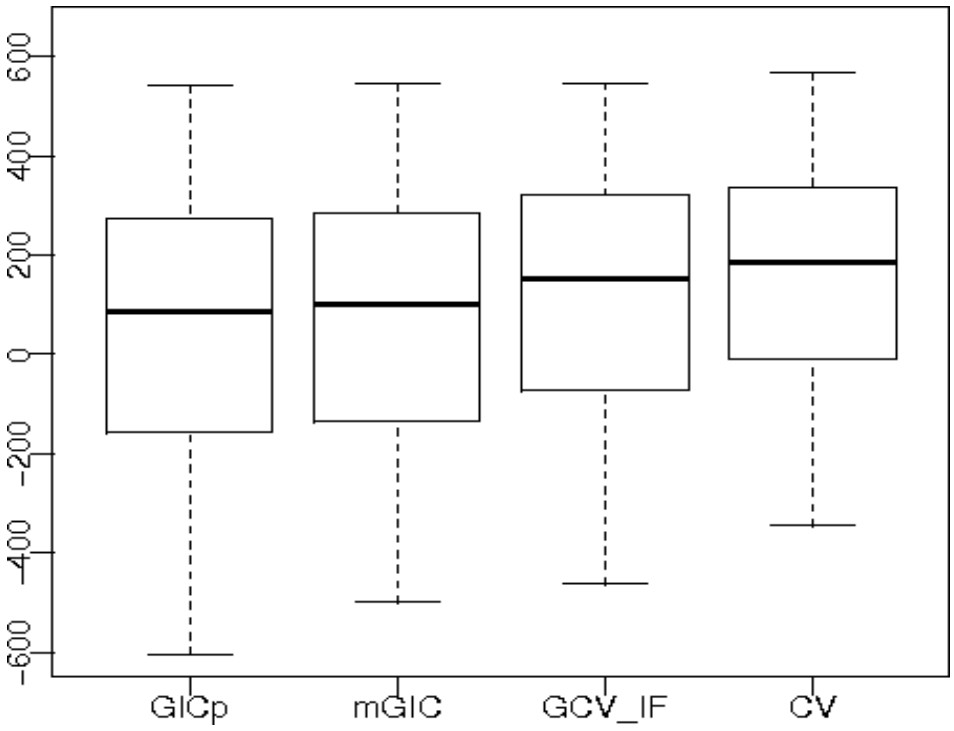
Figure 3.13: Comparison of four criteria (ascending order of CV)

Table 3.13: Five number summary of information criteria for $\beta = 10^{-4}$ of surface II

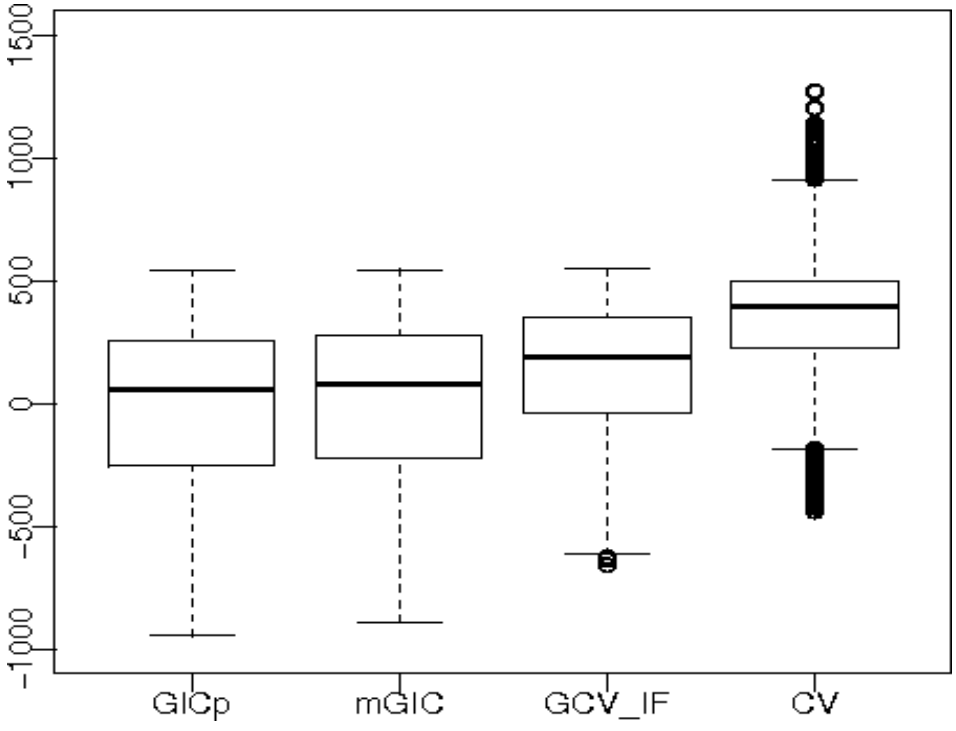
	GICp		mGIC		GCV_{IF}		CV
Min. :	359.2	Min. :	361.9	Min. :	364	Min. :	367.4
1st Qu.:	420.5	1st Qu.:	422.6	1st Qu.:	424.8	1st Qu.:	427.1
Median :	439.3	Median :	441.1	Median :	443.8	Median :	445.2
Mean :	446.1	Mean :	448	Mean :	450.6	Mean :	452
3rd Qu.:	468.6	3rd Qu.:	470.2	3rd Qu.:	473.2	3rd Qu.:	473.8
Max. :	548.4	Max. :	549.4	Max. :	551.5	Max. :	653.4

Table 3.14: Five number summary of information criteria for $\beta = 10^{-5}$ of surface II

	GICp		mGIC		GCV_{IF}		CV
Min. :	33.99	Min. :	44.76	Min. :	60.32	Min. :	74.37
1st Qu.:	133.27	1st Qu.:	141.58	1st Qu.:	152.68	1st Qu.:	165.93
Median :	232.01	Median :	238.59	Median :	251.86	Median :	256.99
Mean :	242.16	Mean :	248.99	Mean :	259.9	Mean :	268.71
3rd Qu.:	347.87	3rd Qu.:	353.34	3rd Qu.:	363.91	3rd Qu.:	369.02
Max. :	539.16	Max. :	541.22	Max. :	542.81	Max. :	854.3

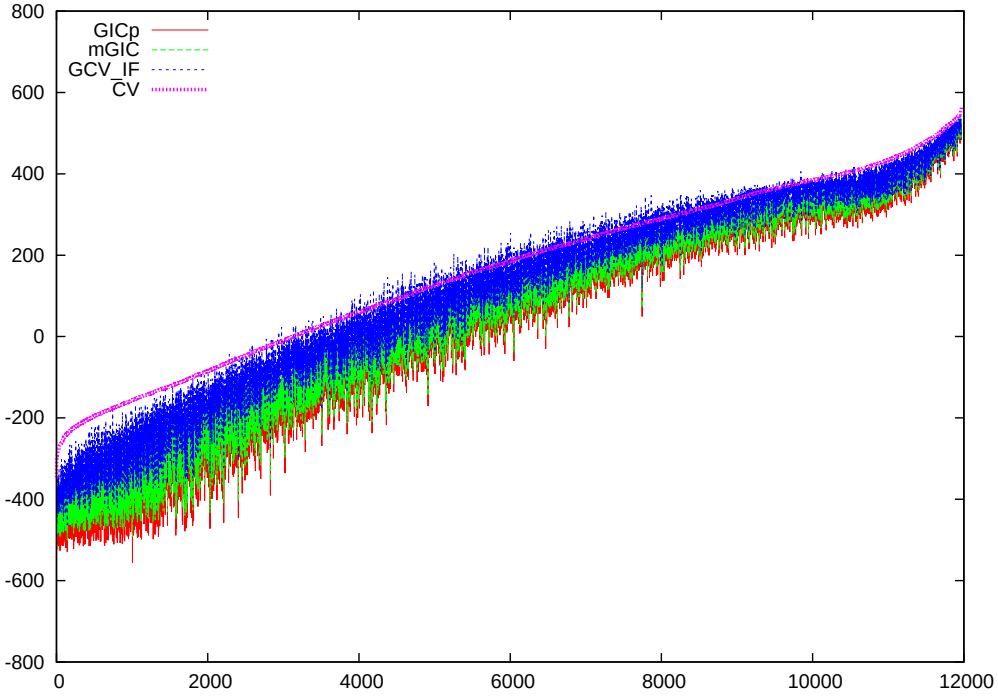


(a) $\beta = 10^{-6}$

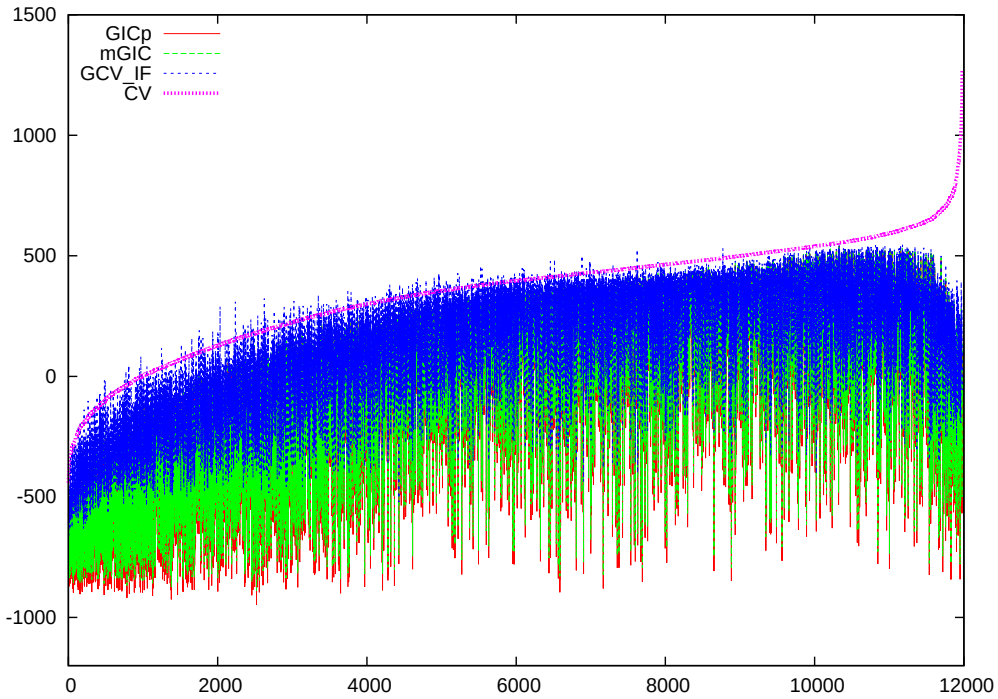


(b) $\beta = 10^{-7}$

Figure 3.14: *Boxplots* for four criteria of surface II



(a) $\beta = 10^{-6}$



(b) $\beta = 10^{-7}$

Figure 3.15: Comparison of four criteria (ascending order of CV)

