Science Journal of Applied Mathematics and Statistics
2013; 1(5): 38-46
Published online October 20, 2013 (http://www.sciencepublishinggroup.com/j/sjams)
doi: 10.11648/j.sjams.20130105.11

A new method for the model selection in *B*-spline surface approximation with an influence function

Hongmei Bao[1], Kaoru Fueda[2]

[1]Graduate School of Environmental Science Okayama University, Okayama Japan
[2]Graduate School of Environmental and Life Science Okayama University, Okayama Japan

Email address:
gev421256@s.okayama-u.ac.jp(H. Bao), fueda@ems.okayama-u.ac.jp (K. Fueda)

To cite this article:
Hongmei Bao, Kaoru Fueda. A New Method for the Model Selection in *B*-Spline Surface Approximation with an Influence Function. *Science Journal of Applied Mathematics and Statistics.* Vol. 1, No. 5, 2013, pp. 38-46. doi: 10.11648/j.sjams.20130105.11

Abstract: In model selection, the most effective method requires much time.The analysis of the bivariate *B*-spline model with a penalized term has many difficulties.It has many factors and parameters such 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. The effect of the influence function and the techniques of the generalized cross validation (CV) are considered. The influence function is related to the first term of a Taylor expansion. Some alternative methods are tested and a new method is proposed. For the verification of this method theoretical proof and the computational results are shown.

Keywords: *B*-Spline Surface, Generalized Information Criterion, Influence Function, Generalized Cross-Validation, Cross-Validation, Kullback-Leibler Divergence, Surface Model Selection

1. Introduction

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

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

In order to determine the 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 chose the method of integration as the most favorable one.

An AIC-type criterion, which is the estimation of Kullback-Leibler divergence for the MPLE, is a generalized information criterion (GIC[1])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[2]) 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[3] was considered which utilizes the influence function. The first order influence function is useful because it requires less time. On the other hand the second order influence function takes too much time in comparison to its small profits. But mGIC is not sufficient to determine the optimal parameters. For better accuracy, we use the generalized CV (GCV[4]) and we proposed GCVIF, which is an improved GCV with the influence function. It is better for the model selection than the CV, AIC, GIC and mGIC. GCVIF is the criterion that includes the residual sum of squares, the number of samples 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 GCVIF.

2. 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:

$$\mathrm{KL}(p;q) = \int p(x)\log\frac{p(x)}{q(x)}dx$$

$$= \int p(x)\log p(x)dx - \int p(x)\log q(x)dx$$

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

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

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

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

where $q^{(-\alpha)}(x)$ is the probability 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 theα-th data point (z_α, x_α) is removed and the parameter vector $\theta = (w', \sigma^2)'$ is estimated based on the remaining *n*-1 observations. We denote the parameter as $\hat{\theta}^{(-\alpha)} = (\hat{w}^{(-\alpha)'}, \hat{\sigma}^{2(-\alpha)})'$ .The corresponding estimated regression function is denoted as $\hat{u}^{(-\alpha)}(x)$ .We use the log-likelihood for Cross Validation(IC_{CV}) as

$$IC_{CV} = -2\sum_{\alpha=1}^{n}\log\left(f\left(x_\alpha, \theta^{(-\alpha)}\right)\right)$$

$$= \sum_{\alpha=1}^{n}\left\{\log\left(2\pi\hat{\sigma}^{2(-\alpha)}\right) + \frac{\left(z_\alpha - \hat{u}^{(-\alpha)}\right)^2}{\hat{\sigma}^{2(-\alpha)}}\right\}. \quad (1)$$

Minimizing the equation (1) is the method of selecting the optimal model.Various alternative schemes are considered for the reduction of its computational cost, and another scheme is called the generalized CV (GCV)[4], which estimate the value of $u^{(-\alpha)}(x_\alpha)$ directly, as follows :

