Kriging Interpolation Methods in Geostatistics and DACE Model

Je-Seon Ryu
Ph. D. Student, Department of Mathematics, Hanyang University, Seoul 133-791, Korea
Min-Soo Kim
Research professor, Center of Innovative Design Optimization Technology, Hanyang University, Seoul 133-791, Korea
Kyung-Joon Cha
Department of Mathematics, Hanyang University, Seoul 133-791, Korea
Tae Hee Lee
School of Mechanical Engineering, Hanyang University, Seoul 133-791, Korea
Dong-Hoon Choi*
Director, Center of Innovative Design Optimization Technology, Hanyang University, Seoul 133-791, Korea

In recent study on design of experiments, the complicate metamodeling has been studied because defining exact model using computer simulation is expensive and time consuming. Thus, some designers often use approximate models, which express the relation between some inputs and outputs. In this paper, we review and compare the complicate metamodels, which are expressed by the interaction of various data through trying many physical experiments and running a computer simulation. The prediction model in this paper employs interpolation schemes known as ordinary kriging developed in the fields of spatial statistics and kriging in Design and Analysis of Computer Experiments (DACE) model. We will focus on describing the definitions, the prediction functions and the algorithms of two kriging methods, and assess the error measures of those by using some validation methods.

Key Words : Kriging, Ordinary Kriging, DACE Model, Semivariogram, Correlation Function, BLUP

1. Introduction

Kriging is based on the field of geostatistics, for example, hybrid discipline of mining engineering, geology, mathematics and statistics. The approach to prediction advocated in this paper has been known as kriging after Dr. D. G. Krige's work (1951) on the Rand gold deposit, in southern Africa. He developed an empirical method for determining a true ore grade distribution from distributions based on sampled ore grade in the 1950's. Matheron (1963) developed this kriging technique in France. This performed well in predicting the value of a possible but actually not taken observation of a spatially distributed variable such as a mine grade (Krige, 1951), a soil characteristic (Webster, 1985), rain fall (Bacchi and Kottegota, 1995), gene frequency (Piazza, et al., 1983), or image sequence coding (Deceneiere, et al., 1998). There are other texts and papers that describe kriging, the spatially correlated data and the mining related with kriging (Volpi and Gambolati, 1978; Gambolati and Galcati, 1985; Kitanidis, 1983; Barendrengt, 1987; Cressie, 1991). Kriging was also compared with splines in

* Corresponding Author,
E-mail : dhchoi@hanyang.ac.kr
TEL : +82-2-2290-0478; **FAX** : +82-24-2290-1639
Director, Center of Innovative Design Optimization Technology, Hanyang University, 17 Haengdang-dong, Sungdong-gu, Seoul 133-791, Korea. (Manuscript **Received** July 14, 2001; **Revised** February 4, 2002)

Dubrule (1983) and Sasena (1998). Recently, kriging goes by a variety of names including DACE model, which is the title of the inaugural paper by Sacks, et al. (1989).

The kriging method is derived from geostatistics and used for fitting the model of the deterministic output as the realization of a random process for efficient predicting. Since computer experiments typically lack random error, an appropriate method for approximating deterministic computer experiments utilizes kriging, which is also referred to as Design and Analysis Computer Experiments (DACE) model. There have been some studies for DACE modeling methods at AIAA (American Institute of Aeronautics and Astronautics), in particular. Sacks, et al. introduced firstly kriging as a tool of interpolation of deterministic computer experiments in 1989. Giunta (1997) and Giunta, et al. (1998) performed a preliminary investigation into the use of kriging for the multidisciplinary design optimization of a high speed civil transport aircraft. Sasena (1998) and Mardia, et al. (1996) compared and contrasted kriging with splines (nonparametric regression model). Booker (1996) and Booker, et al. (1996) used a kriging approach to construct the approximate model for the aeroelastic and dynamic response of a helicopter rotor during structural design. Simpson (1998) provided several kriging algorithms and a source code in FORTRAN.

A kriging includes the mean of "optimal predictor" or "optimally predicting," and this is the very kriging interpolation method. The predictor from an untried point in a region of interest will be a linear combination of the observations with positive or negative weights plus departure. These methods provide not only a best linear unbiased predictor, but also an estimate of the variance of the predictor error. After predicting, one can produce a map of this variable, that is, contour plot, isopleth shaded area, or graphical techniques for surfaces. Kriging method is the only method with a sound statistical base, which is to be applied when an uncertainty exists and the variation is a function of the distance between measurements.

A kriging requires understanding of the principles of spatial autocorrelation and is used when the variation of data is so irregular that simple methods of interpolation may give unreliable predictions. In general, the spatial structure of stochastic dependence between outputs at different tried inputs is represented by covariance, semivariogram or the standardized measure of the correlation coefficient. Semivariogram is the half of variance of the difference between input data and is usually used as a measure of association in geostatistics. This is dependent only on the distance between the inputs, which is so-called intrinsically stationary, and is efficiently used for isotropic and ergodic data. Correlation function in DACE model is defined that the component of its matrix is the inverse of exponential function of the distance between inputs product unknown coefficient of correlation. This provides the best predictor by using maximum likelihood estimator (MLE) of the coefficient of correlation. One of several theoretical semivariogram and correlation function in respective kriging methods is built with many ways which is chosen for building the prediction model.

We will describe the definition, the prediction function, and the algorithm of ordinary kriging. Kriging in DACE model is described with the same those of ordinary kriging. This model uses several correlation functions, which are special forms of the distance matrices. The difference between ordinary kriging and DACE model is described with data and the measure of association. The method for assessing kriging model accuracies is presented, and accuracy of two kriging methods is compared using the validation method with data from the computer simulation.

2. Ordinary kriging

There have been studies for several types of kriging used in geostatistics such as ordinary kriging (Matheron, 1962; Gandin, 1963), universal kriging (Goldberger, 1963; Matheron, 1963), median polish (Cressie, 1986), etc. Ordinary kriging is one of them and the basis of geostatistics. Ordinary kriging gives the optimal

predictions under the assumption that the process is second-order stationary and is distributed with normal, and that observed values are realization of a stationary stochastic process of fairly simple structure. This kriging model is the weighted linear combination of the observation with white noise process. Thus the optimal predictor will be accepted by minimizing mean squared prediction error. The observed values are first used to estimate the unknown parameters of the process and to compute empirical semivariogram. These observed values, parameters and semivariogram are used to produce the best linear unbiased predictor of the unobserved point.

2.1 Model construction

The random process $Z=(Z(x_1), \cdots, Z(x_n))'$, which is n observed values at the known spatial location $\{x_1, \cdots, x_n\} \in D \subset R^d$, presents the realization of points in $D \subset R^d$. This is modeled as the partial realization of the random process

$$\{Z(x);\ x \in D \subset R^d\} \tag{1}$$

It refers to making inferences on unobserved values of the random process $Z(\cdot)$ given by equation (1) from output Z. That is, a spatial predicting is to predict unknown function $g(\{Z(x);\ x \in D\})$ from n observed data in the visited location $x_1, \cdots, x_n$.