Table 3.15: Five number summary of information criteria for $\beta = 10^{-6}$ of surface II

	GICp		mGIC		GCV_{IF}		CV
Min. :	-604.09	Min. :	-500.4	Min. :	-463.1	Min. :	-346.12
1st Qu.:	-157.55	1st Qu.:	-136.52	1st Qu.:	-73.4	1st Qu.:	-10.16
Median :	86.47	Median :	101.6	Median :	152.8	Median :	184.03
Mean :	47.09	Mean :	64.08	Mean :	112.8	Mean :	162.77
3rd Qu.:	272.98	3rd Qu.:	283.65	3rd Qu.:	322	3rd Qu.:	337.65
Max. :	541.6	Max. :	544.17	Max. :	545.5	Max. :	15626.06

Table 3.16: Five number summary of information criteria for $\beta = 10^{-7}$ of surface II

	GICp		mGIC		GCV_{IF}		CV
Min. :	-604.09	Min. :	-889.99	Min. :	-659.31	Min. :	-441.2
1st Qu.:	-158.11	1st Qu.:	-222.01	1st Qu.:	-41.15	1st Qu.:	224.4
Median :	84.61	Median :	79.01	Median :	191.47	Median :	396.2
Mean :	47.94	Mean :	-0.86	Mean :	132.28	Mean :	378.3
3rd Qu.:	274.12	3rd Qu.:	273.93	3rd Qu.:	349.06	3rd Qu.:	499.1
Max. :	541.04	Max. :	544.28	Max. :	546.63	Max. :	348848.6

Chapter 4

Topographical Estimation of the Lake Bottom by B-spline Surface

4.1 Estimation of the Selected Areas

For the topographical estimation we have got the experimental data of the lake. The area of the data is 2.79 km^2, and the total number of data is 1178. At first we tested the two subdomains of the whole data. After the verification of the scheme to these two subdomains we applied the scheme to the whole domain. We chose two subdomains where the distributions of the data are relatively uniform compared with other domains. The area I is an eastern area and the area II is a northern area.

We used the small number of knots for x and y directions respectively because of the irregularity of locations of samples. We tested from 10 to 18 knots for x and y directions respectively. And for every combination of the number of knots, we generated 100 sets of the uniformly randomized x and y coordinates according to the density of samples respectively. But some of them didn't satisfy the Schoenberg-Whitney condition then we generated another set of coordinates again. Furthermore the equations of matrices made from ill conditioned coordinates could not be solved properly then we also generated other coordinates of knots again. After solving the equations (3.25), we have applied generalized cross-validation with influence function GCV_{IF}[2] as the information criterion.

The results are shown in Tables 4.1,4.2,4.3,4.4. The optimal models are shown in Figures 4.3,4.5. For the small values of β the selected models are not adequate. Those are shown in Figures 4.4,4.6.

The results of these criteria shows that the smaller values of β are

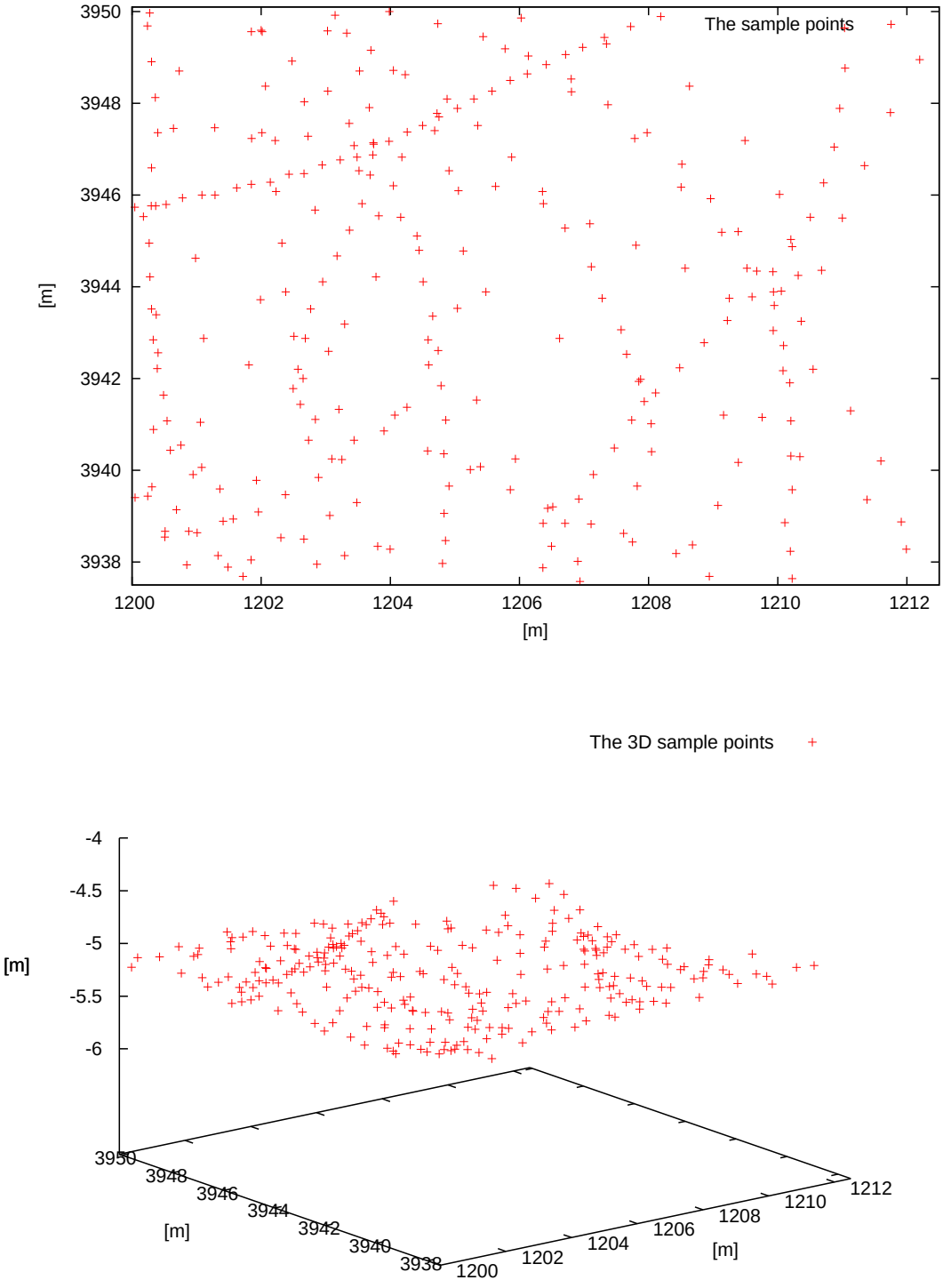


Figure 4.1: Distribution of data over area I

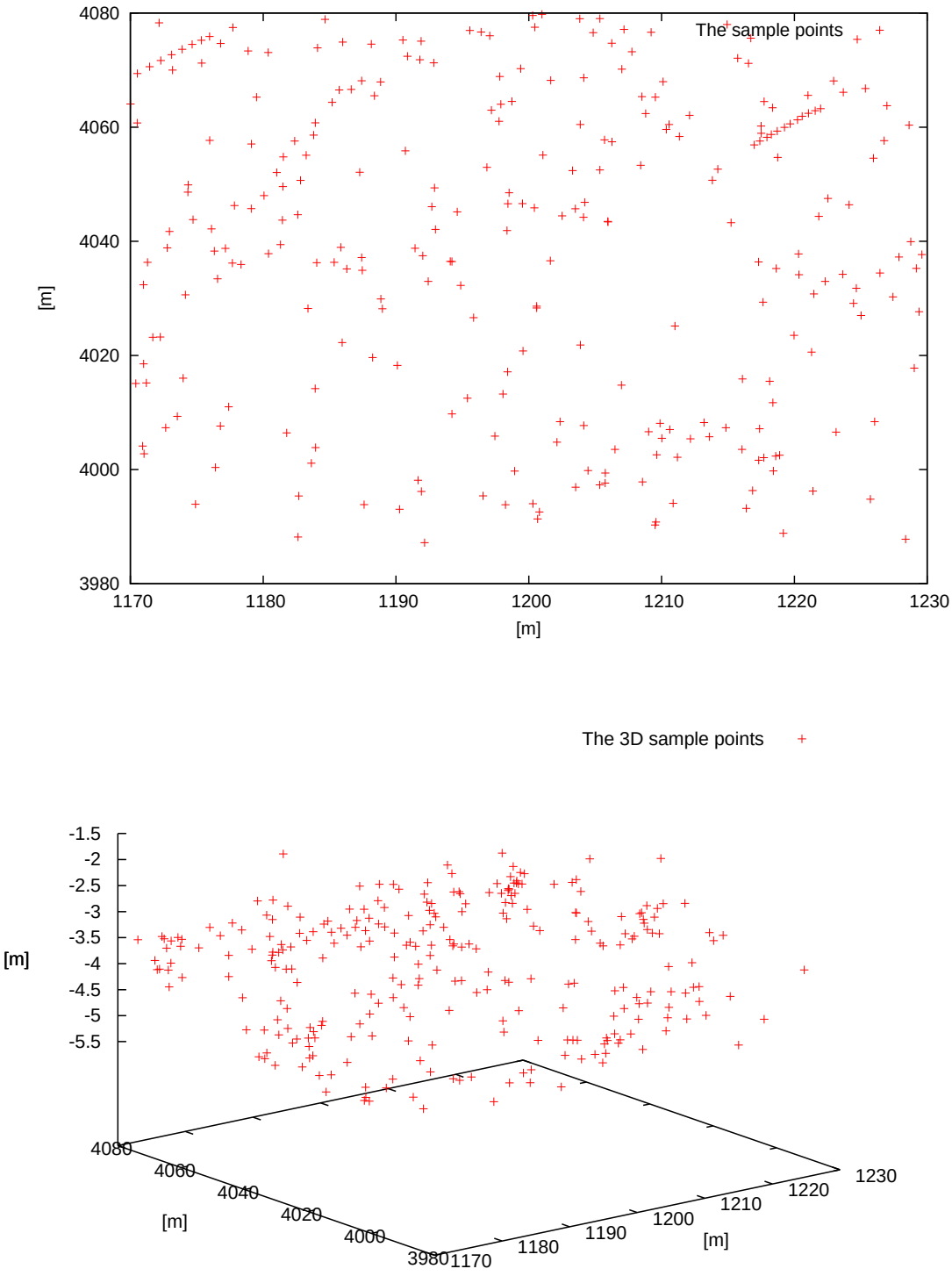


Figure 4.2: Distribution of data over area II

Table 4.1: CV results of area I

total number of knots					
x-axis	y-axis	β	σ^2	λ	CV
17	16	1.00E-00	0.003458	2.8918E+02	-766.40
11	**10**	**1.00E-01**	**0.003292**	**3.0377E+01**	**-776.76**
10	10	1.00E-02	0.003166	3.1586E+00	-776.52
10	10	1.00E-03	0.003118	3.2067E-01	-768.27
11	10	1.00E-04	0.002872	3.4819E-02	-768.20
11	10	1.00E-05	0.002665	3.7525E-03	-765.94
11	10	1.00E-06	0.002661	3.7580E-04	-761.93
10	10	1.00E-07	0.002878	3.4745E-05	-747.79
10	10	1.00E-08	0.002878	3.4746E-06	-746.17
10	10	1.00E-09	0.002878	3.4746E-07	-745.97
10	10	1.00E-10	0.002878	3.4746E-08	-745.95