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

where the $h_{\alpha\alpha}$ is the $(\alpha,\alpha)^{\mathrm{th}}$component of the smoother matrix *H*. The matrix *H* transforms observed data *z* to predicted values $\hat{z}$ where *H* does not depend on the data z, and it is referred to as a hat matrix, a 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. Next, the generalized cross validation with influence function GCV_{IF} is calculated by

$$GCV_{IF} = \sum_{\alpha=1}^{n}\left\{\log\left(2\pi\hat{\sigma}^{2(-\alpha)}\right) + \left[\frac{z_\alpha - \hat{u}(x_\alpha)}{\hat{\sigma}^{(-\alpha)}(1-\frac{1}{n}\mathrm{tr}H)}\right]^2\right\}, \quad (2)$$

Where $h_{\alpha\alpha}$ is replaced with 1/n tr$H$, which is its average, and tr(*H*),which is called the effective number of parameters. The estimation $\hat{\sigma}^{2(-\alpha)}$ is approximated by the influence function $T^{(1)}(z_\alpha; \hat{G})$ as follows [4]

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

3. Method of Regularization

For the nonlinear statistical modeling, the maximum penalized likelihood methods are often used [5-7]. Suppose that we have *n*observations $\{(z_\alpha, x_\alpha); \alpha = 1, \cdots, n\}$, where z_α are the response variables generated from unknown true distribution *G*(*z*|*x*)having a probability density of *g*(*z*|*x*)and x_α are the vectors of explanatory variables.We estimate w, which is a vector consisting of the unknown parameters and determines the model $z = u(x|w)$.Let $f(z_\alpha|x_\alpha;\theta)$ be a specified parametric model, where θ is a vector of unknown parameters included in the model.The regression model with Gaussian noise is denoted as

$$z_\alpha = u(x_\alpha|w) + \varepsilon_\alpha\,, \varepsilon_\alpha \sim N(0,\sigma^2), \alpha = 1, \cdots, n$$

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

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

$$\ell_\lambda(\theta) = \sum_{\alpha=1}^{n}\log f(z_\alpha|x_\alpha;\theta) - \frac{n}{2}\lambda H(w) \quad (3)$$

As the regularized term or penalized terms$H(w)$ with an *m*-dimensional parameter vector w, various types are used depending on the dimension 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 the sum of the squares of w_i are used, and those are

$$H_1(w) = \frac{1}{n}\sum_{\alpha=1}^{n}\sum_{i=1}^{d}\left\{\frac{\partial^2 u(x_\alpha|w)}{\partial x_i^2}\right\}^2,$$

$$H_2(w) = \sum_{i=k+1}^{m}(\Delta^k w_i)^2,$$

$$H_3(w) = \sum_{i=1}^{m} w_i^2.$$

For the three dimensional approximation we use [8]

$$H(w) = \iint \left\{ \left(\frac{\partial^2 u}{\partial x^2}\right)^2 + \left(\frac{\partial^2 u}{\partial y^2}\right)^2 \right\} dxdy\,, \qquad (4)$$

and it is represented in the quadratic form

$$H(w) = w'Kw\,. \qquad (5)$$

Therefore the equation (3) will be

$$\ell_\lambda(\theta) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(z - Bw)'(z - Bw) - \frac{n}{2}\lambda w'Kw,$$

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

$$B = (b(x_1)', \cdots, b(x_n)')'\,.$$

With respect to θ,differentiating $\ell_\lambda(\theta)$ and setting the result equal to zero obtainstheir solution.As a result, the estimations of the parameters are

$$\hat{w} = (B'B + n\lambda\hat{\sigma}^2 K)^{-1}B'z\,,$$

$$\hat{\sigma}^2 = \frac{1}{n}(z - B\hat{w})'(z - B\hat{w})\,. \qquad (6)$$

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

3.1. B-splines

We consist the *B*-spline function $M_{m,i}(x)$of required degree *r*-1 (order *r*) by the algorithm of de Boor-Cox[8-11].This calculation can be begun by the first step:

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

and the successive recurrence formula is below:

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

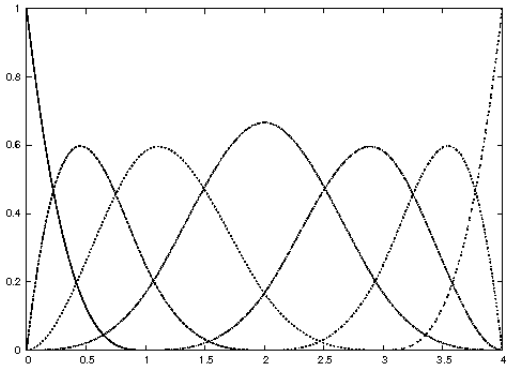


Figure 1.*Spline functions (order four)*

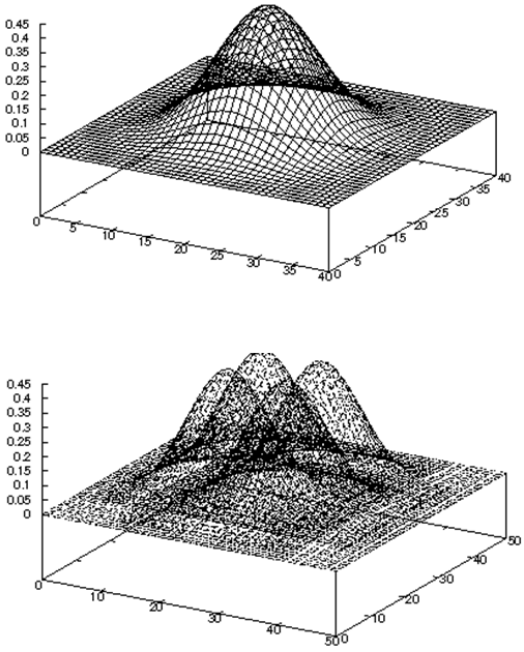


Figure 2.*Three dimensional spline functions (order four)*

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

We set the approximation for the three dimensional surface as

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

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 the *x*,*y* direction respectively.The shape of the three dimensional *B*-splines are shown in Fig. 2. The upper figure shows a one function with $\xi_i = (i-1) \times 10, \eta_i = (i-1) \times 10, i = 1, 2, \cdots 5$ .The lower figure shows four functions with p_1=p_2=2, $\xi_i = (i-1) \times 10, \eta_i = (i-1) \times 10$, *i*=1,2,…,6.The Schoenberg-Whitney condition [12] has to be satisfied, 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 the parameter w_{ij} cannot be determined.

In the equation of integration (4)

$$\left(\frac{\partial^2 u}{\partial x^2}\right)^2 = \left(\sum_{i=1}^{p_1}\sum_{j=1}^{p_2} w_{ij}\frac{d^2M_i(x)}{dx^2}N_j(y)\right)^2, \qquad (7)$$

$$\left(\frac{\partial^2 u}{\partial y^2}\right)^2 = \left(\sum_{i=1}^{p_1}\sum_{j=1}^{p_2} w_{ij}M_i(x)\frac{d^2N_j(y)}{dy^2}\right)^2. \qquad (8)$$

We set $w_{ij} = \tilde{w}_k, i = i_k, j = j_k, p_1p_2 = m$, $\frac{d^2M_i(x)}{dx^2}N_j(y) =$

$\tilde{B}_{1,k}, M_i(x)\frac{d^2N_j(y)}{dy^2} = \tilde{B}_{2,k}$.Then, the equation (7) (8) can be rewritten as

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

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

As a result, the integrations become

$$\iint \tilde{B}_{1,p}\,\tilde{B}_{1,q}\,dxdy =$$

$$\int \frac{d^2M_{i_p}(x)}{dx^2}\frac{d^2M_{i_q}(x)}{dx^2}dx \int N_{j_p}(y)N_{j_q}(y)\,dy, \quad (9)$$

$$\iint \tilde{B}_{2,p}\,\tilde{B}_{2,q}\,dxdy =$$

$$\int M_{i_p}(x)M_{i_q}(x)\,dx \int \frac{d^2N_{j_p}(y)}{dy^2}\frac{d^2N_{j_q}(y)}{dy^2}dy. \quad (10)$$

The sum of equations (9) and (10) will be K_{pq} which is the component of $m \times m$ nonnegative matrix K that is represented in the equation (5).