Ordinary kriging refers to a spatial prediction under the following two assumptions. The model is assumed as $Z(x)=\mu+\delta(x)$ for $x \in D$ and $\mu \in R$ and $\delta(x)$ has zero mean and is intrinsically stationary. Also, its predictor is assumed as $p(z;x)=\sum_{i=1}^{n} \lambda_i z(x_i)$, $\sum_{i=1}^{n} \lambda_i=1$, which guarantees an uniform unbiasedness. Denoting the generic predictor of $g(Z(\cdot))$ by $p(z;g)$ for an unvisited sample point x, the optimal $p(\cdot;x)$ will minimize the mean-squared predictor error $\sigma^2=E(z(x)-p(z;x))^2$ over the class of linear predictors $\sum_{i=1}^{n} \lambda_i z(x_i)$ that satisfy $\sum_{i=1}^{n} \lambda_i=1$.

If two points are close together, their dependency will typically be large. Otherwise, their dependency will get smaller. In a given area, semivariogram, between the values assumed by the process $Z(x)$ at two observed points x_1 and x_2, can be described by matrix form $\Gamma(x_1, x_2)=1/2\ \mathrm{Var}[z(x_1)-z(x_2)]=1/2E[\{z(x_1)-z(x_2)\}-\{m(x_1)-m(x_2)\}]^2$. In this matrix form, the m denotes the mean of the process. For more detailed comparison between covariance and semivariogram, one may refer to Cressie (1991).

For a given area, within which one needs to understand the structure of the random process, it is obviously impossible to know the true form of association. Thus, the estimator of semivariogram should be obtained using appropriate methods. In the definition of semivariogram under the constant-mean assumption, a natural estimator based on the methods of moments, is $2\hat{\Gamma}(h) \equiv \frac{1}{|N(h)|} \sum_{N(h)} (z(x_i)-z(x_j))^2$ for $h \in R$, where $N(h) \equiv \{(x_i, x_j);\ x_i-x_j=h,\ i, j=1, \cdots, n\}$ and $|N(h)|$

Table 1 Summary of theoretical semivariogram model

	theoretical semivariogram model
Linear model	$\begin{cases} A_0\delta(h)+A_1h & \text{for } h<a \\ A_0+A_1a & \text{for } h \geq a \end{cases}$
Spherical model	$\begin{cases} A_0\delta(h)+w/2[3h/a-(h/a)^3] & \text{for } h<a \\ A_0+w & \text{for } h \geq a \end{cases}$
Exponential model	$A_0\delta(h)+w[1-\exp(-h/a)]$
Gaussian model	$A_0\delta(h)+w[1-\exp(-h/a)^2]$

$\delta(h)=1$ for $h>0$; $\delta(h)=0$ for $h=0$

A_0 is nugget effect caused by possible errors of measurement,

A_1 is the rate of decrease of the spatial covariance in the field for the linear model,

A_0+w is sill, which is the variance of the field less the discontinuity A_0,

a is range, or the correlation distance, and is in practce the maximum distance for which observations are correlated.

is the number of the distance pairs in $N(h)$ (Matheron, 1962). From the analysis of the empirical semivariogram, it is possible to formulate the theoretical models suitable for representing the visited spatial variability by determining the lag h. Some of the theoretical semivariogram models of the homogeneous and isotropic semivariogram, used in practice, are shown in Table 1.

After estimating the semivariogram model, structural analysis and interpolation for the unobserved points in the interested region can be provided because the semivariogram determines the optimal weights for the interpolation.

2.2 Predicting the spatial distribution

If n observed values of the stochastic process $Z(x_i)$ are known at the given points x_1, $i=1, \cdots, n$, the predicted value $\hat{Z}(x)$ needs to be obtained at the unobserved point x. For the kriging coefficient λ_i, $i=1, \cdots, n$, linear combination $\hat{Z}(x)=\sum_{i=1}^{n}\lambda_i Z(x_i)$ is the optimal estimate when two conditions of unbiasedness and minimum variance are satisfied. The estimator with these conditions is called BLUP (Best Unbiased Linear Predictor). The unbiasedness condition gives $E[\hat{Z}(x)]=E\left[\sum_{i=1}^{n}\lambda_i Z(x_i)\right]=\sum_{i=1}^{n}\lambda_i E[Z(x_i)]=\sum_{i=1}^{n}\lambda_i m=m$, where m is constant mean. We can then minimize

$$E(Z(x)-\sum_{j=1}^{n}\lambda_j Z(x_j))^2-2\mu\left(\sum_{j=1}^{n}\lambda_j-1\right) \quad (2)$$

with respect to λ_i, $i=1, \cdots, n$, and μ, where μ is a Lagrange multiplier that ensures $\sum_{j=1}^{n}\lambda_j=1$. With this condition, Eq. (2) becomes

$$-\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j\gamma(x_i-x_j)+2\sum_{i=1}^{n}\lambda_i\gamma(x-x_i)-2\mu\left(\sum_{i=1}^{n}\lambda_i-1\right) \quad (3)$$

where $\gamma(\cdot)$ is a semivariogram of the values between the processes in the two observed points. After differentiating Eq. (3) with respect to λ_i, $i=1, \cdots, n$, and μ, and equating their result to zero, the optimal parameters are obtained to satisfy

$$\sum_{j=1}^{n}\lambda_j\gamma(x_i-x_j)+\mu=\gamma(x_i-x), \quad i=1, \cdots, n$$

$$\sum_{i=1}^{n}\lambda_i=1$$

The kriging coefficients are determined through equating matrix multiplication and summation

$$\begin{aligned}\lambda&=\Gamma^{-1}\gamma+\frac{\Gamma^{-1}1_n 1_n'\Gamma^{-1}\gamma-\Gamma^{-1}1_n}{-1_n'\Gamma^{-1}1_n}\\&=\Gamma^{-1}\left(\gamma+1_n\frac{1-1_n'\Gamma^{-1}\gamma}{1_n'\Gamma^{-1}1_n}\right)\end{aligned} \quad (4)$$

where $\gamma=(\gamma_1, \cdots, \gamma_n)$, and γ_i is the semivariogram between the observed points and the unobserved points, the value of 1_n is the vector whose elements are all ones, and Γ is semivariogram matrix of the values $z(x)$ at the observed points.

This kriging coefficient vector computed from Eq. (4) minimizes the mean squared estimator error, or $E(z(x)-\hat{z}(x))^2$. Minimum mean squared estimator error is thus sometimes called kriging variance $\hat{\sigma}^2$, as

$$\hat{\sigma}^2(x)=\gamma'\Gamma^{-1}\gamma-\frac{(1_n'\Gamma^{-1}\gamma-1)^2}{1_n'\Gamma^{-1}1_n}$$

The spatial variable $z(s_0)$ is distributed with normal, and the prediction interval 100 $(1-\alpha)$% can be constructed.