Table 4.2: CV results of area II

total number of knots					
x-axis	y-axis	β	σ^2	λ	CV
16	18	1.00E+00	0.096383	1.0375E+01	199.88
17	10	1.00E-01	0.090808	1.1012E+00	203.61
11	10	1.00E-02	0.084684	1.1809E-01	215.32
12	12	1.00E-03	0.072609	1.3772E-02	219.93
12	11	1.00E-04	0.070483	1.4188E-03	195.25
12	**11**	**1.00E-05**	**0.070152**	**1.4255E-04**	**194.72**
12	11	1.00E-06	0.070117	1.4262E-05	196.03
12	11	1.00E-07	0.070116	1.4262E-06	196.97
12	11	1.00E-08	0.070116	1.4262E-07	197.12
12	11	1.00E-09	0.070116	1.4262E-08	197.13
12	11	1.00E-10	0.070116	1.4262E-09	197.13

Table 4.3: GCV_{IF} results for the area I

total number of knots					
x-axis	y-axis	β	σ^2	λ	GCV_{IF}
17	16	1.00E+00	0.0032969	3.0300E+02	-768.49
10	16	1.00E-01	0.0032059	3.1200E+01	-778.16
16	**10**	**1.00E-02**	**0.003125**	**3.1997E+00**	**-779.12**
10	10	1.00E-03	0.003118	3.2067E-01	-772.00
10	10	1.00E-04	0.003063	3.2646E-02	-772.00
10	10	1.00E-05	0.002968	3.3690E-03	-762.05
11	10	1.00E-06	0.002678	3.7343E-04	-761.34
10	10	1.00E-07	0.002726	3.6680E-05	-760.30
10	10	1.00E-08	0.002726	3.6680E-06	-760.23
10	10	1.00E-09	0.002726	3.6680E-07	-760.22
10	10	1.00E-10	0.002726	3.6680E-08	-760.22

Table 4.4: GCV_{IF} results for the area II

total number of knots					
x-axis	y-axis	β	σ^2	λ	GCV_{IF}
16	18	1.00E+00	0.096383	1.0375E+01	190.63
18	18	1.00E-01	0.072638	1.3767E+00	167.43
15	**14**	**1.00E-02**	**0.062674**	**1.5956E-01**	**158.61**
11	15	1.00E-03	0.061989	1.6132E-02	165.76
11	11	1.00E-04	0.070245	1.4236E-03	173.68
11	10	1.00E-05	0.077283	1.2940E-04	177.41
10	10	1.00E-06	0.077254	1.2944E-05	177.73
10	10	1.00E-07	0.077254	1.2944E-06	177.74
10	10	1.00E-08	0.077254	1.2944E-07	177.74
10	10	1.00E-09	0.077254	1.2944E-08	177.74
10	10	1.00E-10	0.077254	1.2944E-09	177.74

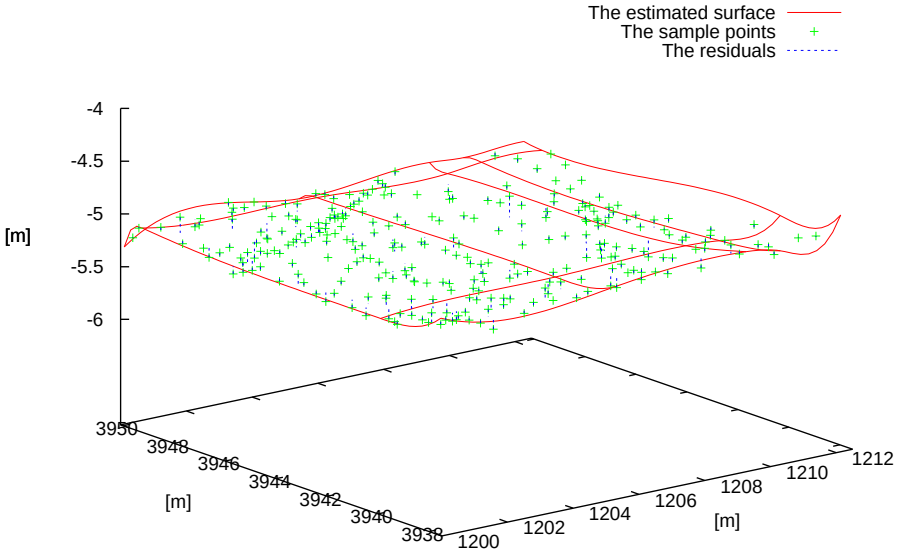
not important for the determination of the models.

4.2 Estimation of the Whole Area

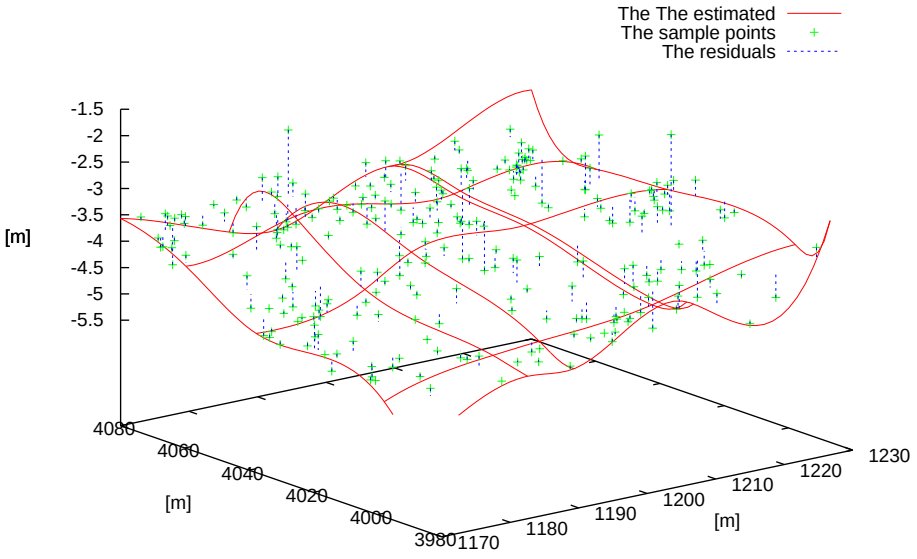
We tested the estimation with various β's which were from 10^{-1} to 10^{-7} according to the results of two subareas. For the evaluation of the obtained parameters we use the GCV_{IF} as the information criterion. We can determine the optimal parameters which minimize this GCV_{IF}. Those results are shown in the Table 4.5. And the optimal model is shown in Figure 4.8. Figure 4.9 shows the selected model for $\beta = 10^{-7}$ and that is not adequate in shape.

Table 4.5: GCV_{IF} results over the whole area

total number of knots					
x-axis	y-axis	β	σ^2	λ	GCV_{IF}
22	**19**	**1.00E-01**	**0.049689**	**2.0125E+00**	**292.61**
21	18	1.00E-02	0.051459	1.9433E-01	333.00
22	18	1.00E-03	0.066054	1.5139E-02	347.73
22	18	1.00E-04	0.050174	1.9931E-03	337.37
22	21	1.00E-05	0.095211	1.0503E-04	318.57
22	18	1.00E-06	0.042305	2.3638E-05	303.21
22	18	1.00E-07	0.041752	2.3951E-06	299.24

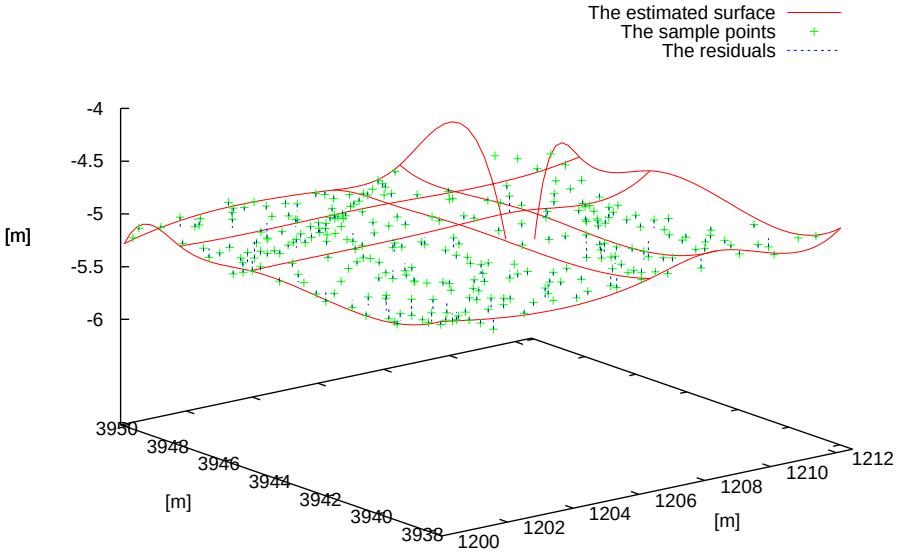


(a) area I ($\beta = 10^{-1}$)

