# Derivative-Enhanced Variable Fidelity Surrogate Modeling for Aerodynamic Functions

Wataru YAMAZAKI[1]
*Assistant Professor, Nagaoka University of Technology, Nagaoka, Niigata, 940-2188, Japan*

*and*

Dimitri J. MAVRIPLIS[2]
*Professor, University of Wyoming, Laramie, Wyoming, 82072, USA*

**In this paper, a derivative-enhanced, variable fidelity surrogate model approach is developed based on a direct Kriging formulation. In this approach, the absolute values of a high-fidelity function as well as the trends obtained by low-fidelity function values are utilized to develop an accurate surrogate model. Derivative information of arbitrary fidelity levels can be also utilized to develop a more accurate surrogate model. The efficiencies of the developed approaches are investigated by analytical function fitting, aerodynamic data modeling and 2D airfoil drag minimization problems. The derivative-enhanced variable fidelity surrogate model approach is shown to be useful for efficient aerodynamic database construction and in the development of both efficient design optimization and uncertainty analysis.**

## Nomenclature

| | | |
|---|---|---|
| $EI$ | = | expected improvement value |
| $\mathbf{F}$ | = | regression matrix |
| $M_\infty$ | = | freestream Mach number |
| $m$ | = | dimensionality of design space (number of design variables) |
| $n_l$ | = | number of sample points at $l$-th fidelity level |
| $\mathbf{r}$ | = | correlation vector |
| $\mathbf{R}$ | = | correlation matrix |
| $R_{l_1 l_2}$ | = | correlation function between fidelity levels $l_1$ and $l_2$ |
| $s_{(\mathbf{x})}$ | = | standard error of Kriging model at $\mathbf{x}$ |
| $scf$ | = | spatial correlation function |
| $w$ | = | Kriging weight coefficients |
| $\mathbf{x}$ | = | location vector in design variable space |
| $y$ | = | exact function value |
| $\hat{y}_{(\mathbf{x})}$ | = | estimated function value at $\mathbf{x}$ |
| $Z_{(\mathbf{x})}$ | = | random process model |
| $\alpha$ | = | angle of attack [deg] |
| $\beta$ | = | constant (regression) model |
| $\Delta\mathbf{x}$ | = | step size vector for indirect Kriging approach |
| $\mu$ | = | Lagrange multiplier |
| $\sigma^2$ | = | model variance |
| $\sigma_D$ | = | standard deviation of all parameters |
| $\theta$ | = | hyper parameter for spatial correlation function |

[1] Assistant Professor, Department of Mechanical Engineering, Nagaoka University of Technology, Member AIAA
[2] Professor, Department of Mechanical Engineering, University of Wyoming, Associate Fellow AIAA

## I. Introduction

Surrogate model approaches have attracted increased attention recently in aerospace engineering since they offer substantial benefits for design optimization, aerodynamic database construction, and uncertainty analysis. The idea of a surrogate model approach is to replace expensive functional evaluations (i.e. high-fidelity computational fluid dynamics (CFD) simulations) with an analytical model which is constructed through selective sampling of the high-fidelity model. When a surrogate model is constructed with given exact function data, a designer can efficiently explore the approximated design space at very low computational cost. Uncertainty analysis can be also executed efficiently on the surrogate model, which is referred to as inexpensive Monte-Carlo (IMC) simulation approach. In Ref.[1], the performances of major surrogate models, such as least square polynomial, multi-layer perceptron, radial basis function (RBF) and Kriging, have been compared for a two dimensional turbomachinery problem. The Kriging and RBF models showed the best performance during this investigation. The Kriging model, which was originally developed in the field of geological statistics, has often been found to perform well in other engineering fields and has thus gained popularity in aerospace engineering and design [1-14]. This surrogate model predicts the function value by using stochastic processes, and has the flexibility to represent multimodal/nonlinear functions.

For accurate aerodynamic data modeling and uncertainty analysis, more accurate surrogate model should be essential. One of the major approaches to increase the accuracy of surrogate model is to utilize derivative information of aerodynamic functions. Because efficient gradient evaluation methods based on adjoint formulations have been developed and applied successfully in the field of aerodynamics [15-21], the introduction of gradient information within surrogate models as additional input data has also attracted attention. Two gradient-enhanced Kriging (called cokriging or GEK) approaches have been developed and beneficial results have been shown in the literature [6,9,10]. Additionally, efficient calculation methods of CFD Hessian and Hessian vector product have been developed recently [22-24]. In this approach, the Hessian components of a CFD functional output with respect to design variables can be efficiently calculated by using the adjoint method and automatic differentiation (AD) tools such as TAPENADE [25]. Thus it is promising to utilize derivative information within surrogate model to enhance the accuracy of the surrogate model.