3. DACE Model

In many computer experiments, the observation is formulated for the response function implementing a computer model with choosing various input factors. For the given specific input x, one can accept the response value for Y by solving the differential equation computationally. Since the equation solver is highly expensive and time consuming, the response value y needs to be estimated at the untried input points with solving the relationship between some tried points and response values through the reasonable number of computation. In practical situation, many computer codes are deterministic and are not subject to measurement error. Deterministic means that the responses from computer code with same inputs will be identical, and are different from the random error derived from the case of physical experiments.

Table 2 Summary of Correlation Functions

	Correlation Function
Exponential	$\prod_{k=1}^{d} \exp(-\theta_k \mid d_k \mid)$
Gaussian	$\prod_{k=1}^{d} \exp(-\theta_k \mid d_k \mid^2)$
Cubic spline	$\prod_{k=1}^{d} \begin{cases} 1-6(\theta_k \mid d_k \mid)^2+6(\theta_k \mid d_k \mid)^3 & \theta_k \mid d_k \mid < 1/2 \\ 2(1-(\theta_k \mid d_k \mid)^3) & 1/2 \le \theta_k \mid d_k \mid < 1 \\ 0 & \theta_k \mid d_k \mid \ge 1 \end{cases}$
Matern linear function	$\prod_{k=1}^{d} [(1+\theta_k \mid d_k \mid)\exp(-\theta_k \mid d_k \mid)]$
Matern cubic function	$\prod_{k=1}^{d} [(1+\theta_k \mid d_k \mid+\theta_k^2 \mid d_k^2 \mid/3)\exp(-\theta_k \mid d_k \mid)]$

d is the number of design variable,
θ_k is the unknown vector of correlation parameters used to be fitted this kriging model,
d_k is the distance between any two sample points x_k^i and x_k^j of the k^{th} components.

3.1 Model construction

Kriging in DACE model is formulated as the Eq. (5) that is a combination of a linear regression model and departure.

$$y(x)=\beta f(x)+Z(x) \tag{5}$$

where $y(x)$ is the unknown function of interest, $f(x)$ is a known linear function of input x and $Z(x)$ is the realization of stochastic process with zero mean, σ^2 variance, and nonzero covariance. Also, β is the unknown vector that should be estimated from n observed response values from the properties of kriging interpolation in the DACE model. In Eq. (5), $\beta f(x)$ provides a global model in design space, which is similar to the polynomial type response surface model. It is noted that $Z(x)$ creates a localized deviation between the global model and the exact model. Therefore, kriging model can successfully interpolate the n sampled data points. The covariance matrix of $Z(x)$ is expressed as

$$Cov[Z(x^i), Z(x^j)]=\sigma^2 R[R(x^i, x^j)],$$
$$i, j=1, \cdots, n$$

where R is a correlation matrix and $R(x^i, x^j)$ is a correlation function between any two points x^i and x^j in the n sampled points. R is an $(n\times n)$ symmetric and positive definite matrix of which diagonal elements are one values. The correlation function $R(x^i, x^j)$ can be specified to reflect the association effectively among the sampled points. For a smooth response value, a correlation function with some derivatives would be preferred, whereas an irregular response might call for a correlation function with no derivatives.

Several correlation functions are listed in Table 2. Sacks et al (1989) proposed the exponential and the Gaussian functions. Mitchell and Morris (1992) used the three correlation functions such as a cubic spline, a matern linear and a matern cubic correlation. In this study, a unique θ value for each dimension is considered based on past difficulties with scaling the design space to $[0, 1]^d$ during the model fitting process. Although this study uses a unique θ value for each dimension, it is worth noting that using a single correlation parameter gives sufficiently good results in some cases (Booker, et al., 1995; Osio and Amon, 1996). The exponential correlation function, used in this study, is rewritten as

$$R(x^i, x^j)=\exp\left[-\theta\sum_{k=1}^{n}\left|d_k\right|\right] \tag{6}$$

Another term of interest is the correlation vector, $r(x)$, between the response values at the observed points $x^i, \cdots, x^n$ and the response at a given location x that requires an estimation. This correlation vector can be expressed as

$$r(x)=[R(x, x^1), \cdots, R(x, x^n)] \tag{7}$$

3.2 Kriging prediction

When the symbol $\hat{y}(x)$ is the estimated DACE model for the unknown function $y(x)$ in Eq. (5), this estimate can be defined from the statistical notation as $\hat{y}(x)=E[y(x)\mid y(x^1), \cdots, y(x^n)]$, where the terms $y(x^1), \cdots, y(x^n)$ are the observed response values of the n sampled points. Also, $y(x)$ is the true function of untried input x and $\hat{y}(x)$ is the actual estimate of x to be expected to close to real $y(x)$.

Considering the linear predictor $\hat{y}(x)=c'(x)\,y_x$ of $y(x)$ at an untried x, we can conceptually replace $y(x)$ by the corresponding random quantity $Y_x=\{Y(x^1), \cdots, Y(x^n)\}$ in order to treat $\hat{y}(x)$ as random, and compute the mean squared error of this predictor averaged over the random process. For the random quantities, the best linear unbiased predictor (BLUP) is obtained by choosing the $n\times 1$ vector $c(x)$ to minimize mean squared error (MSE), $MSE[\hat{y}(x)]=E[c'(x)\,Y_x-Y(x)]^2$ while satisfying the unbiasedness constraint $E[c'(x)\,Y_x]=E[Y(x)]$. In order to give some technical details connected with implementing the BLUP of the response at an untried input, the notation $f(x)=[f_1(x), \cdots, f_d(x)]'$ can be introduced for the d functions in the regression such as

$$F=\begin{pmatrix} f'(x^1) \\ \vdots \\ f'(x^n) \end{pmatrix}$$

where F is the $n\times d$ expanded design matrix. Also, $R=\{R(x^i, x^j)\mid 1\le i, j\le n\}$ becomes the $n\times n$ matrix of stochastic-process correlations, and $r(x)$ is the vector of correlations with this definition. Then, MSE of $\hat{y}(x)$ becomes

$$E[c'(x)\,Y_x-Y(x)]^2 = \sigma^2[c'(x), -1]\begin{bmatrix} R & r(x) \\ r'(x) & 1 \end{bmatrix}\begin{bmatrix} c(x) \\ -1 \end{bmatrix} \quad (8)$$

Typically, an unbiased requirement leads to the constraints $F'c(x)=f(x)$. The BLUP of $Y(x)$ is obtained by minimizing $MSE[\hat{y}(x)]$ subject to $F'c(x)=f(x)$. Using Lagrange multiplier for the constrained minimization of the MSE produces $Rc(x)=F\lambda(x)+r(x)$ and the coefficient $c(x)$ of the BLUP must satisfy

$$\begin{pmatrix} 0 & F' \\ F & R \end{pmatrix}\begin{pmatrix} -\lambda(x) \\ c(x) \end{pmatrix}=\begin{pmatrix} f(x) \\ r(x) \end{pmatrix} \quad (9)$$

The BLUP can be obtained by inverting the partitioned matrix as

$$\hat{y}(x)=f'(x)\,\hat{\beta}+r'(x)\,R^{-1}(Y_x-F\hat{\beta}) \quad (10)$$

where $\hat{\beta}=(F'R^{-1}F)^{-1}F'R^{-1}Y_x$ is the usual generalized least squared estimate of β. In Eq. (10), two terms on the right side are uncorrelated, and the second term can be interpreted as a smoothing surface of the residuals. Therefore, the DACE can be interpreted as two stages; fit and then interpolate the residuals as if there were no regression model.