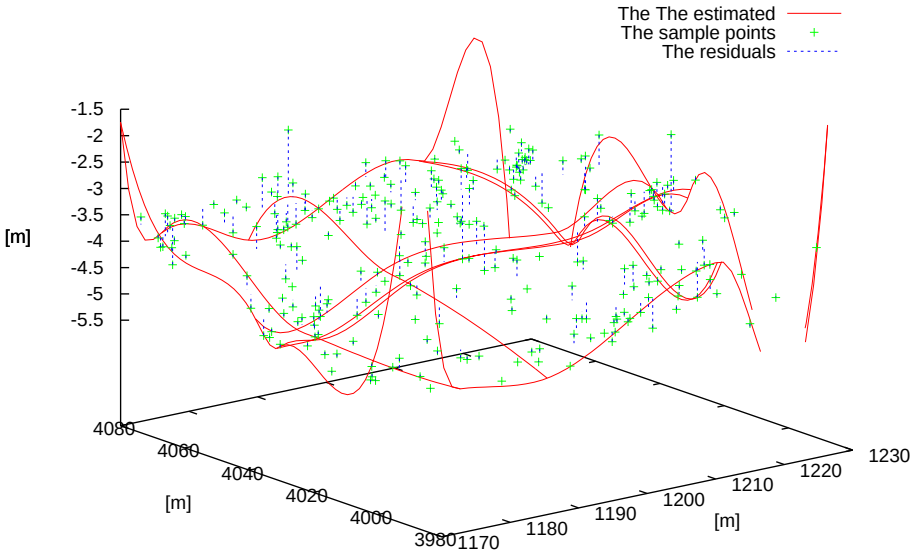


(b) area II ($\beta = 10^{-5}$)

Figure 4.3: Selected model(CV)

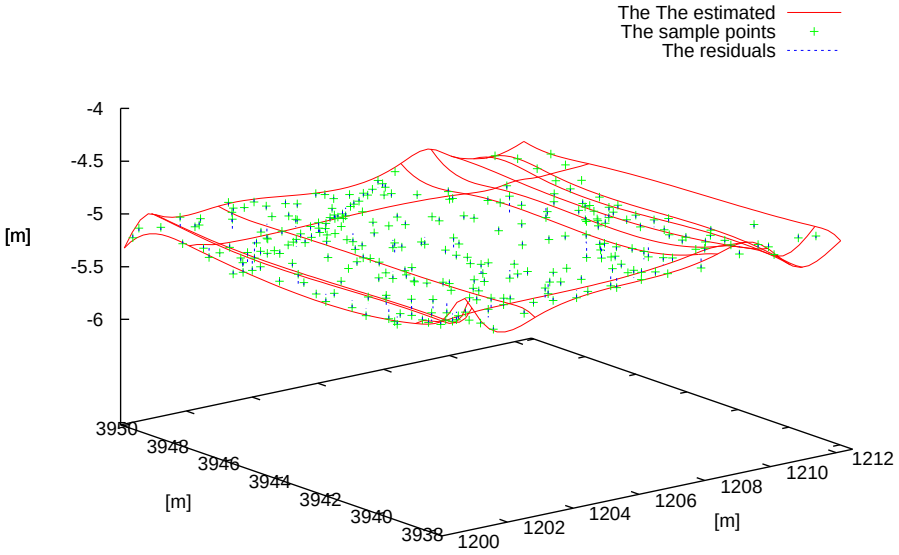


(a) area I ($\beta = 10^{-7}$)

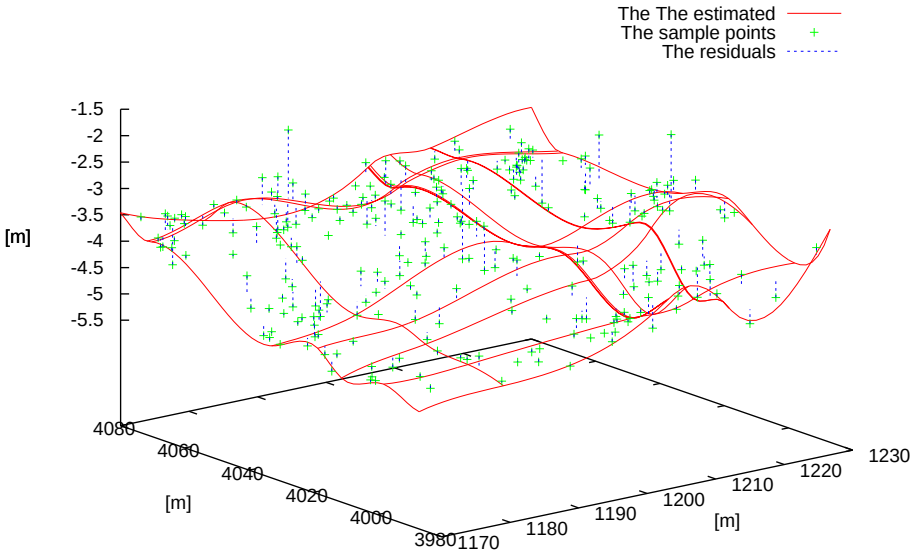


(b) area II ($\beta = 10^{-7}$)

Figure 4.4: Selected model(CV)

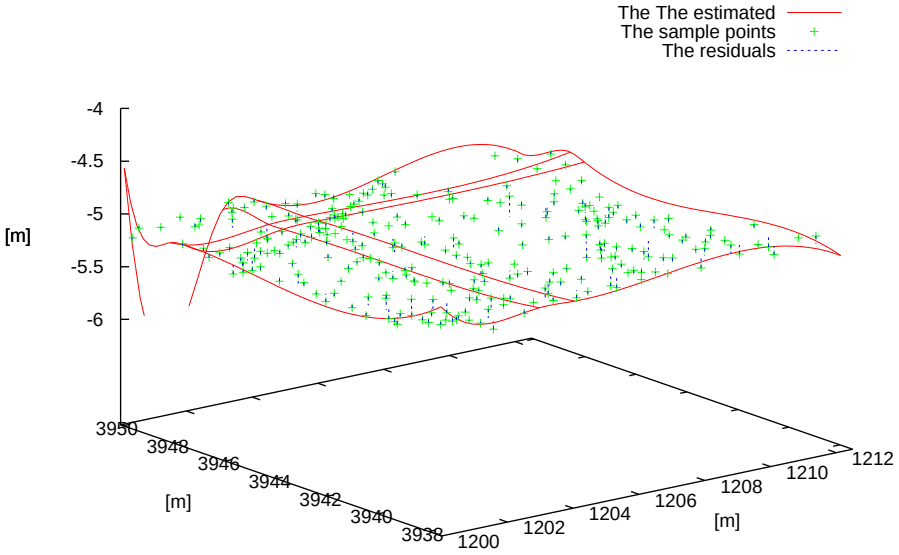


(a) area I ($\beta = 10^{-2}$)

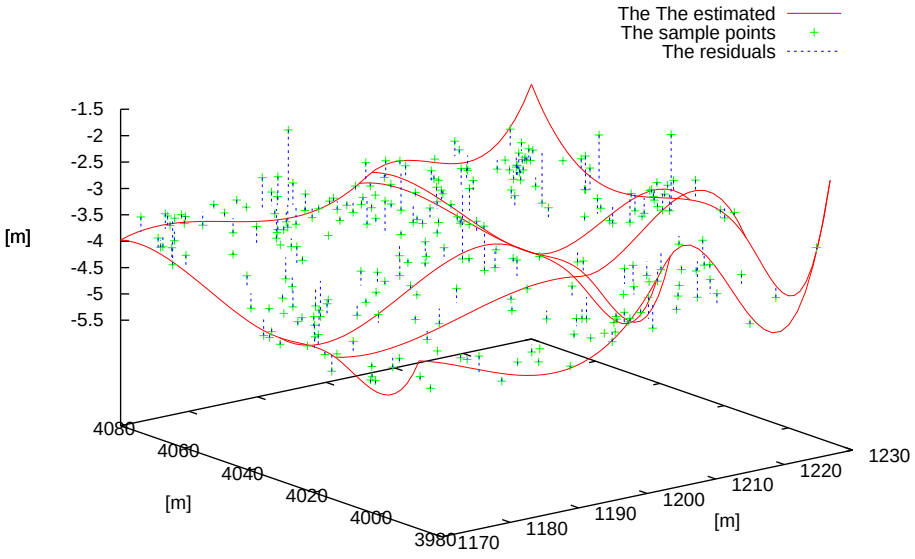


(b) area II ($\beta = 10^{-2}$)

Figure 4.5: Selected model(GCV_{IF})



(a) area I ($\beta = 10^{-7}$)



(b) area II ($\beta = 10^{-7}$)

Figure 4.6: Selected model(GCV_{IF})