3.2Higher Order Empirical Influence Function

We can denote equations (6) as follows

$$\sum_{\alpha=1}^{n} \psi_i(x_\alpha;\theta) = 0 \quad (i = 1,2,\cdots,p), p = p_1p_2 + 1, \quad (11)$$

where $\theta = (w', \sigma^2)'$. When we denote $\psi = \left(\psi_1, \cdots, \psi_p\right)'$, the solution$\hat{\theta}$of the equation (11) is given by $\hat{\theta} = T(\hat{G})$ which is the vector of the functional with degree p defined with distribution G as follows:

$$\int \psi(x, T(G))dG(x) = 0.$$

By replacing the distribution $G$with $(1-\varepsilon)G + \varepsilon\delta_x$,we obtain

$$\int \psi(z, T((1-\varepsilon)G + \varepsilon\delta_x)\, d\{(1-\varepsilon)G(z) + \varepsilon\delta_x(z)\} = 0. \quad (12)$$

The first order influence function is given in [4].For the higher order influence function, we differentiate the equation (12) with respect to εtwice and let $\varepsilon = 0$, then we obtain

$$2\int \frac{\partial\psi(x,T(G))'}{\partial\theta} d\{\delta_x(z) - G(z)\} \cdot \frac{\partial}{\partial\varepsilon}\{T(H_\varepsilon)\}|_{\varepsilon=0} + \int \frac{\partial^2\psi}{\partial\theta\partial\theta}\frac{\partial T(H_\varepsilon)}{\partial\varepsilon}|_{\varepsilon=0} dG \cdot \frac{\partial T(H_\varepsilon)}{\partial\varepsilon}|_{\varepsilon=0} + \int \frac{\partial\psi}{\partial\theta} dG \cdot \frac{\partial^2 T(H_\varepsilon)}{\partial\varepsilon^2}|_{\varepsilon=0} = 0. \quad (13)$$

Therefore the influence function contains a second order as

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

Recall the equation

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

and the equation (13) can be rewritten as:

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

Thus, we obtained

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

4. Other Information Criteria

An information criterion for the model$f\left(z|x;\hat{\theta}\right)$, obtained by maximizing the penalized log-likelihood function (3)is given by

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

where$R(\psi,\hat{G})$ and $Q(\psi,\hat{G})$ are (m+1)×(m+1) matrices [4].

We adopted the next approximation for the alternative CV [4]

$$T\left(\hat{\mathrm{G}}^{(-\alpha)}\right) \approx T(\mathrm{G}) + \frac{1}{n-1}\sum_{i\neq\alpha}^{n} T^{(1)}(z_i;\mathrm{G}) \approx \mathrm{T}\left(\hat{\mathrm{G}}\right) - \frac{1}{n}T^{(1)}(z_\alpha;\hat{\mathrm{G}}).$$

In the equation (1) of IC_{CV} we replaced the $\hat{\theta}^{(-\alpha)}$ with $\hat{\theta} - \frac{1}{n}T^{(1)}(z_\alpha;\hat{\mathrm{G}})$ and its scheme was called the modified GIC (mGIC) [13],

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

5. Numerical Example

5.1. Surfaces and Samples

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

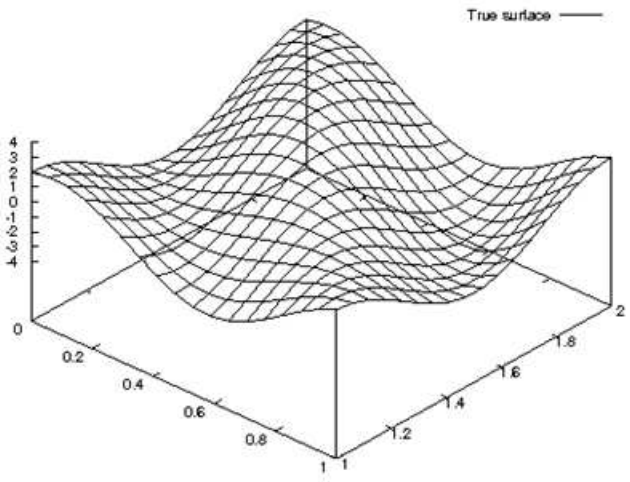
$$\mathrm{I}: z = \sin(2\pi x) + 2\cos\left(2\pi(x+y)\right)$$