A convenient representation for the MSE $[\hat{y}(x)]$ is obtained by substituting Eq. (9) in Eq. (8) togive

$$MSE[\hat{y}(x)] = \sigma^2\left(1-[f'(x), r'(x)]\begin{bmatrix} 0 & F' \\ F & R \end{bmatrix}^{-1}\begin{bmatrix} f(x) \\ r(x) \end{bmatrix}\right)$$

Assuming the Gaussian process, the likelihood is a function of the β's in the regression model, the process variance σ^2 and the correlation parameters. Given the correlation parameters, the estimate of the β's is the generalized least squared estimate, and MLE of σ^2 is

$$\hat{\sigma}^2=\frac{1}{n}(y_x-F\hat{\beta})'R^{-1}(y_x-F\hat{\beta}) \quad (11)$$

With the definitions of $\hat{\beta}$ and $\hat{\sigma}^2$, the problem can be simplified to maximize

$$\phi(\theta)=-(\det R)^{\frac{1}{n}}\hat{\sigma}^2 \quad (12)$$

This is the function of only the correlation parameter and the data. Any value for the unknown parameter θ in Eq. (6) creates an interpolative approximation model. However, the best kriging model can be obtained by solving the unconstrained nonlinear optimization problem of the Eq. (12) with respect to the value of θ.

4. Spatial Design and DACE Model

In the ordinary kriging, the random process is often modeled by using the semivariogram $\Gamma(x_1, x_2)=\frac{1}{2}Var[Z(x_1)-Z(x_2)]$ rather than the covariance function. Thus, the ordinary kriging requires the theoretical semivariogram for prediction of the spatial process at the unobserved locations. Hence, it is necessary to decide on a theoretical semivariogram based on the sample semivariogram. While there are several methods of fitting semivariogram models such as the least squares, the maximum likelihood and the robust methods (Cressie, 1991), it seems that these methods are not appropriate for data sets

resulting in a small number of semivariogram points. Instead, a visual fit of the semivariogram point to a few standard models is widely used. Even when there are sufficient semivariogram points, a visual check against a fitted theoretical model is more appropriate. After a suitable theoretical model is determined with the number of lags, the experimental semivariogram is computed. A suitable theoretical model is thus found visually and one can perform ordinary kriging using semivariogram. Although this visual fit gives more accurate model, it prevents the ordinary kriging from being automated because an experienced human interaction is required during its statistical process.

For DACE modeling, the correlation parameters θ_j in Table 2 are estimated by MLE. The MLE, based on the underlying Gaussian process model, is a form of cross-validation in which subsets of data are used to predict the remaining data. The correlation parameters can be numerically chosen to minimize a summary measure of errors from these predictions. Ripley (1988) discussed problems with optimizing likelihood to estimate correlation parameters. Other approaches besides MLE are discussed in Besac (1997), Currin, et al. (1988), and Torczon (1998).

The difference between the ordinary kriging and the DACE model is as follows; first, ordinary kriging is used for random data with the measurement error, while data in DACE model is deterministic. Second, the measures of the correlation of both models are the semivariogram in ordinary kriging and the correlation function in DACE model, respectively. Nugget, sill and range are used in determining semivariogram and the value of θ is directly used as the coefficient of correlation function in DACE model.

As the ordinary kriging has usually treated only two and three-dimensional space in geostatistical situation, it is not obvious that the methods of estimating semivariogram well extend from the low dimensional spatial coordinate to the high dimensional inputs of computer experiments. These data, sometimes appear to have measurement error or may be more erratic than response from computer codes. Hence, the geostatistical models used often in corporate a so-called nugget effect, or micro-scaled variations, which is caused by a discontinuity at the origin of the plot of semivariogram vs. lag. The only possible reason for nugget effect is measurement error. This occurs when a measurement is taken several times (duplicate) and different results are obtained.

The DACE model describes that correlation function of the form $R(x^i, x^j) = \prod \exp(-\theta |x^i - x^j|^p)$ with $0 < p \leq 1$ may be useful for modeling such the computational experiments that have high dimension inputs and deterministic results. The DACE model can also be extended to settings where systematic and random errors are both important and should be useful for physical experiments.

5. Validating a Kriging Model

Since kriging model interpolates the data, residual plots and R^2 statistic, which are the usual model assessments for response surface, are meaningless for DACE model (Myers and Mongomery, 1995; Simpson, 1998). Therefore, the additional validation points are needed to validate kriging model. If additional validated points can be afforded, the maximum absolute error, average absolute error and root mean squared error for these points can be calculated to assess model accuracy. These measurements are shown in Table 3 where the above three measures for model assessment methods are named by abbreviated words of Maxerr, Aveerr and Rtmse, respectively. The values of n, y_i and $\hat{y}_i$ denote the number of the untried points used to test, the real values from input and the predicted value from the kriging model, respectively.

Table 3 Error Measures for assessing kriging model accuracy

	Error Measure
Maxerr	Max. $\lvert y_i - \hat{y}_i \rvert \quad i = 1, \cdots, n$
Avgerr	$\frac{1}{n}\sum_{i=1}^{n} \lvert y_i - \hat{y}_i \rvert$
Rtmse	$\sqrt{\frac{\sum_{i=1}^{n} \lvert y_i - \hat{y}_i \rvert^2}{n}}$

6. Numerical Studies

In order to compare the accuracy of both kriging models, two test functions are estimated such as bird function and Rosenbrock's function using computer simulation. For these test functions, the number of design variable $d=2$ and inputs are sampled at equally spaced locations by $n=7\times7=49$, $n=10\times10=100$, $n=14\times14=196$ and $n=20\times20=400$ sample points on the given area. To assess these kriging model accuracies, each test function is sampled at $n=30\times30=900$ equally spaced points on the given area. In the problem of design of experiment for building kriging model, "space filling" experimental design with no blocking, replication and randomization may be better suited for computer experiments. In this example, we use equally spaced sampling method, but several other unequal spacing experimental designs exist (Myers and Mongomery, 1995). Simpson (1998) summarized a wide variety of design including the Latin hypercubes, the minimax/maximin designs and orthogonal arrays, etc. The predicted models with 49, 100, 196, 400 sampled points and exact model (constructed by 900 sample points) are plotted for each test function, respectively.