Utilizing low fidelity function values as secondary information is an alternative approach to further improve the accuracy of surrogate models [11,12]. A same kind of concept was originally investigated in optimization communities, which is often referred to as variable fidelity (VF) model approach or model management optimization (MMO) [26]. In the VF surrogate model approach, the trends of low-fidelity function values as well as a small number of high-fidelity function absolute values are simultaneously utilized to construct a surrogate model. This approach is promising in the field of aerospace engineering since one can define many sets of different fidelity models as shown in Table.1.

In this paper, a derivative-enhanced variable fidelity surrogate model approach is proposed based on a Kriging formulation. The efficiency is investigated using analytical function fitting, aerodynamic data modeling and 2D airfoil drag minimization problems.

**Table.1 Examples of Different Fidelity Models**

| High-fidelity model | Low-fidelity model |
|---|---|
| Experimental data | CFD result |
| Accurate physical model (e.g. RANS) | Cheap physical model (e.g. Inviscid Euler) |
| Finer mesh CFD result | Coarser mesh CFD result |
| Converged solution | Loose converged solution |

## II. Kriging Formulations

In this section, a derivative-enhanced VF Kriging approach is proposed based on a direct formulation. First, the basics of conventional Kriging, as well as direct and indirect derivative-enhanced Kriging approaches are reviewed. Then, a VF model formulation is introduced.

### A. Conventional Kriging, Derivative-enhanced Kriging Approaches

The Kriging method is a statistical prediction of a function from a set of exact function values obtained with arbitrary design variables. Kriging prediction depends on spatial correlations between given sample points. The correlation is given by a correlation function which is only dependent on the distance between two points. These

correlations are considered in a matrix (called correlation matrix) which appears in the Kriging formulation. The size of matrix is $n_1 \times n_1$, where $n_1$ is the number of (high-fidelity) sample points. The details of formulation are given in Subsection II.B for the case of a VF Kriging model.

For derivative-enhanced Kriging, derivative information at several sample points is also used for the construction of the Kriging model. In the case of direct approach, the derivative information is directly used in the modified Kriging formulation by including the correlations between function/derivative and derivative/derivative. These correlations can be modeled by differentiating the correlation function. If all sample points have function and gradient information, the size of the correlation matrix is $n_1(m+1)$, where $m$ is the dimensionality of the design space (i.e. number of design variables).

The formulation of indirect approach is exactly the same as that of the original Kriging model. A difference in this approach is to construct additional sample points around real sample points that have derivative information. When a sample point has gradient information, the additional sample points are defined by a first-order Taylor approximation as follows:

$$\begin{aligned} \mathbf{x}^{add} &= \mathbf{x} + \Delta\mathbf{x} \\ y_{(\mathbf{x}^{add})} &= y_{(\mathbf{x})} + \Delta\mathbf{x}^T \left[ \frac{\partial y_{(\mathbf{x})}}{\partial \mathbf{x}} \right] \end{aligned} \tag{2.1}$$

where $\mathbf{x}$ and $y_{(\mathbf{x})}$ are the location and exact function value of a real sample point, respectively. $\mathbf{x}^{add}$ and $y_{(\mathbf{x}^{add})}$ are the location and approximated function value of an additional sample point. $\Delta\mathbf{x}$ is a user-specified step size vector for the relative location of additional sample points. Usually, one additional point is created in each direction of the design variable space for gradient-enhanced cases (in total $m$ additional points per real sample point). After augmentation of the sample points, the indirect Kriging model is created by using both real and additional sample point information. If all sample points have function and gradient information, the total number of sample points is $n_1(m+1)$, which is the same as the size of the direct GEK correlation matrix. The advantage of the indirect approach over the direct approach is the ease of implementation. On the other hand, the major disadvantage of indirect approach is the issue associated with the step size $|\Delta\mathbf{x}|$ used for creating additional sample points. Very small step sizes can lead to an ill-conditioned correlation matrix in the Kriging formulation due to the closely spaced sample points. The step size is usually determined by considering both the accuracy of the Taylor approximation and the matrix conditioning.

According to Ref.[9], direct and indirect GEK models are identical when they are constructed by a same set of parameters and a reasonable step size. However, ill-conditioning can be avoided only by the direct GEK approach. Gradient/Hessian-enhanced Kriging approaches have been developed by present authors [13] in which direct and indirect formulations were compared. We concluded that the direct approach is preferable, because there is no sensitive parameter in the direct formulation and derivative information can be exactly enforced at real sample locations. Furthermore, the matrix conditioning is much better than that of the indirect approach since additional sample points are not created in the direct formulation.