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

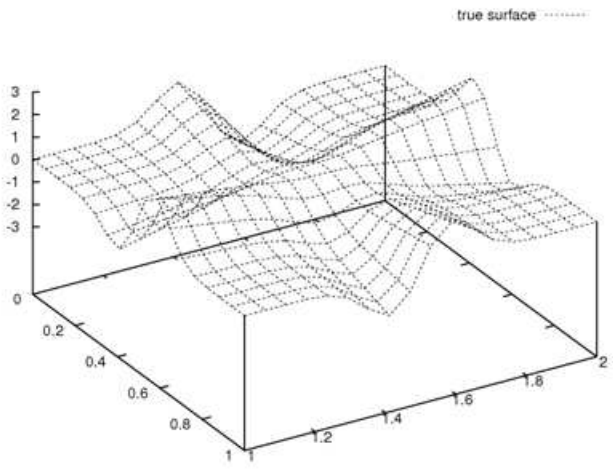
Those topographies are shown in Figure 3.For the estimation, we generated 300 sample coordinates data with the Gaussian noise according to the normal distribution $N(0,\sigma^2)$.

5.2. Estimation of Parameters

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



(a) Surface I



(b) Surface II

Figure 3.*The topographies of two surfaces*

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

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

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) where $n_1$and n_2 are the total number of knots along x and $y$directions respectively.Also, for every (n_1,n_2), we tested 100 sets of randomized knots generated uniformly.However, some of them didn't satisfy the Schoenberg-Whitney condition so another set of knots was generated again. Furthermore the equations of matrices made from ill-conditioned sets cannot be solved properly, so we also generated another sets of knots again, testing 100 solvable sets for every (n_1,n_2).The total number of the basis will be $(n_1\text{-}4)(n_2\text{-}4)$ and the total number of the parameters will be $(n_1\text{-}4)(n_2\text{-}4)+1$ which consists of the coefficients of the basis and the variance.For the regularization term, we used (4) and for the numerical calculation, we used (9) and(10). We tested the estimation with various β's which are from10^{-1} to 10^{-10} in principle.

5.3. Evaluation of Models

For the evaluation of the obtained parameters we test some criteria such as GICP, mGIC, CV and GCVIF.Those results are shown in Figures and Tables below. Fig. 4 summarize the results of GICP, mGIC, CV and GCVIF over the various values of β for surface I. And Fig. 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.Table 1 and 2 summarize the results of CV in the various values of β.The optimal value, which minimizes the information criterion Cross-Validation (1), determines 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 for every surface.

In the calculation of mGIC we use the influence function to estimate the value of parameters.This method can obtain almost the same result in parameters as CV and it takes very little time. It takes almost 35 minutes for every β and