In this study, the algorithms for ordinary kriging and DACE model are coded with S-plus 2000. In order to fit kriging in DACE model, the initial design sample are scaled to $[0, 1]^2$. With the given inputs and responses, three semivariogram parameters are accepted for ordinary kriging predictor and the coefficient of correlation is also accepted for DACE model predictor of each test function. In ordinary kriging, sill, range and nugget are firstly chosen from the plot of semivariogram vs. lag for formulating the theoretical model as described previously. We then fit a theoretical semivariogram model to an empirical one. In the case of ordinary kriging, the theoretical exponential model with 20 lags is employed in these two test functions. The number of lags can be subjectively determined by the user at the design space and we use 20 lags that is the default in S-plus SPATIALSTATS(1996). Finally, we determine the exponential model as an appropriate semivariogram model from the plot of semivariogram vs. lag.

Table 4 Bird function; Error measures of ordinary kriging and DACE model

		Number of Sampling Points			
	Error	49	100	196	400
Ordinary kriging	Maxer	3.657747	2.358567	2.542853	2.130335
	Aveerr	0.406870	0.321746	0.287571	0.265200
	Rtmse	0.687343	0.549449	0.495159	0.461488
DACE model	Maser	3.266171	2.519098	2.477272	1.423881
	Aveerr	0.411055	0.280579	0.165186	0.081779
	Rtmse	0.682882	0.504521	0.336560	0.197590

Maxerr=maximum absolute error; Aveerr=average absolute error; Rtmese=root mean squared error

Table 5 Rosenbrock's valley function; Error measures of ordinary kriging and DACE model

		Number of Sampling Points			
	Error	49	100	196	400
Ordinary kriging	Maxerr	1310.936	1849.761	1521.834	1369.704
	Aveerr	221.6596	199.9768	153.6563	152.1085
	Rtmse	338.1948	336.4224	275.2284	273.6638
DACE model	Maxerr	573.0329	295.7348	192.5026	106.6619
	Aveerr	123.9449	48.48755	20.78803	8.652694
	Rtmse	193.5972	85.09829	40.45158	18.71458

Maxerr=maximum absolute error; Aveerr=average absolute error; Rtmese=root mean squared error

After predicting empirical semivariogram, kriging predictions at a set of unobserved spatial locations are computed using weights and plotted on three-dimensional space. In DACE model, we use the exponential model in Table 2. It is most suitable for estimating the smooth surface (Sacks et al., 1989) unlike that of semivariogram. The scalar coefficient of correlation θ in Eq. (6) is estimated from a one-dimensional analytic function maximizing $\phi(\theta)$ in Eq. (12). In fact, the accuracy of kriging model largely depends on the value of θ and the determination of θ requires another optimization process. The golden section refinement is used in this study to solve it. After computing correlation matrix with θ, kriging predictions at untried inputs are predicted and plotted as in the case of ordinary kriging. Once two kriging models are created, their model accuracies are assessed using three error measures in Table 3. The results of validating error of these two kriging methods are summarized in Tables 4 and 5.

6.1 Bird function

The following function is named as "bird"

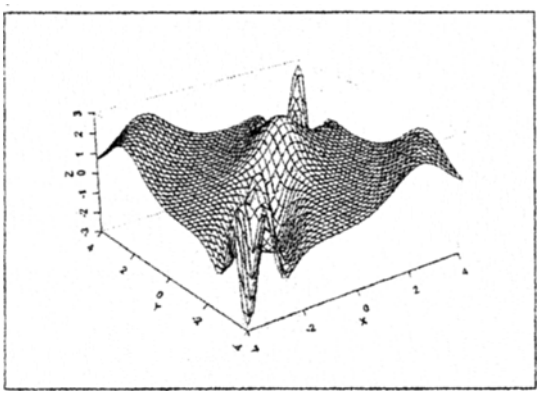

Fig. 1 Exact bird function

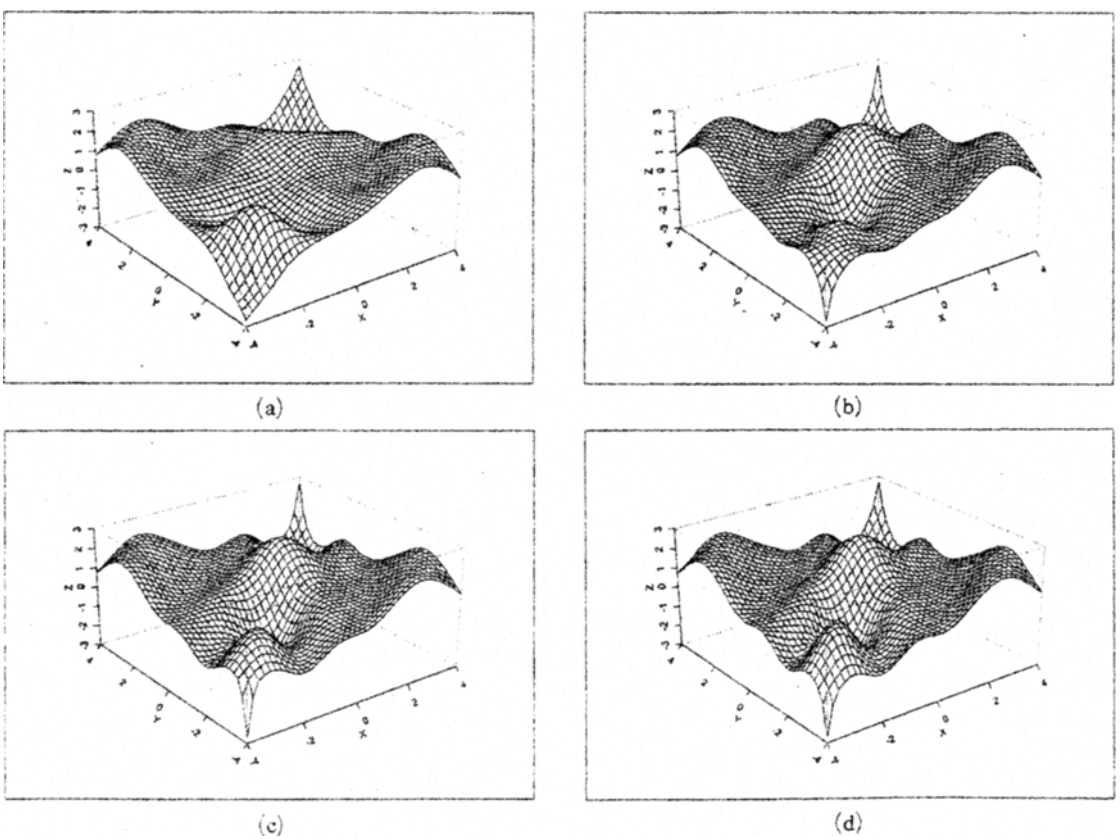

Fig. 2 Approximate bird function using ordinary kriging with (a) 49, (b) 100, (c) 196 and (d) 400 equally space sample points, respectively