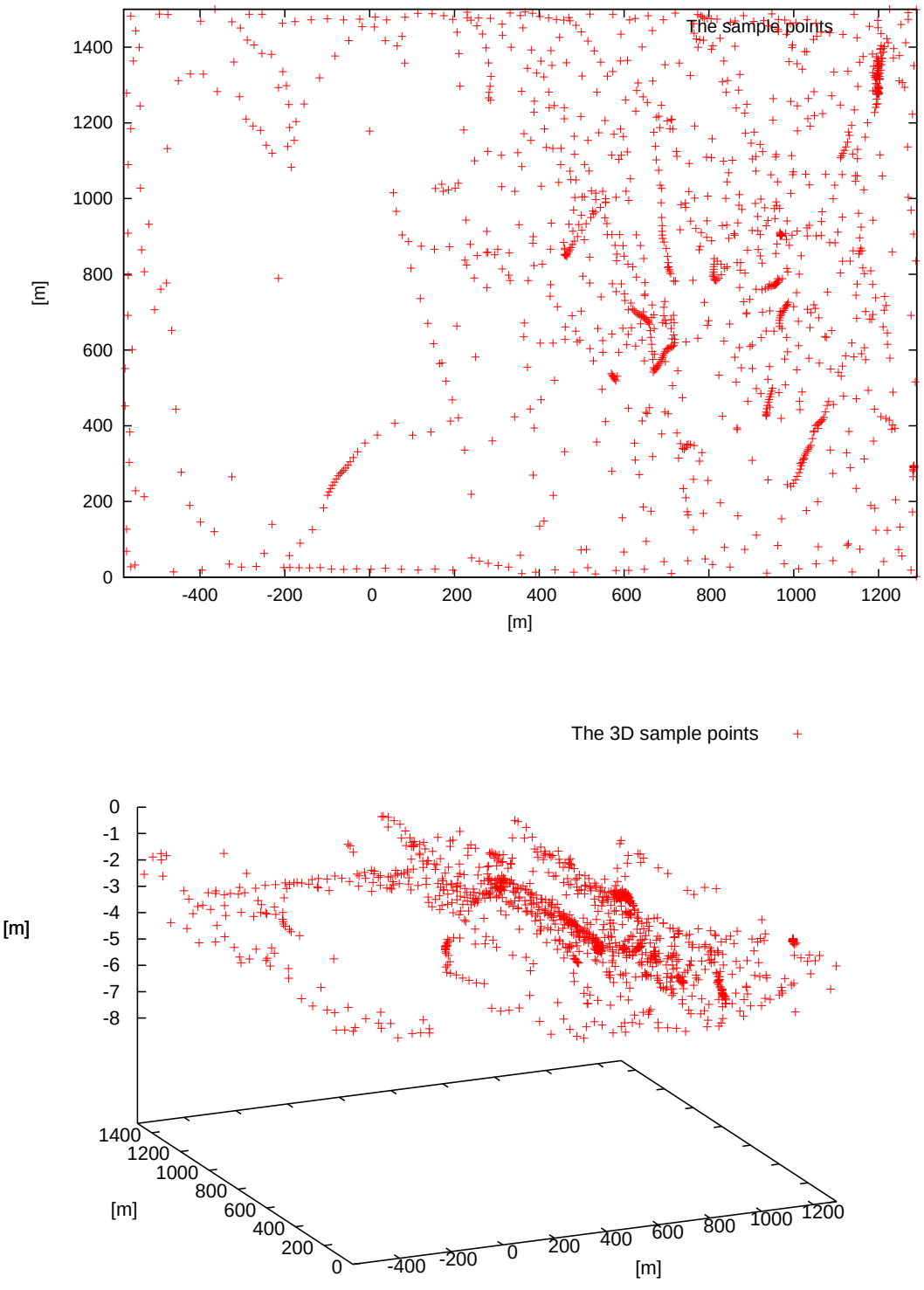


Figure 4.7: Distribution of data over whole area

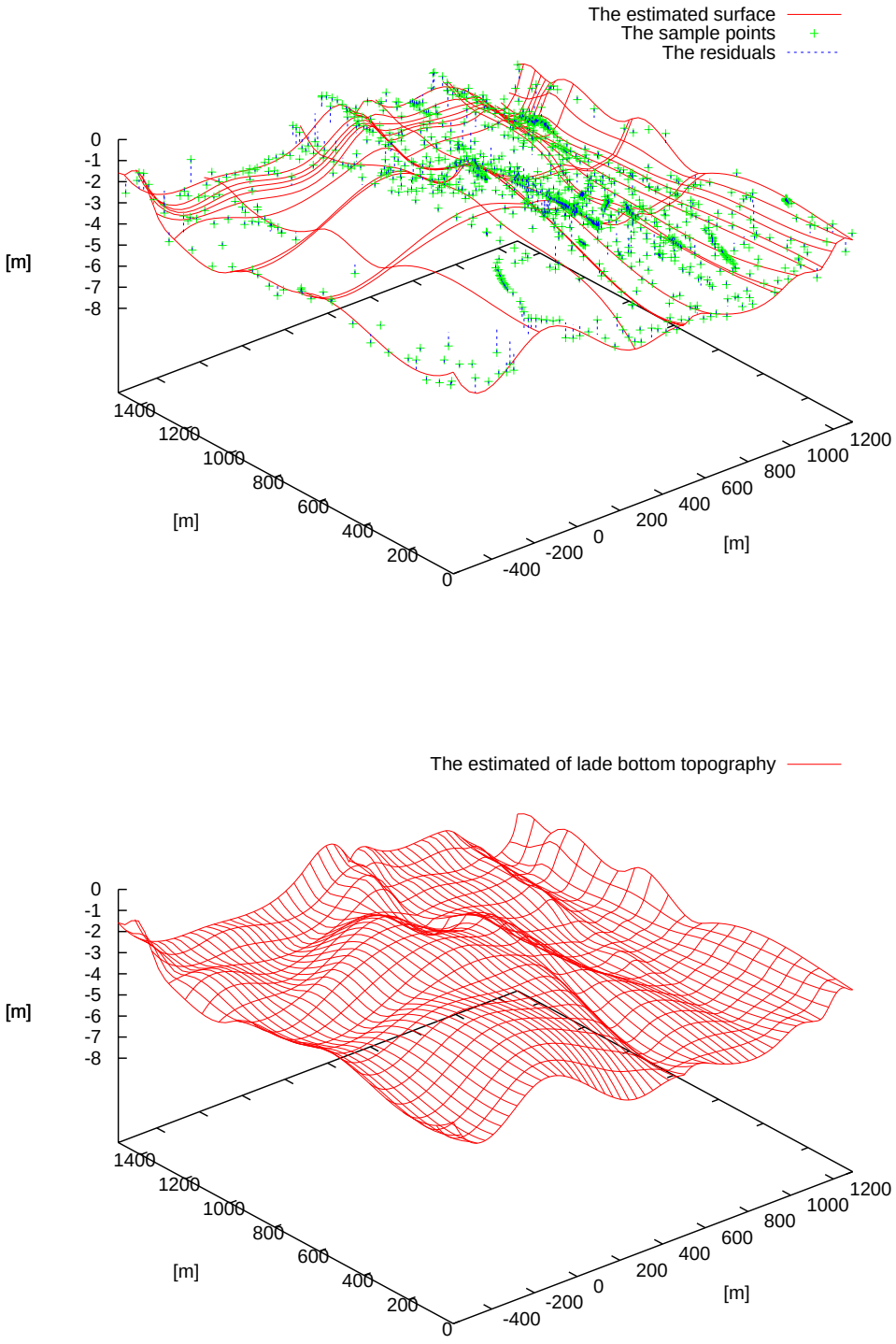


Figure 4.8: Estimation of lake bottom topography

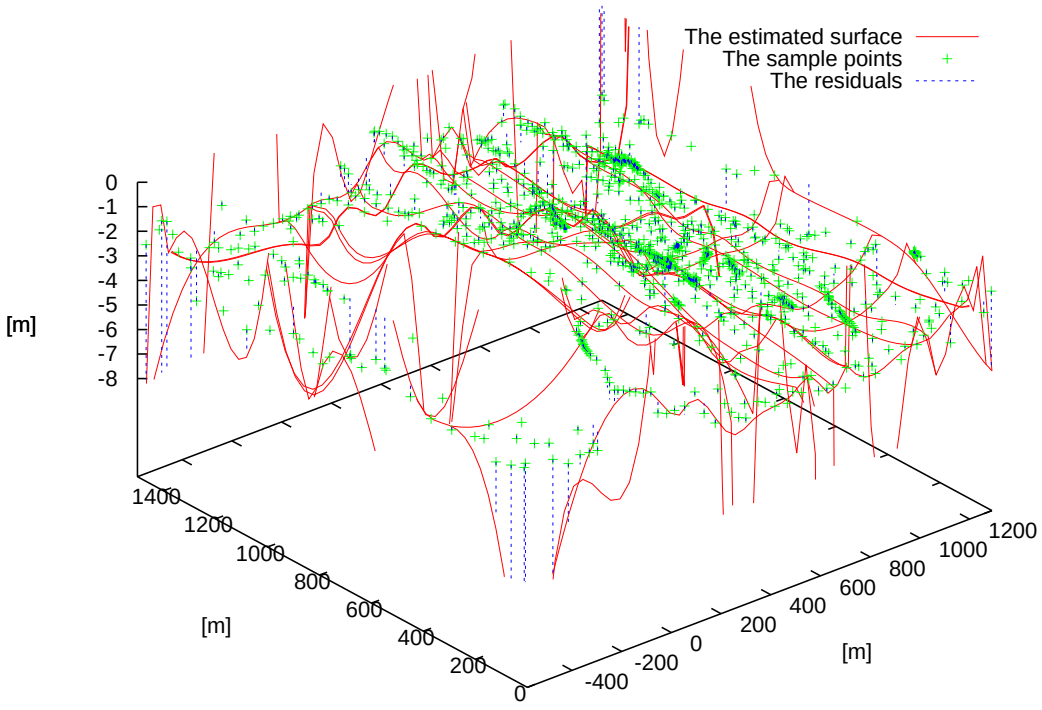


Figure 4.9: GCV_{IF} results over the whole area ($\beta = 10^{-7}$)

4.3 Numerical Simulation of Influence Function

There are two alternative methods of LOOCV. One estimates the value of $\hat{u}^{(-\alpha)}$ of (3.14) directly such as GCV_{IF}. The other estimates the values of parameters $\hat{\boldsymbol{\theta}}^{(-\alpha)}$ and it calculates $\hat{u}^{(-\alpha)}$ using those parameters such as mGIC. In section 3.5.4 we have written that mGIC does not make a good approximation of CV compared with GCV_{IF}. We simulate numerically where the errors occur at and how large those are.

4.3.1 Samples and Conditions

The sample area is area I of the lake in Section 4.1. The total number of data used in this simulation is 278. The total number of knots of B-spline along the x and y axis is 20 respectively.

$$u(x,y) = \sum_{i=1}^{20} \sum_{j=1}^{20} w_{ij} M_i(x) N_j(y), \tag{4.1}$$

where $M_i(x), N_j(y)$ are the spline functions with order four along the x and y axis respectively and w_{ij} are the coefficients of products of the spline functions.

The total number of estimated coefficients of the spline function is $(20-4) \times (20-4) = 256$. And the variance is also to be estimated. The result of surface estimation is shown in Fig. 1. The curved line is the estimated surface, cross points are the location of samples and vertical lines show the size of residuals.

In this simulation we set $\lambda\hat{\sigma}^2 = 10^{-5}$ as the experiment.

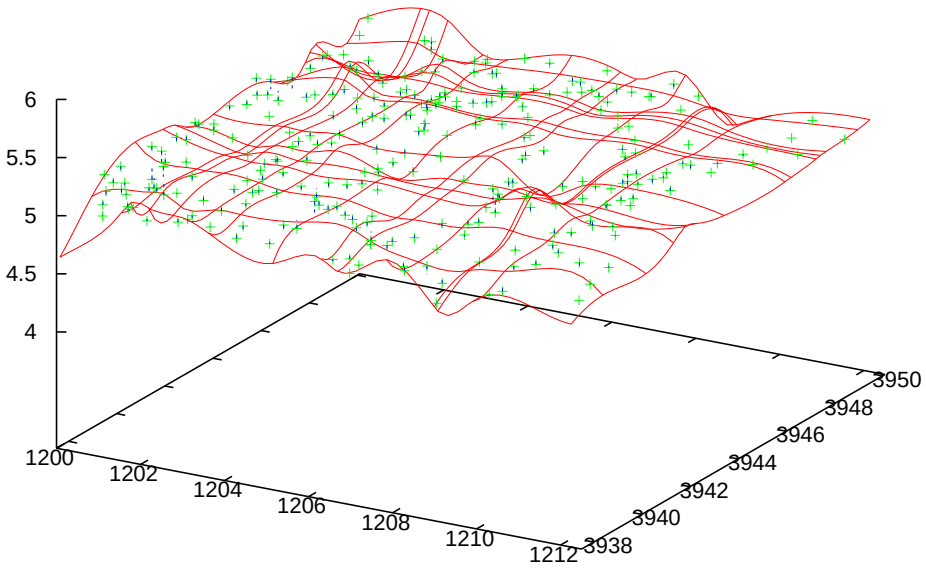


Figure 4.10: Estimated surface

4.3.2 Information Criteria

LOOCV is calculated as below

$$\begin{aligned} \mathrm{CV} &= -2\sum_{\alpha=1}^{n} \log(f(\boldsymbol{x}_\alpha, \hat{\boldsymbol{\theta}}^{(-\alpha)})) \\ &= \sum_{\alpha=1}^{n} \left\{ \log(2\pi\hat{\sigma}^{(-\alpha)^2}) + \frac{(z_\alpha - \hat{u}^{(-\alpha)})^2}{\hat{\sigma}^{(-\alpha)^2}} \right\}. \end{aligned} \tag{4.2}$$