it is 1/23 of the CV.In this approximation of parameters the

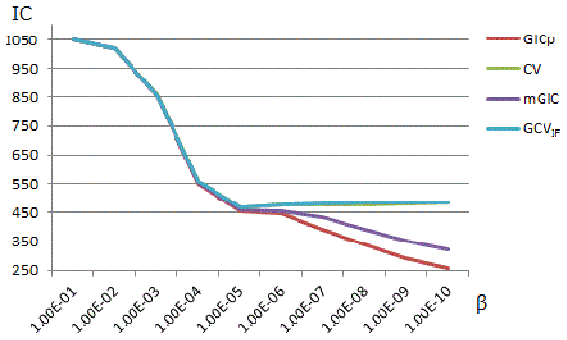


Figure 4. *Four criteria of Surface I*

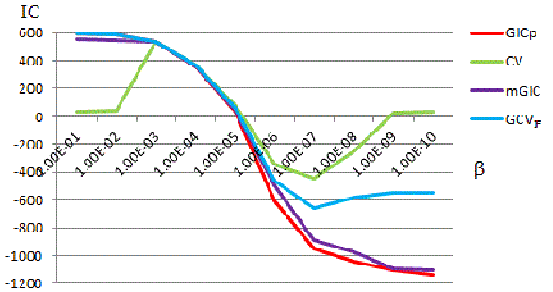


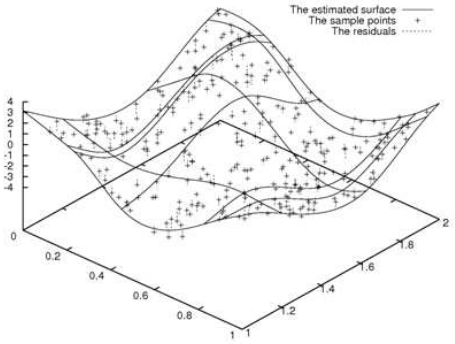
Figure 5. *Four criteria of Surface II*

difference between the mGIC and CV is quite small. The correlation coefficient is almost 1.0.For example, when β =10-5, n1= n2=20, the average of the correlation coefficients over 100 sets of knots is 0.99999611.However, the values of mGIC are monotone decreasing so we cannot determine the optimal parameters in this case.

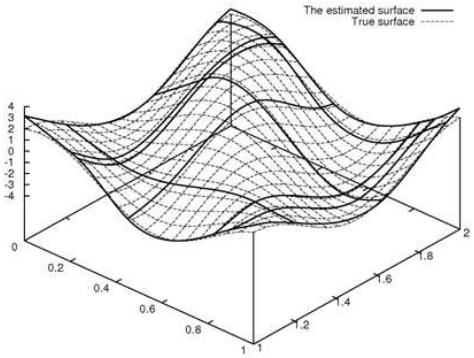
The results of an alternative method GCVIF with influence function (2) for model selection are shown in Table 3 and 4. For the estimation of variance we use the influence function.The selected models with optimal parameters determined by the GCVIF are shown in Fig. 6 and Fig. 7 for surfaces I and II respectively.In Fig. 6 the estimated surface is based on the (13,11) knots. In Fig. 7 the estimated surface is based on the (14,17) knots. In those figures the locations of knots and samples and the residuals are also shown.

5.4. Comparison between the Distributions of Criteria

We compare the distributions of four information criteria for the two surfaces I andII.The criteria are improved versions of GIC_P, CV, mGIC and GCV_{IF}.On the surface I, we show the Boxplots of four criteria overβ=10^{-4},10^{-5},10^{-6},10^{-7} inFig.8and 9. Similarly on the surface II, we show the Boxplots of four criteria overβ=10^{-4},10^{-5},10^{-6},10^{-7} inFig. 10 and 11. We can find from these figures that mGIC is larger than GIC_P and GCV_{IF} is closer to the CV than others.