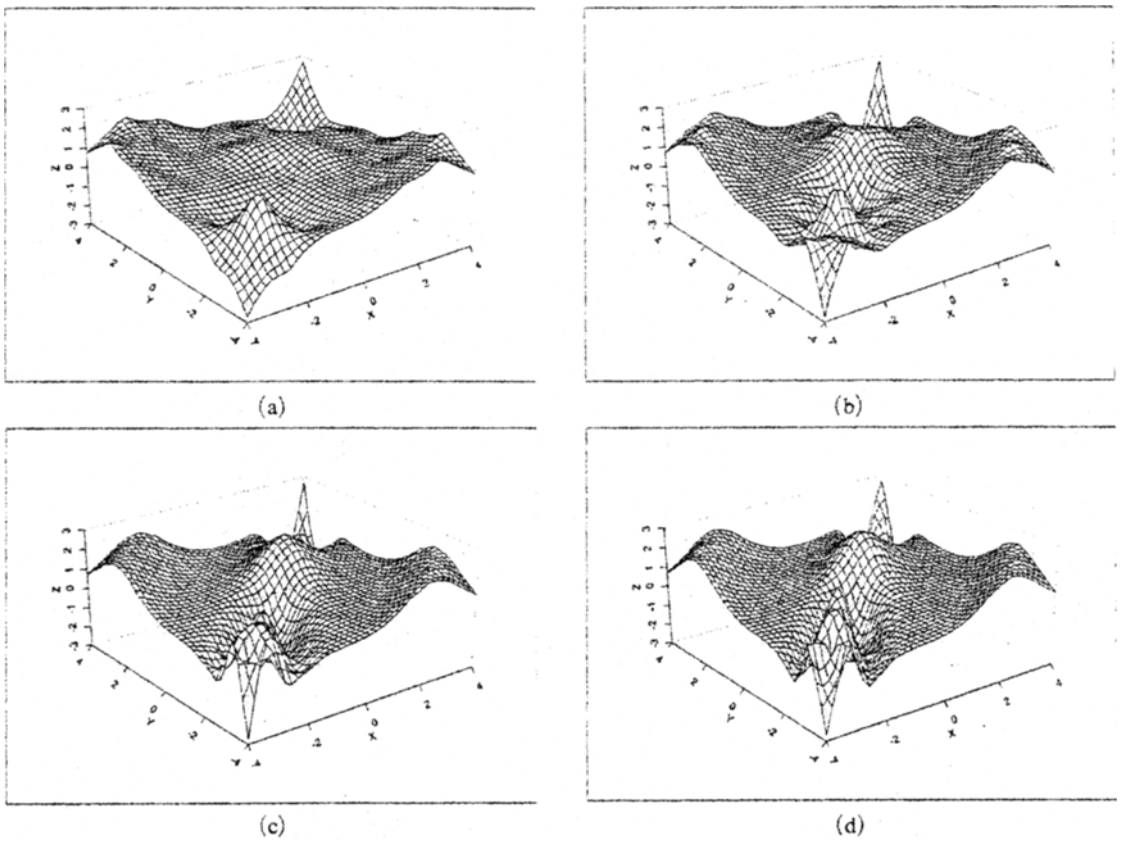

Fig. 3 Approximate bird function using DACE model with (a) 49, (b) 100, (c) 196 and (d) 400 equally space sample points, respectively

function because its surface feature is similar to bird as shown in Fig. 1 and the analytical form of the bird function is given by

$$f(x_1, x_2) = e^{\cos(x_1-x_2)}\sin\left(\frac{(x_1-x_2)^2+x_1+x_2}{1+(x_1-x_2)^2}\right),$$
$$-4 \le x_i \le 4,\ i=1, 2$$

Figure 1 shows exact model of this bird function sampled at $30\times30=900$ equally spaced sample points on $[-4, 4]^2$. To predict the response vales at these sites, ordinary kriging and DACE model are built with equally spaced 49, 100, 196 and 400 sample points as in Fig. 2 and Fig. 3, respectively. Considering the overall model construction, semivariogram parameters are accepted from 900 equally spaced untried sample points. Here, sill=1.0, range=1.4 and nugget=0.2 are selected from the visual fittings. For DACE model, each optimal coefficient of correlation is determined as $\theta=0.96$, $\theta=1.32$, $\theta=1.30$ and $\theta=0.56$ for the same inputs in semivariogram, respectively. The results for validating ordinary kriging and DACE model are listed in Table 4, where maximum absolute error with 100 sample points in ordinary kriging is better than DACE model. In the different cases, the error measures of DACE model are better than ordinary kriging. Thus, if the number of sampled points is increased, the accuracy of the model in the region of interest is greatly improved.

6.2 Rosenbrock's valley function

As a second example of kriging method, kriging prediction of Rosenbrock's valley function is considered

$$f(x_1, x_2) = 100(x_2-x_1^2)^2+(1-x_1)^2,$$
$$-2.5 \le x_i \le 2.5,\ i=1, 2$$

Figure 4 shows exact model of this function sampled at 30×30 equally spaced sample points on $[-2.5, 2.5]^2$. With equally spaced 49, 100, 196 and 400 sample points, two kriging models in Figs. 5 and 6 are constructed and semivariogram parameters are accepted from the $30\times30=900$ sample points for this function as in the case of bird function. With, sill=4e6, range=6 and nugget=1e4. For DACE model, each optimal correlation parameter is determined as $\theta=0.06$, $\theta=0.03$, $\theta=0.0002$ and $\theta=0.0002$ for these sample points, respectively. Because the correlation matrix R in equation (6) is poorly conditioned, the small values of θ are estimated. Sacks, et al. (1989) stated the computation of D-optimal or other efficient design for experiments would share this difficulty. The results for validating ordinary kriging and DACE model are in Table 5. For this case, the error measures of DACE model are better than ordinary kriging as shown in Table 3. Although maximum absolute error in 7×7 sample points is smaller than that in the different number of sample points, the accuracy of the model in the region of interest is greatly improved when the number of sample points is increased.