### B. Variable Fidelity Kriging Approach

In this subsection, a VF Kriging formulation is derived. Our VF Kriging formulation is quite similar with that of Ref.[12]. In this paper, the major derivations are introduced for a case in which function values of three fidelity levels are used to construct a Kriging model. We also extend the VF approach to include derivative information.

The formulation appearing in this paper is based on the "ordinary Kriging" model for the simplicity of the description. The high and low fidelity functions are replaced by the following random functions:

$$\hat{y}_{l(\mathbf{x})} = \beta_l + Z_{l(\mathbf{x})} \tag{2.2}$$

where $l$ (=1,2,3) means the index of fidelity level. $\hat{y}_{l(\mathbf{x})}$ is the prediction of an function at an arbitrary location of $\mathbf{x}$. The first term $\beta_l$ is a constant model (or low-order polynomial regression model in the formulation of "universal Kriging" model [7]) and the second term $Z_{l(\mathbf{x})}$ represents a random process model with zero mean, variance $\sigma_l^2$ and the covariance of two locations $\mathbf{x}_i$ and $\mathbf{x}_j$ is given as follows:

$$\mathrm{cov}_{ij}^{l_1 l_2} = Cov\left[Z_{l_1(\mathbf{x}_i)}, Z_{l_2(\mathbf{x}_j)}\right] = \sigma_{l_1}\sigma_{l_2} R_{l_1 l_2}\left(\mathbf{x}_i, \mathbf{x}_j\right) \quad \left(l_1, l_2 = 1,2,3\right) \tag{2.3}$$

$R_{l_1 l_2}$ is the correlation function between fidelity levels $l_1$ and $l_2$. Then, a linear combination of the high and low-fidelity information at given sample points is considered as follows:

$$\hat{y}_{(\mathbf{x})} = \sum_{i=1}^{n_1} w_{1i} y_{1i} + \sum_{j=1}^{n_2} w_{2j} y_{2j} + \sum_{k=1}^{n_3} w_{3k} y_{3k} = \mathbf{w}_1^T \mathbf{y}_1 + \mathbf{w}_2^T \mathbf{y}_2 + \mathbf{w}_3^T \mathbf{y}_3 \tag{2.4}$$

where $\mathbf{y}_l$ and $\mathbf{w}_l$ are respectively the observed function and their weight coefficients. The fidelity level 1 is considered as high-fidelity data, and $n_1 << n_2, n_3$ is the most general situation in our practical applications. The Kriging approach finds the best linear unbiased predictor which minimizes the mean square error (MSE):

$$MSE\left[\hat{y}_{(\mathbf{x})}\right] = E\left[\sum_{i=1}^{n_1} w_{1i} y_{1i} + \sum_{j=1}^{n_2} w_{2j} y_{2j} + \sum_{k=1}^{n_3} w_{3k} y_{3k} - \hat{y}_{1(\mathbf{x})}\right] \tag{2.5}$$

subject to the unbiasedness constraint as follows:

$$E\left[\sum_{i=1}^{n_1} w_{1i} y_{1i} + \sum_{j=1}^{n_2} w_{2j} y_{2j} + \sum_{k=1}^{n_3} w_{3k} y_{3k}\right] = E\left[\hat{y}_{1(\mathbf{x})}\right] \tag{2.6}$$

The weight coefficients can be found by solving this constrained minimization problem with the Lagrange multiplier approach as follows:

$$J = MSE\left[\hat{y}_{(\mathbf{x})}\right] + \mu_1\left(\sum_{i=1}^{n_1} w_{1i} - 1\right) + \mu_2 \sum_{j=1}^{n_2} w_{2j} + \mu_3 \sum_{k=1}^{n_3} w_{3k} \tag{2.7}$$

where $\mu_l$ are the Lagrange multipliers. The weight coefficients are solved from $\partial J/\partial w_{li} = \partial J/\partial \mu_l = 0$ as:

$$\begin{aligned}
&\sum_{i=1}^{n_1} w_{1i}\,\mathrm{cov}_{hi}^{11} + \sum_{j=1}^{n_2} w_{2j}\,\mathrm{cov}_{hj}^{12} + \sum_{k=1}^{n_3} w_{3k}\,\mathrm{cov}_{hk}^{13} + \frac{1}{2}\mu_1 &&= Cov\left[Z_{1(\mathbf{x}_h)}, Z_{1(\mathbf{x})}\right] && \left(h = 1,\cdots,n_1\right)\\
&\sum_{i=1}^{n_1} w_{1i}\,\mathrm{cov}_{hi}^{12} + \sum_{j=1}^{n_2} w_{2j}\,\mathrm{cov}_{hj}^{22} + \sum_{k=1}^{n_3} w_{3k}\,\mathrm{cov}_{hk}^{23} + \frac{1}{2}\mu_2 &&= Cov\left[Z_{2(\mathbf{x}_h)}, Z_{1(\mathbf{x})}\right] && \left(h = 1,\cdots,n_2\right)\\
&\sum_{i=1}^{n_1} w_{1i}\,\mathrm{cov}_{hi}^{13} + \sum_{j=1}^{n_2} w_{2j}\,\mathrm{cov}_{hj}^{23} + \sum_{k=1}^{n_3} w_{3k}\,\mathrm{cov}_{hk}^{33} + \frac{1}{2}\mu_3 &&= Cov\left[Z_{3(\mathbf{x}_h)}, Z_{1(\mathbf{x})}\right] && \left(h = 1,\cdots,n_3\right)\\
&\sum_{i=1}^{n_1} w_{1i} = 1 &&&&\\
&\sum_{j=1}^{n_2} w_{2j} = 0 &&&&\\
&\sum_{k=1}^{n_3} w_{3k} = 0 &&&&
\end{aligned} \tag{2.8}$$