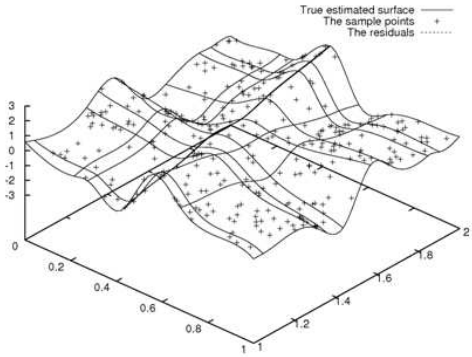


(a) Sample and estimated surface

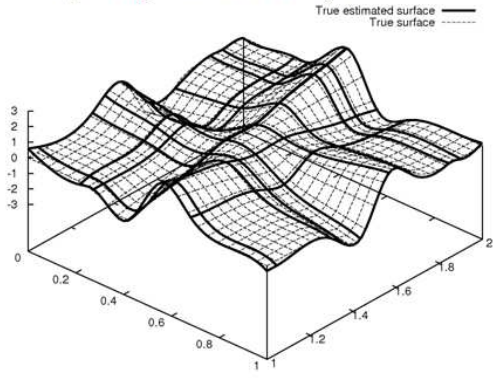


(b) Comparison of true surfaceand estimated surface

Figure 6. *SurfaceI*



(a) Sample and estimated surface



(b) Comparison of true surfaceand estimated surface

Figure 7. *SurfaceII*

Table1. *CV results for Surface I*

Total number of knots *x*-axis	*y*-axis	β	σ^2	λ	CV
19	13	1.000E-01	1.85E+00	5.4025E-02	1050.1
19	13	1.000E-02	1.67E+00	5.9798E-03	1021.1
20	17	1.000E-03	9.38E-01	1.0665E-03	859.3
20	17	1.000E-04	3.16E-01	3.1645E-04	560.3
13	**11**	**1.000E-05**	**3.16E-01**	**4.4643E-05**	**468.1**
10	11	1.000E-06	2.29E-01	4.3673E-06	475.1
10	11	1.000E-07	2.28E-01	4.3939E-07	478.0
10	11	1.000E-08	2.24E-01	4.4721E-08	478.8
10	11	1.000E-09	2.20E-01	4.5520E-09	479.8
10	10	1.000E-10	2.20E-01	4.5354E-10	486.1

Table2. *CV results for Surface II*

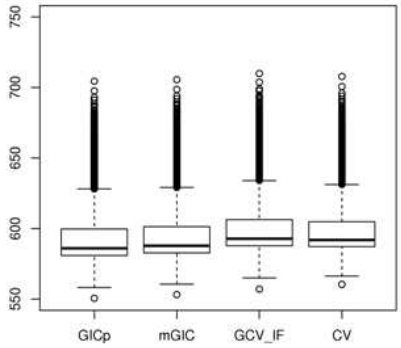
Total number of knots *x*-axis	*y*-axis	β	σ^2	λ	CV
13	13	1.000E-01	3.85E-03	2.60E+01	31.0
18	18	1.000E-02	5.35E-02	1.87E-01	40.2
20	15	1.000E-03	3.22E-01	3.10E-03	537.9
17	20	1.000E-04	1.66E-01	6.03E-04	367.4
20	15	1.000E-05	4.87E-02	2.05E-04	74.3
16	14	1.000E-06	6.37E-03	1.57E-04	-346.1
16	**14**	**1.000E-07**	**2.30E-03**	**4.34E-05**	**-441.1**
19	13	1.000E-08	2.77E-03	3.61E-06	-244.2
13	13	1.000E-09	3.85E-03	2.60E-07	25.9
13	13	1.000E-10	3.85E-03	2.60E-08	31.0

Table3. *GCV_{IF} results for Surface I*

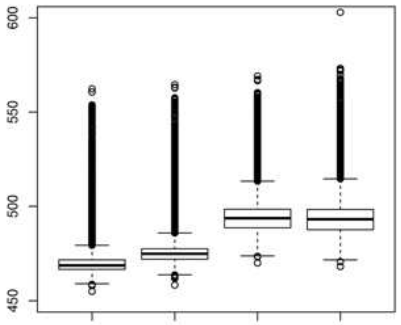
Total number of knots		β	σ^2	λ	GCV_{IF}
x-axis	*y*-axis				
19	13	1.000E-01	1.850993	5.40E-02	1050.1
19	13	1.000E-02	1.672307	5.98 E-03	1021.4
20	17	1.000E-03	0.937645	1.07 E-03	857.2
20	17	1.000E-04	0.316010	3.16 E-04	556.9
13	**11**	**1.000E-05**	**0.224000**	**4.46 E-05**	**470.0**
12	10	1.000E-06	0.214460	4.66 E-06	478.4
10	11	1.000E-07	0.227586	4.39 E-07	481.9
10	11	1.000E-08	0.223611	4.47 E-08	484.3
10	10	1.000E-09	0.220486	4.54 E-09	487.3
10	10	1.000E-10	0.220486	4.54 E-10	487.3

Table4. *GCV_{IF} results for Surface II*

Total number of knots		β	σ^2	λ	GCV_{IF}
x-axis	*y*-axis				
18	11	1.000E-01	0.407434	2.45 E-01	596.9
20	19	1.000E-02	0.397532	2.52 E-02	591.1
20	15	1.000E-03	0.355688	2.81 E-03	536.8
17	20	1.000E-04	0.165862	6.03 E-04	363.9
10	10	1.000E-05	0.227627	4.39 E-05	60.3
14	17	1.000E-06	0.005844	1.71 E-04	-463.0
14	**17**	**1.000E-07**	**0.001890**	**5.29 E-05**	**-659.3**
14	17	1.000E-08	0.001712	5.84 E-06	-582.7
13	13	1.000E-09	0.003851	2.60 E-07	-549.8
13	13	1.000E-10	0.003851	2.60 E-08	-549.5