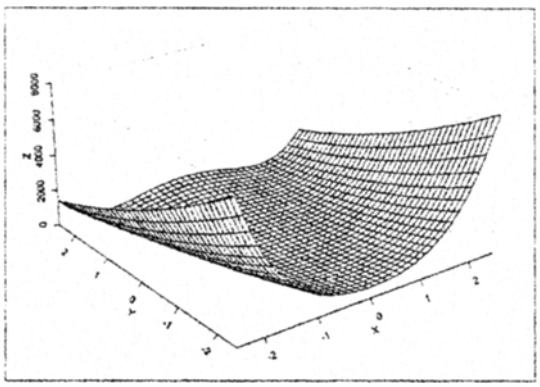

Fig. 4 Exact Rosenbrock's valley function

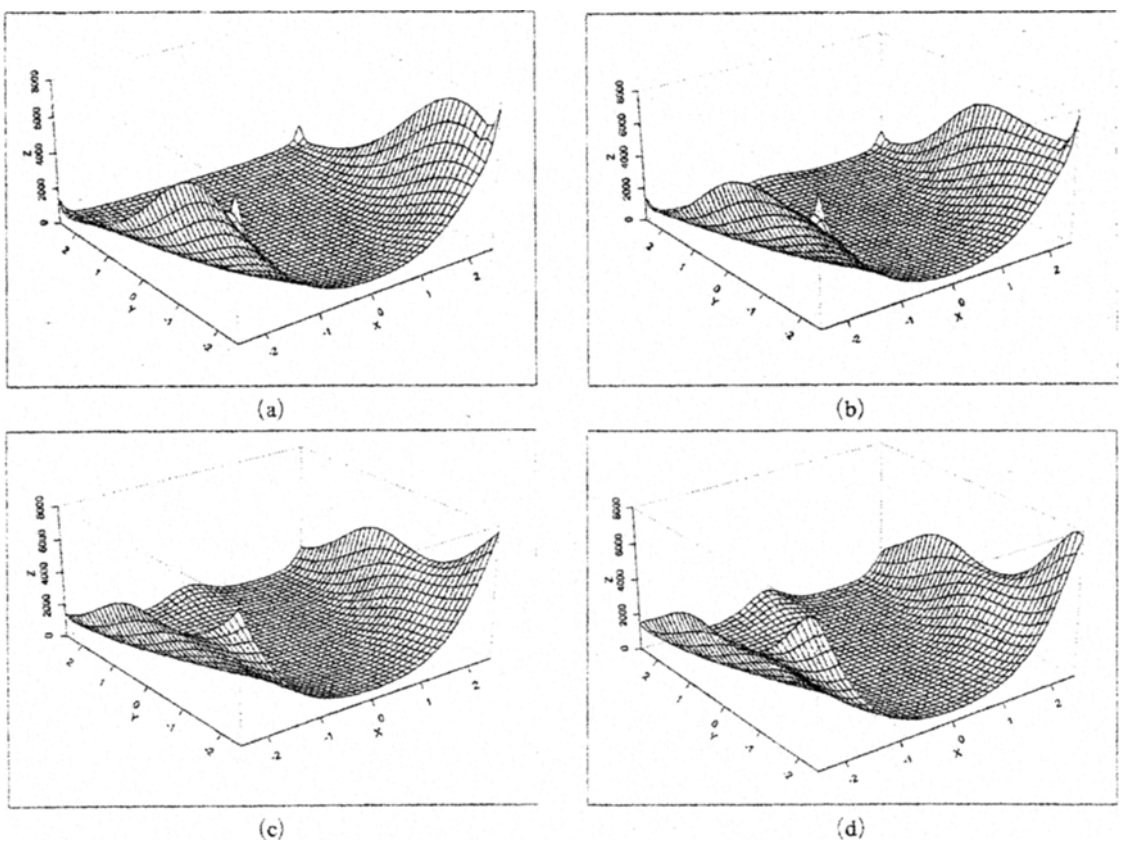

Fig. 5 Approximate Rosenbrock's valley function using ordinary kriging with (a) 49, (b) 100, (c) 196 and (d) 400 equally space sample points, respectively

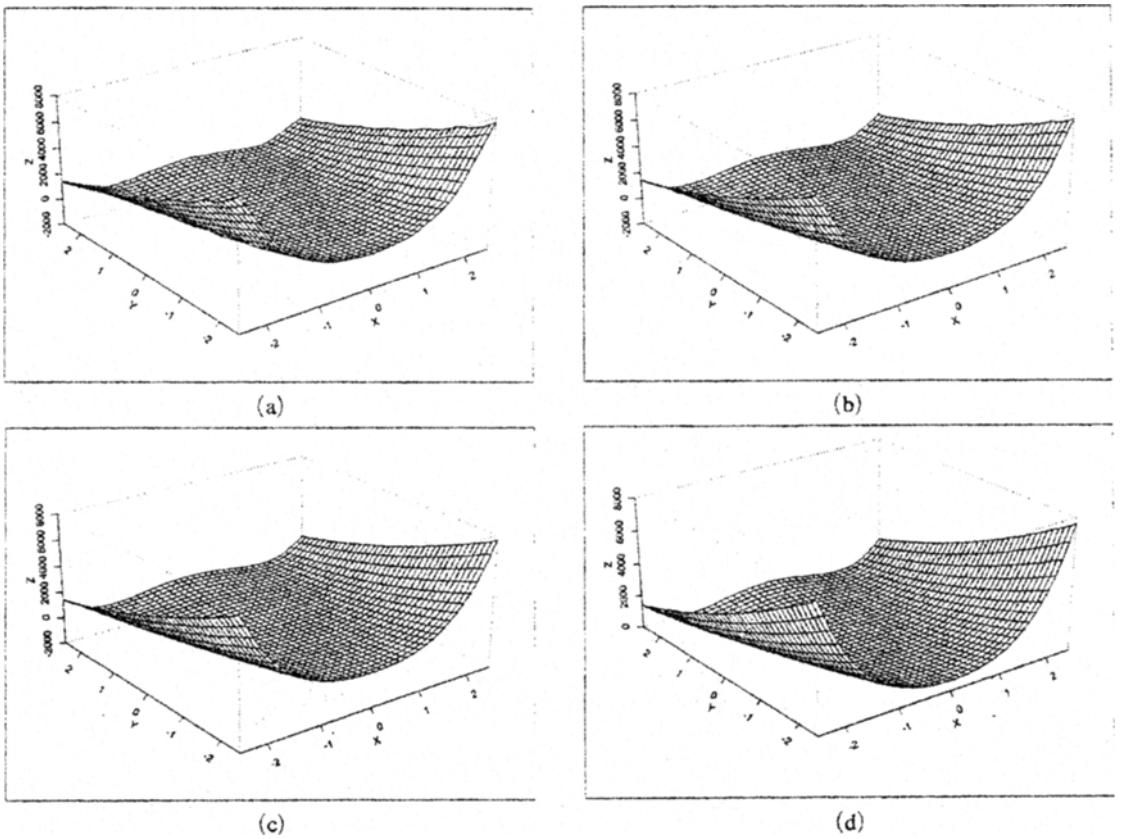

Fig. 6 Approximate Rosenbrock's valley function using DACE model with (a) 49, (b) 100, (c) 196 and (d) 400 equally space sample points, respectively

7. Summary

In this work, we presented the difference between semivariogram in ordinary kriging and correlation function form in DACE model. Ordinary kriging was used for random data with the measurement error, while data in DACE model was deterministic. Nugget, sill and range were used in determining empirical semivariogram and correlation coefficient θ was used in DACE model. In numerical studies, we used bird and rosenbrock's functions with equally spaced sample points. The accuracy of these two kriging models was evaluated through examination of two test functions. The results showed that DACE model was more accurate than ordinary kriging. We speculated that these results were obtained from the visual investigation of the empirical semivariogram and from the deterministic data of the computer simulation suitable for DACE model. We also found that the accuracy of the model in the region of interest was greatly improved if the number of input points was increased.

Finally, it is noted that kriging can be used for prediction, approximation, interpolation, or smoothing. We show that kriging can be used in prediction process. Future work will be included for design problem with more design variable, other sampling technique, and global optimization problem.

Acknowledgements

This research was supported by the center of Innovative Design Optimization Technology (iDOT), Korea Science and Engineering

Foundation.