And we adopt the next approximation for CV

$$\begin{aligned} \boldsymbol{T}(\hat{G}^{(-\alpha)}) &\approx \boldsymbol{T}(G) + \frac{1}{n-1}\sum_{i\neq\alpha}^{n}\boldsymbol{T}^{(1)}(z_i;G) \\ &\approx \boldsymbol{T}(\hat{G}) - \frac{1}{n}\boldsymbol{T}^{(1)}(z_\alpha;\hat{G}). \end{aligned} \tag{4.3}$$

In the equation(3.14) of CV we replace the $\hat{\boldsymbol{\theta}}^{(-\alpha)}$ with $\tilde{\boldsymbol{\theta}}_\alpha = \hat{\boldsymbol{\theta}} - \frac{1}{n}\boldsymbol{T}^{(1)}(z_\alpha;\hat{G})$ and its scheme is called as modified GIC (mGIC)[28]. The calculation is shown as below

$$\begin{aligned} \text{mGIC} &= -2\sum_{\alpha=1}^{n}\log(f(\boldsymbol{x}_\alpha, \hat{\boldsymbol{\theta}} - \frac{1}{n}\boldsymbol{T}^{(1)}(z_\alpha;\hat{G}))) \\ &= \sum_{\alpha=1}^{n}\left\{\log(2\pi\tilde{\sigma}_\alpha^2) + \frac{(z_\alpha - \tilde{u}_\alpha)^2}{\tilde{\sigma}_\alpha^2}\right\}, \end{aligned} \tag{4.4}$$

where $\tilde{\boldsymbol{\theta}}_\alpha = (\tilde{\boldsymbol{w}}'_\alpha, \tilde{\sigma}_\alpha^2), \tilde{u}_\alpha = u(\boldsymbol{x}_{\boldsymbol{\alpha}}|\tilde{\boldsymbol{w}}'_\alpha)$.

4.3.3 Numerical Result

The numerical result shows that the first terms (variance) of (3.14),(3.52) are almost same and the second terms (depth) of them are quite different.

Table 4.6: Comparison of CV and mGIC

	CV	mGIC
variance term	-1299.783	-1298.879
depth term	1227.449	544.995

About the variance term, the maximum difference is 0.024155 and the average is 0.003258. These are quite small. But about the depth term, there are large differences.

By the influence function we estimate not the depth but the parameter. We calculate 278 samples and every sample has 257 parameters. So the total number of parameters which we estimated is 71446. The ratios of the parameters estimated by CV and estimated by mGIC are calculated and those results are shown in Table 4.7. The average of the ratios is 0.99999 and the variance is 5.2410×10^{-6}.

Table 4.7: Ratio of the estimated parameters (mGIC / CV)

ratio	total number	percentage
0.995 - 1.005	70416	98.55%
0.990 - 1.010	71010	99.38%
0.975 - 1.025	71342	99.85%
0.950 - 1.050	71419	99.96%
all	71446	

The total number of parameters which have the differences larger than 5.0% is only 27. The smallest ratio is 0.87720 and the largest is 1.10598. The estimation of the parameter is quite accurate. But the estimation of the depth using these parameters cause large differences in some samples. Consequently the value of the information criterion has the large difference.

4.3.4 Difference between Smaples

About the depth term the half of 278 samples have the difference smaller than 0.71 and the sum of those differences is only 30.91 . But the rest samples have the large differences and the sum of them is 651.56 . The samples which have the large differences are distributed near the boundary area. Those locations are shown in Fig.4.11. The

maximum difference is 43.06839 and the average is 2.45486. The ratio of the parameters of the sample which has the largest difference are shown in Table 4.8.

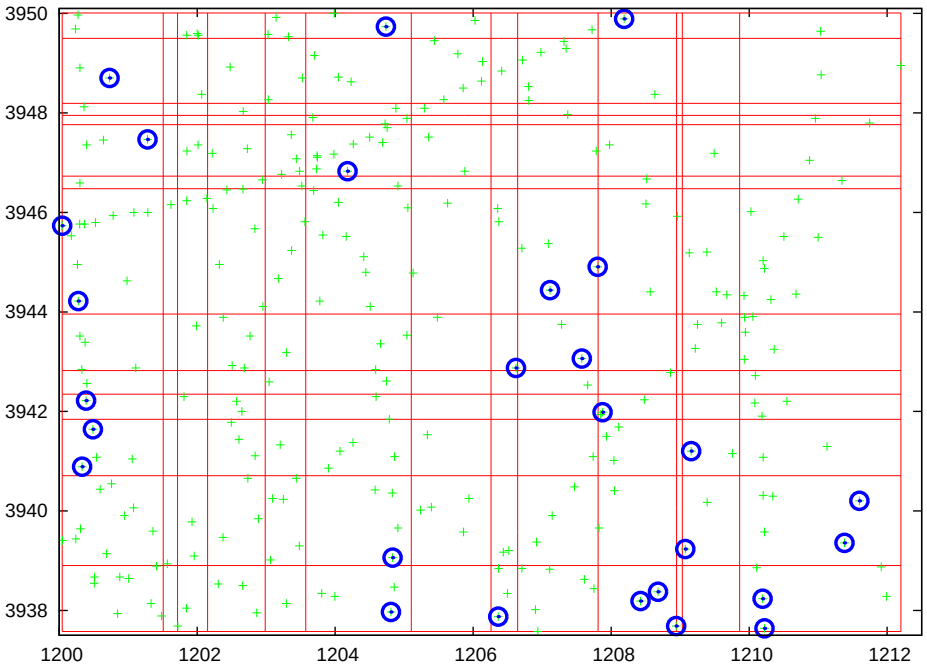


Figure 4.11: locations of samples with large errors

Table 4.8: ratio of parameters of the sample which has the largest difference

	average	maximum	minimum
all 257 parameters	0.99958		
valid 16 spline coefficients	0.99342	1.03259	0.88941
other 241 parameters	1.00002	1.00428	0.99701

4.3.5 Difference between Parameters

On the other hand the parameters which have large errors are shown in Table 4.9.

Table 4.9: Parameters with large variance

function number(i)	function number(j)	minimum ratio	maximum ratio	variance	standard deviation
3	15	0.87720	1.01827	0.0000735	0.0085736
16	12	0.99835	1.10571	0.0000514	0.0071676
1	14	0.88941	1.00357	0.0000499	0.0070617
8	11	0.98520	1.09219	0.0000455	0.0067451
4	16	0.97364	1.10598	0.0000445	0.0066704
4	15	0.98694	1.08300	0.0000363	0.0060287
3	14	0.95697	1.04481	0.0000330	0.0057406
15	12	0.99693	1.08498	0.0000324	0.0056886
2	14	0.97906	1.07677	0.0000296	0.0054376
3	15	0.99109	1.07392	0.0000288	0.0053688

The function number in Table 4.9 means the number of spline function in (4.1). The supports of the spline functions related with the coefficients in Table 4.9 are shown in Fig. 4.12. Those areas are distributed southeast mainly.

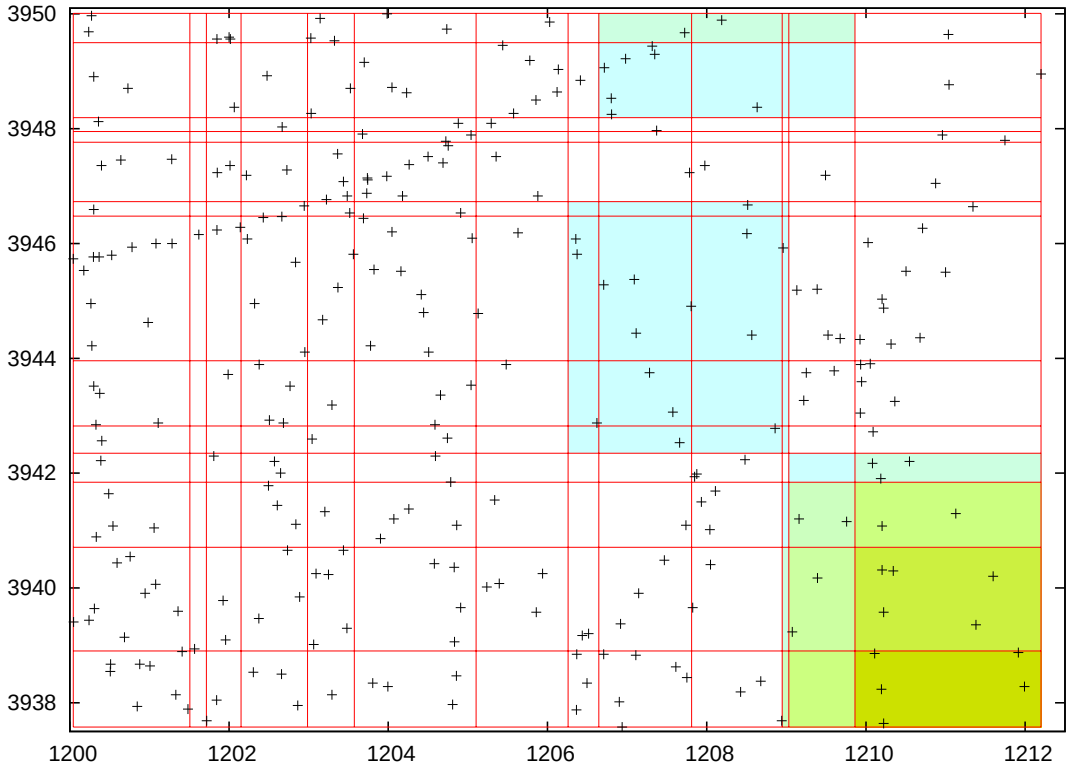


Figure 4.12: Supports of Spline function related with the sample points in Table 4.9