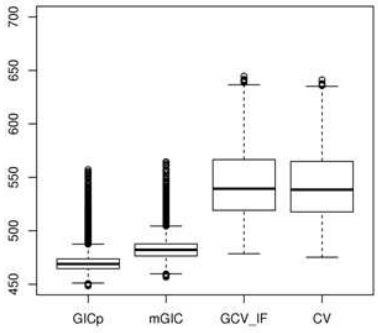


(a) $\beta=10^{-4}$

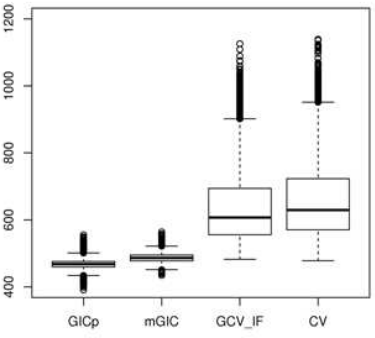


(b) $\beta=10^{-5}$

Figure 8.*Boxplotsfor four criteria of surface I*

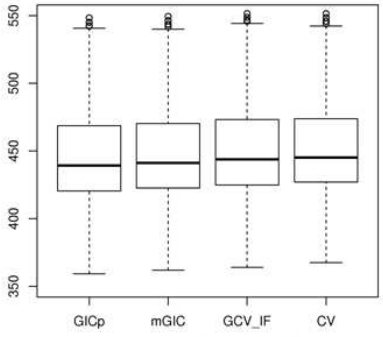


(a) $\beta=10^{-6}$

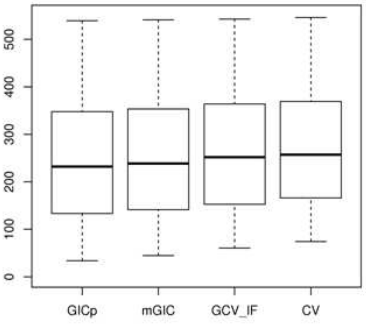


(b) $\beta=10^{-7}$

Figure 9. *Boxplotsfor four criteria of surface I*

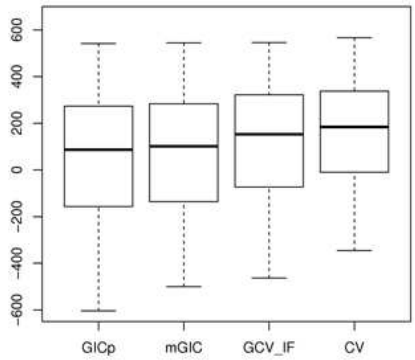


(a) $\beta=10^{-4}$

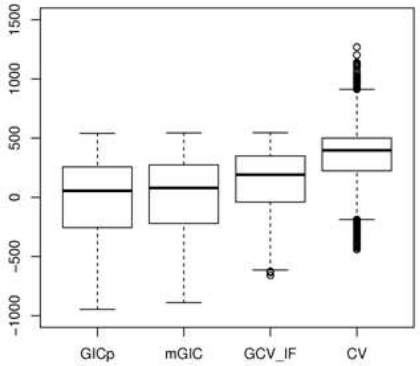


(b) $\beta=10^{-5}$

Figure 10. *Boxplotsfor four criteria of surface II*



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



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

Figure 11. *Boxplots for four criteria of surface II*

7. Conclusion

As the value of β decreases, the residual variance reduces and the information criterion GIC_p also reduces monotonously. As a result, the GIC_p cannot determine the optimum model in both surfaces.The method for 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 computational costs.

For the parameter estimation the alternative method mGIC by the information function works well.However, the total number of parameters is so many (more than 36 and less than or equal to 257) 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 the CV very well and it requires small computation.

We propose GCV_{IF}as an improved GCV criterion. This conclusion is a theory 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 the CV, GIC and mGIC, and we can also see that the GCV_{IF} criterion includes their informations.

References

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

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

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

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

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

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

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

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

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

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

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

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

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