References

Bacchi, B. and Kottegoda, N. T., 1995, "Identification and Calibration of Spatial Correlation Patterns of Rain Fall," *Journal of Hydrology*, Vol. 165, pp. 311~348.

Barendrengt, L. G., 1987, "The Estimation of the Generalized Covariance When it is a Linear Combination of Two Known Generalized Covariances," *Water Resour. Res.*, Vol. 23, No. 4, pp. 583~590.

Besag, J. E., 1977, "Efficiency of Pseudo Likelihood Estimation for Simple Gaussian Fields," *Biometrika*, Vol. 64, No. 3, pp. 616~618.

Booker et al., 1995, "Global Modeling for Optimization: Boeing/IBM/Rice Collaborative Project," 1995 Final Report, ISSTECH-95-032, The Boeing Company, Seattle, WA.

Booker et al, 1996, "Multi-Level Design Optimization: A Boeing/IBM/Rice Collaborative Project," 1996 Final Report, ISSTECH-96-031, The Boeing Company, Seattle, WA.

Booker, A. J., 1996, "Case Studies in Design and Analysis of Computer Experiments," *Proceedings of the Section on Physical and Engineering Sciences*, American Statistical Association.

Chung, H. S. and Alons, J. J., 2000, "Comparison of Approximation Models with Merit Functions for Design Optimization," *American Institute of Aeronautics and Astronautics*, 8th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, September 6-8, 2000, Long Beach, CA, AIAA-2000-4754, pp. 381~391.

Cox, D. D. and John, S., 1995, "SDO: A Statistical Method for Global Optimization," *Proceedings* of the ICASE/NAS Langley Workshop on Multidisciplinary Optimization, Hampton, March 13-16, VA, SIAM, pp. 315~329.

Cressie, N., 1986, "Kriging Nonstaionary Data," *Journal of the American Statistical Association*, Vol. 81, pp. 625~634.

Cressie, N, 1991, *Statistics for Spatial Data*, John Wiley & sons, New York, pp. 1~143.

Currin et al., 1988, "A Bayesian Approach to the Design and Analysis of Computer Experiments," Technical Report ORNL-6498, Oak Ridge National Laboratory.

Deceneiere et al., 1998, "Applications of Kriging to Image Sequence Cooling," *Signal Processing; Image Communication*, Vol. 13, pp. 227~249.

Dubrule, O., 1983, "Two Methods with Different Objections, Splines and Kriging," *Mathematical Geology*, Vol. 15, pp. 245~257.

Gambolati, G. and Galcati, G., 1985, "Modello Efficiente di Interpolazione Basato Sulta Teoria del Kriging," RapportoTecnico Enel. SI-723/85, CRIS-Serv. Idrol. ENEL-CRIS Laboratory, Mestre, Italy.

Gandin, L. S., 1963, "Objective Analysis of Meteorlolgical Ffelds," Gidrometeorologicheskoe Izdatelstvo (GIMIZ).

Guinta, A. A., 1997, *Aircraft Multidisciplinary Design Optimization Using Design of Experiments Theory and Response Surface Modeling*, Ph. D. Dissertation and MAD Center Report No. 97-05-01, Department of Aerospace and Ocean Engineering, Virginia, Polytechnic Institute and State University, Blacksburg, VA.

Giunta et al., 1998, "A Comparison of Approximation Modeling technique: Polynomial Versus Interpolating Models," 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis & Optimization, September 2-4, St, Louis, MI, AIAA, AIAA-98-4758.

Goldberger, A. S., 1962, "Best Linear Unbiased Prediction in the Generalized Linear Regression Model," *Journal of the American Association*, Vol. 57, pp. 369~375.

Kitanidis, P. K., 1983, "Statistical Estimation of Polynomial Generalized Covariance Functions and Hydrologic Method," *Journal of Hydrology Science*, Vol. 36, No. 3, pp. 223~240.

Krige, D. G., 1951, "A Statistical Approach to Some Basic Mine Valuation Problems on the Witwatersrand," *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, Vol. 52, pp. 119~139.

Mardia et al., 1996, "Kriging and Splines with Derivative Information," Biometrica, Vol. 83, pp. 207~221.

Matheron, G., 1962, "Traite de Geotatistique Appliquee," *Tome I. Memoires du Burean de Recherches Geologigues et Minieres*, No. 14.

Matheron, G., 1963, "Principles of Geostatistics," *Economic Geology*, Vol. 58, pp. 1246~1266.

Matheron, G., 1969, "Le Krigeage Universel," *Les Cahiers du Centre de Morpholgie Mathematique*, Vol. 1, Ecole de Mines, Fonteinbleu, Fronce.

Mitchell, T. J. and Morris, M. D., 1992, "The Spatially Correlation Function Approach to Response Surface Estimation," *Proceedings of the 1992 Winter Simulation Conference*, December 13-16, Arlington, VA, IEEE, pp. 565~571.

Myers, R. H. and Montgomery, D. C., 1995, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, John Wiley & Sons, NY.

Osio, I. G. and Amon, C. H., 1996, "An Engineering Design Methodology with Multistage Bayesian Surrogates and Optimal Sampling," *Research in Engineering Design*, Vol. 8, No. 4, pp. 189~206.

Piazza, et al., 1983, "The Making and Testing of Geographic Gene-Frequency Maps," *Biometrics*, Vol. 37, pp. 635~659.

Ripley, B. D., 1988, *Statistical Inference for Spatial Processes*, Cambridge: Cambridge University Press.

Sacks, et al., 1989, "Design and Analysis of Computer Experiments," *Statistical Science*, Vol. 4, No. 4, pp. 409~435.

Sasena, M. J., 1998, *Optimization of Computer Simulations via Smoothing Splines and Kriging Metamodels*, M. S. Thesis, Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI.

Simpson, T. W., 1998, *A Concept Exploration Method for Product Family Design*, Ph. D. Dissertation, Department of Mechanical Engineering, Georgia Institute of Technology.

Simpson, et al., 1998, "Comparison of Response Surface and Kriging Models for Multidisciplinary Design Optimization," *American Institute of Aeronautics and Astronautics*, AIAA-98-4755, pp. 381~391.

Stephen, et al., 1996, "S+SPATIALSTATS User's Manual," MathSoft, Inc., Seattle, Washington, pp. 73~120.

Torczon, V., 1993, *On the Convergence of Pattern Search Methods,* Technical Report 93-10, Department of Computational and Applied Mathematics, Rice University, Houston, TX 77251-1892.

Trosset, M. W. and Torczon, V., 1997, *Numerical Optimization Using Computer Experiments,* Report No. TR97-02, Department of Computational and Applied Mathematics, Rice University, Houston, TX.

Volpi, G. and Gambolati, G., 1978, "On the Use of a Main Trend for the Kriging Technique in Hydrology," *Adv. Water Resourvoir.*, Vol. 1, No. 6, pp. 345~349.

Webster, R., 1985, *Quantitative Spatial Analysis of Soil in the Field*, Advances in Soil science, 3, B. A. Stewart(ed.), New York, Springer-Verlag, pp. 1~70.