Chapter 5

Effective Scheme

5.1 Evaluation of the Model

For the B-spline the total number and the locations of the knots are important. When the knots at the both ends are four-folded, the least total number of knots is ten along x and y axis respectively. The total number of parameters is $(n_x - 4) \times (n_y - 4) + 1$ which is consist of the coefficients of the basis and the variance, and this value should be less than the total number of the sample points at least. We have set 300 sample points and the adequate total number of knots along every axis is less or equal to 20. And for every (n_x, n_y), we prepared 100 sets of randomized knots generated uniformly. But some of them don't satisfy the Schoenberg-Whitney condition [23] then we generate other set of knots again. Furthermore the equations of matrices made from ill conditioned sets cannot be solved properly then we also generate another sets of knots again so we tested 100 solvable sets for every (n_x, n_y). We denote a set of knots as $S_{j,k}\,(j = 1, 2, \cdots, 121,\ k = 1, 2, \cdots, 100)$ where $j = 11(n_x - 10) + (n_y - 10) + 1$ represents the number of knots and k represents the number of the sets which has the same total number of knots. On the other hand, we tried eight values of the smoother parameter β which is the coefficient of the penalized term. We set the values of β from 10^{-1} to 10^{-8}, so we considered 96,800 models and determined the parameters of those by the regulaization method. The evaluations of the models are done by the value of CV. We use the log-likelihood for Cross-Validation(CV) as

$$\begin{aligned} \mathrm{CV} &= -2\sum_{\alpha=1}^{n}\log(f(\boldsymbol{x}_\alpha,\boldsymbol{\theta}^{(-\alpha)})) \\ &= \sum_{\alpha=1}^{n}\left\{\log(2\pi\hat{\sigma}^{2(-\alpha)})+\frac{(z_\alpha-\hat{u}^{(-\alpha)})^2}{\hat{\sigma}^{2(-\alpha)}}\right\} \end{aligned} \tag{5.1}$$

where $\hat{\sigma}^{2(-\alpha)}$, $\boldsymbol{\theta}^{(-\alpha)}$, $\hat{u}^{(-\alpha)}$ are determined by the data without α'th sample.

We only tried quite rough values of β. The result is shown in Table 5.1-5.2.

Table 5.1: All values of CV by previous method

n_x	n_y	j	k	β							
				10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}
10	10	1	1	628.70	619.04	572.46	484.71	409.37	385.67	424.62	446.15
10	10	1	2	628.79	619.96	581.09	506.56	464.77	470.70	502.74	514.65
	⋮										
10	10	1	100	628.79	619.92	579.63	507.01	464.46	453.35	480.04	487.96
10	11	2	1	628.58	617.90	565.17	463.64	351.10	339.66	402.22	414.08
10	11	2	2	628.95	621.43	590.44	534.84	523.31	541.29	581.61	645.75
	⋮										
10	11	2	100	628.85	620.57	585.10	509.99	469.65	471.24	482.55	486.20
	⋮										
20	20	121	1	628.55	617.59	561.84	415.92	106.62	-145.01	367.30	3028.74
	⋮										
20	20	121	99	628.57	617.81	564.13	438.16	216.48	-10.95	127.41	1393.35
20	20	121	100	628.56	617.74	563.23	425.62	152.27	-32.05	173.46	3192.88

In order to determine better model we need more accurate value of β. But it will take many time to calculate the CV's. We studied

to get more efficient scheme. In these models, there are many sets of knots and β's which are not useful to determine the best model. We save the time for calculation and tried to obtain more correct value of β and verify the validity of our scheme.

Table 5.2: Minimum value of CV for each β by previous method

n_x	n_y	j	k	β	σ^2	λ	CV
13	13	37	19	10^{-1}	0.003851	2.59673E+01	31.07
18	18	97	37	10^{-2}	0.053463	1.87045E-01	40.25
20	15	66	92	10^{-3}	0.322460	3.10116E-03	537.93
17	20	118	95	10^{-4}	0.165861	6.02913E-04	367.40
20	15	66	92	10^{-5}	0.048688	2.05391E-04	74.36
16	14	51	21	10^{-6}	0.006367	1.57050E-04	-346.11
16	**14**	**51**	**21**	10^{-7}	**0.002303**	**4.34181E-05**	**-441.16**
19	13	43	4	10^{-8}	0.002769	3.61129E-06	-244.25

5.2 Optimal Smoother Parameter

5.2.1 Interpolation by Spline Function

For every set of knots $S_{j,k}$, we have only eight values of CV for $\beta = 10^{-1}, 10^{-2}, \cdots, 10^{-8}$. Fig. 5.1 shows the interpolation of the value of CV by the spline function.

Let $x_i = -\log_{10}\beta_i$ and spline interpolation functions $f_i(x) = a_i(x - x_i)^3 + b_i(x - x_i)^2 + c_i(x - x_i) + d_i$, $(i = 1, 2, \cdots, N-1)$ are defined over the interval $[x_i, x_{i+1}]\,(i = 1, 2, \cdots, N-1)$. We determine the coefficients of these functions as follows.

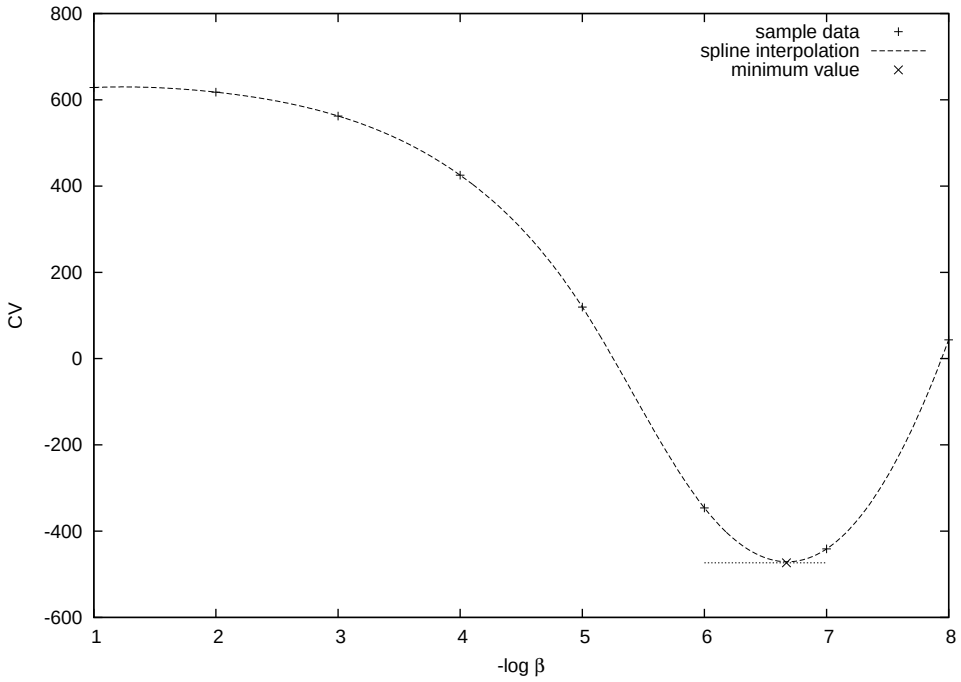


Figure 5.1: Spline Interpolation

$$
\begin{aligned}
f_i(x_{i+1}) &= f_{i+1}(x_{i+1}) = y_{i+1} \\
f_i{}'(x_{i+1}) &= f_{i+1}{}'(x_{i+1}) \\
f_i{}''(x_{i+1}) &= f_{i+1}{}''(x_{i+1}) \quad \text{for } i = 1, 2, \cdots, N-2
\end{aligned}
$$

Directly we find $d_i = y_i \, (i = 2, 3, \cdots, N-1)$.
The boundary condition is the natural spline

$$
\begin{aligned}
f_1(x_1) &= y_1 \\
f_{N-1}(x_N) &= y_N \\
f_1{}''(x_1) &= f_{N-1}{}''(x_N) = 0.
\end{aligned}
$$

Based on these conditions we can obtain the coefficients by solving the next equation of matrix.

$$\begin{bmatrix} 2(h_1+h_2) & h_2 & 0 & \cdots & 0 \\ h_2 & 2(h_2+h_3) & h_3 & \cdots & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & \cdots & h_{N-2} & 2(h_{N-2}+h_{N-1}) \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{N-2} \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_{N-2} \end{bmatrix}$$

where $h_i = x_{i+1} - x_i, v_i = 3\left(\frac{y_{i+1}-y_i}{h_i} - \frac{y_i - y_{i-1}}{h_{i-1}}\right)$. We can determine the all variables from $\{b_i\}$ as follows

$$a_i = \frac{b_{i+1} - b_i}{3h_i},\ c_i = \frac{y_{i+1} - y_i - a_i h_i^3 - b_i h_i^2}{h_i}.$$

To get the value of x which gives the minimum value of function. The spline functions are only three dimensional polynomials, we can easily differentiate and calculate the zero points of them.

$$f_i{}'(x) = 3a_i(x - x_i)^2 + 2b_i(x - x_i) + c_i = 0$$

Let $\tilde{x}_j$ be the zero point for every j. Instead of calculating CV's for all sets $S_{j,k}\,(j = 1, 2, \cdots, 121,\ k = 1, 2, \cdots, 100)$, we select only one $k = k_j$ for every j, we calculate CV's for selected sets $S_{j,k_j}\,(j = 1, 2, \cdots, 121)$. In those sets we make the spline interpolation and determine the minimum estimated values $m_j\,(j = 1, 2, \cdots, 121)$ and $\tilde{x}_j$ which gives m_j for every set. Among them we determine the minimum value of m_j and $\tilde{x}_j$ which gives it. We denote those values as $m_{\min}$ and $x_{\min}$. Then we obtain $\beta_{\min} = 10^{-x_{\min}}$. Using this $\beta_{\min}$, we estimate the coefficients of the estimated surface and calculate CV for all sets $S_{j,k}\,(j = 1, 2, \cdots, 121,\ k = 1, 2, \cdots, 100)$.

The value of $\beta_{\min}$ that is $10^{-x_{\min}}$ varies depend on the selection of k_j, the statistical values of $x_{\min}$ based on 100 experiments are as follows.

Table 5.3: Minimum value of CV for each j

n_x	n_y	j	$\tilde{x}_j$	m_j
10	10	1	5.824489	384.30
10	11	2	5.585633	330.07
	⋮			
20	19	120	6.232998	-150.10
20	20	121	6.200994	-159.55

Table 5.4: Statistical values of $x_{\min}$

average	6.56285
median	6.55971
standard deviation	0.102462
maximum	6.85326
minimum	6.32370

The result of the calculation based on the set of various values of $\beta_{\min}$ which include maximum one and minimum one shows that the same model $S_{71,21}$ is the best.

Table 5.5: Model evaluation for $\beta_{\min} = 10^{-6.85326}$

n_x	n_y	j	k	λ	CV
16	14	71	21	0.00240204	-462.824
15	14	60	80	0.00369185	-429.770
17	14	82	79	0.00271836	-425.079
	$\vdots$				

Table 5.6: Model evaluation for $\beta_{\min} = 10^{-6.67209}$

n_x	n_y	j	k	λ	CV
16	14	71	21	0.00261375	-473.446
15	14	60	80	0.00391828	-457.147
16	15	72	99	0.00325389	-429.550
	$\vdots$				

5.2.2 Estimation of the Optimal β

After the determination of the best set of knots, we have to determine the optimal value of smoother parameter β. We have only estimated vague value of β. To obtain the best value of CV we calculate it based on various values of β only on the best set selected above. The range of β is set from $b-3\sigma$ to $b+3\sigma$, where b is the mean of best ten values of $\tilde{x}_j$ in Table 5.3 and σ is the standard deviation of them. The result of calculations is shown in Table 5.9.

Table 5.7: Model evaluation for $\beta_{\min} = 10^{-6.53701}$

n_x	n_y	j	k	λ	CV
16	14	71	21	0.00288071	-469.079
15	14	60	80	0.00420413	-457.276
16	15	72	99	0.00356967	-441.773
	⋮				

Table 5.8: Model evaluation for $\beta_{\min} = 10^{-6.32370}$

n_x	n_y	j	k	λ	CV
16	14	71	21	0.00364017	-439.822
16	15	72	99	0.00438565	-429.780
15	14	60	80	0.00502962	-421.374
	⋮				

Table 5.9: Values of CV for randmized β

$-\log\beta$	β	λ	CV
6.66160	2.1796×10^{-7}	0.00263041	-473.493
6.65547	2.2106×10^{-7}	0.00264043	-473.490
6.67365	2.1200×10^{-7}	0.00261131	-473.435
6.64192	2.2807×10^{-7}	0.00266333	-473.407
6.64123	2.2843×10^{-7}	0.00266452	-473.400
	⋮		

Chapter 6

Conclusion

For the estimation of a surface we adopt the regularized method. In this method the coefficient of the penalized term is important. The value of β in 3.3 plays an important role. As the value of β decreases, the residual variance reduces and the information criterion GIC_P also reduces monotonously. Then the GIC_P cannot determine the optimum model in both surfaces given in 3.5. The method of minimizing the cross-validation (CV) can determine the optimal values for two surfaces. The result of computation shows the excellence of the criterion CV. But it requires a large amount of computation. For the parameter estimation the alternative method mGIC by the information function works well. But the total number of parameters is so many that occasionally the estimated values are quite different from the sample value. Those samples make the value of mGIC worse and consequently we cannot determine the optimum model by this criterion in both surfaces. To overcome this difficulty the GCV is quite useful. To improve the property of GCV we use the influence function to estimate the variance of $n-1$ samples. We can recognize the superiority of GCV_{IF} which can determine the optimum model and can approximate the distribution of CV very well and it requires small amount of computation.

We propose an improved GCV criterion GCV_{IF}. This conclusion is the theory and obtained through a large number of simulation tests. From the results of these tests of GCV_{IF} criterion on surface I and surface II we can see that the GCV_{IF} criterion is more stable than CV, GIC and mGIC, also we can see that the GCV_{IF} criterion includes their informations.

In the actual measurement data some of information criteria could not determine the optimal parameters. In particular, the solutions to small β's did not determine the shape of surfaces adequately. The major reason of those difficulties is considered as the irregularity of the distribution of data. If there is little data near the boundary of the domain, even if the surface changes sharply, the value of criterion will seldom be influenced. In spite of those difficulties CV and the generalized CV with information function (GCV_{IF}) can determine the optimal values to the various sets of data. The computational cost of GCV_{IF} is 1/50 of CV. Furthermore the selected optimal model by GCV_{IF}(Figure 4.5) is better than that by CV(Figure 4.3). We can assert that GCV_{IF} is just practical method. This approximation method is able to contribute for the improvement of the water quality of Kojima Lake. We can predict the values of spline coefficients of LOOCV using the influence function of order one. But the difference of the information criteria is quite large. In this paper we show the result of computation and the distribution of errors of samples or parameters. It will be our future work to clarify the reason of these errors.

Furthemore we propose a new scheme which can reduce the amount of calculation to almost one eighth. At first we only use one set of each number of knots. Although the estimated value of $\beta_{\min}$ depends on the selection of sets, we need not mind the difference. Because the standard deviation of $x_{\min}$ is quite small. Based on that $\beta_{\min}$ we can determine the best set of knots. And the determination does not depend on the value of estimated $\beta_{\min}$.

To the selected optimal set of knots we finally obtain the optimal smoother parameter β by randomization. And the effective digit of it becomes higher than the previous method. It is our future study how to set the appropriate range of β in the final step.

References

[1] Akaike H., "A new look at the statistical identification model", *IEEE Transactions on Automatic Control*, **19(6)**(1974), pp.716-723.

[2] Bao, H. and Fueda, K., "A New Method for the Model Selection in B-spline Surface Approximation with an Influence Function", *Science Journal of Applied Mathematics and Statistics*, **1(5)**(2013), pp.38-46.

[3] Bao, H. and Fueda, K., "A Method for Topographical Estimation of Lake Bottoms by B-spline Surface", *American Journal of Theoretical and Applied Statistics*, **2(4)**(2013), pp.102-109.

[4] Bao, H. and Fueda, K., "An Effective Scheme for Estimating a Smoother Parameter in the Method of Regularization", *Applied and Computational Mathematics*, **2(6)**(2013), pp.118-123.

[5] Bao, H. and Fueda, K., "Numerical Evaluation of the Approximation by an Influence Function", *Journal of the Faculty of Environmental Science and Technology Okayama University* **19**(2014) (submitted to).

[6] Cox, M.G., "The numerical evaluation of B-splines", *J. Inst. Math. Appl.*, **10**(1972), pp.134-149.

[7] Cox, M.G., "An algorythm for spline interpolation", *J. Inst. Math. Appl.*, **15**(1975), pp.95-108.

[8] de Boor, C., "On calculation with B-splines", *J. Approx. Theory*, **6**(1972), pp.50-62.

[9] Good, I. J. and Gaskins, R.A., "Non parametric roughness penalties for probability densities", *Biometrika*, Vol. **58**(1971). pp. 255-277

[10] Good, I. J. and Gaskins, R.A., "Density estimation and bump hunting by the penalized likelihood method exemplified by scattering and meteorite data", *Journal of American Standard Association*, Vol. **75**(1980). pp. 42-56

[11] Green, P. J., Silverman, B. W., "Nonparametric Regression and Generalized Linear Models", Chapman and Hall, London(1994).

[12] Hampel, F.R. , "Contributions to the theory of robust estimation", Ph.D. Thesis, University of California, Barkeley (1968).

[13] Hampel, F.R. , "The influence curve and its role in robust estimation", *J. Amer. Statist. Assoc.*, **62**(1974b), 1179-1186.

[14] Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals (with discussion). *Ann. Statist.*, **16,** 927-985.

[15] Hall, P. (1992).*The Bootstrap and Edgeworth Expansion.* Springer-Verlag: London.

[16] Konishi, S., Kitagawa, G. (1996). "Generalised information criteria in model selection.", *Biometrika*, **83**, 875—890.

[17] Konishi, S., Kitagawa, G. (2008). "Information Criteria and Statistical Modeling", Springer Science+Business Media, LLC.

[18] Lawless, J. F. and Fredette, M. (2005). Frequentist prediction intervals and predictive distributions. *Biometrika* **92,** 529-542.

[19] Loader, C. R. (1999).*Local regression and likelihood.* Springer, New York.

[20] Meeker, W. Q. and Escobar, L. aAAaaaA

[21] Nonaka, T. and Konishi, S. (2005). Nonlinear regression modeling using regularized local likelihood method. *Ann. Inst. Statist. Math.***57,** 617-635.

[22] Polansky, A. M. and Schucany, W. R. (1997). Kernel smoothing to improve bootstrap xonfidence intervals.*J. Roy. Statist. Soc. Ser. B* **59,** 821-831.

[23] Schoenberg I. J., Whitney A., "On Pólya frequency functions III", *Trans. Amer. Math. Soc*, Vol. **74**(1953). pp. 246-259, pp. 246-259

[24] Schwarz G., "Estimating the dimension of a model", *Annals of Statistics*, **6(2)**(1978), pp.461-464.

[25] Stone,M., "Cross-validatory choice and assessment of statistical predictions (with discussion)", *Journal of the Royal Statistical Society*, Series B, 36 (1974), 111—147.

[26] Sweeting,T.J(1995). A framework of Bayesian and likelihood approximations in statistics. *Biometrika*, **82**, 1-23.

[27] Sakamoto,Y., Ishiguro, M. and Kitagawa,G.(1994).*Information statistics.* Kyoritsu Shuppan (in Japanese).

[28] Ueki, M. and Fueda, K., "Optimal Tuning Parameter Estimaiton In Maximum Penalized Likelihood Method", *Annals of the Institute of Statistical Mathematics*, **62** (2010), 413-438.

[29] Ueki, M. and Fueda, K., "Over close model problem and a modification of information criteria", Proceedings of The 28th symposium of Japanese Society of Applied Statistics, 7-10(2006). (in Japanese)

[30] Umeyama, S., (1996). "Discontinuity extraction in regularization using robust statistics", *Technical report of IEICE.*,PRU95-217 (1996). pp. 9-16

[31] Von Mises, R., "On the asymptotic distribution of differentiable statistical function.", *Annals of Mathematical Statistics*,**18**(1947),309-348.

[32] Wand, M. P. (1999). On the optimal amount of smoothing in penalised spline regression. *Biometrika*, **86**,936—940.

[33] Withers, C. S.,"Expansions for the distribution and quantiles of a regular functional of the empirical distribution with applications to nonparamteric confidence intervals.",*Annals of Mathematical Statistics*(1983),**11**(2),(1983),577-587.

[34] Yuen K.V., Bayesian methods for structural dynamics and civil engineering", John Wiley and Sons, NJ (2010).

Acknowledgements

First above all, I would like to express my sincere gratitude to Associate Professor Kaoru Fueda for his warm guidance and insightful advice throughout my graduate studies at Okayama University. My appreciation goes to for his useful advice and kindness.

I am also grateful to Professor Koji Kurihara and Associate Professor Masaya Iizuka who are my co-supervisor.And, I would like to thank Professor Tomoyuki Tarumi, Professor Wataru Sakamoto, Professor Masaji Watanabe and all the members of statistics seminar group who helped my university life.

Finally, to my husband and my son I wish to thank them for helping my university life. Without them, this thesis could not be completed.