These equations finally yield the following system of equations:

$$\begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \mathbf{R}_{13} & 1 & 0 & 0 \\ \mathbf{R}_{12} & \mathbf{R}_{22} & \mathbf{R}_{23} & 0 & 1 & 0 \\ \mathbf{R}_{13} & \mathbf{R}_{23} & \mathbf{R}_{33} & 0 & 0 & 1 \\ 1^T & 0^T & 0^T & 0 & 0 & 0 \\ 0^T & 1^T & 0^T & 0 & 0 & 0 \\ 0^T & 0^T & 1^T & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{w}}_1 \\ \widetilde{\mathbf{w}}_2 \\ \widetilde{\mathbf{w}}_3 \\ \widetilde{\mu}_1 \\ \widetilde{\mu}_2 \\ \widetilde{\mu}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \mathbf{r}_3 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad \begin{pmatrix} \widetilde{\mathbf{w}}_1 = \mathbf{w}_1 \\ \widetilde{\mathbf{w}}_2 = \mathbf{w}_2 \, \sigma_2 / \sigma_1 \\ \widetilde{\mathbf{w}}_3 = \mathbf{w}_3 \, \sigma_3 / \sigma_1 \\ \widetilde{\mu}_1 = \mu_1 / \left(2\sigma_1^2\right) \\ \widetilde{\mu}_2 = \mu_2 / \left(2\sigma_1 \sigma_2\right) \\ \widetilde{\mu}_3 = \mu_3 / \left(2\sigma_1 \sigma_3\right) \end{pmatrix} \tag{2.9}$$

and then it can be described as the following matrix form:

$$\begin{bmatrix} \mathbf{R} & \mathbf{F} \\ \mathbf{F}^T & 0 \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{w}} \\ \widetilde{\mu} \end{bmatrix} = \begin{bmatrix} \mathbf{r} \\ \mathbf{L} \end{bmatrix} \tag{2.10}$$

where $\mathbf{R} \in \Re^{(n_1+n_2+n_3)\times(n_1+n_2+n_3)}$, $\mathbf{r} \in \Re^{(n_1+n_2+n_3)}$ and $\mathbf{F} \in \Re^{(n_1+n_2+n_3)\times 3}$ are respectively the correlation matrix, correlation vector and regression matrix. $\mathbf{R}$ expresses the correlations between all observed data, and $\mathbf{r}$ is for the correlations between the observed data and location $\mathbf{x}$. The unknown vector $\widetilde{\mathbf{w}}$ is determined by inverting the matrix of Eq.(2.10). The final form of the VF Kriging approach is:

$$\hat{y}_{(\mathbf{x})} = \mathbf{L}^T \widetilde{\beta} + \mathbf{r}_{(\mathbf{x})}^T \mathbf{R}^{-1} \left( \mathbf{Y} - \mathbf{F} \widetilde{\beta} \right) \tag{2.11}$$

where

$$\begin{aligned} &\widetilde{\beta} = \left( \mathbf{F}^T \mathbf{R}^{-1} \mathbf{F} \right)^{-1} \mathbf{F}^T \mathbf{R}^{-1} \mathbf{Y} \in \Re^{3\times 1} \\ &\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 \, \sigma_1/\sigma_2 & \mathbf{y}_3 \, \sigma_1/\sigma_3 \end{bmatrix}^T \in \Re^{(n_1+n_2+n_3)} \end{aligned} \tag{2.12}$$

The matrix form of Eq.(2.11) is quite similar with that of the original Kriging formulation. The factors $\sigma_1/\sigma_2$ and $\sigma_1/\sigma_3$ are additional parameters required for this VF formulation, that take the influence of low-fidelity data into consideration for the VF function prediction. The MSE of Eq.(2.5) can be expressed as:

$$MSE\left[\hat{y}_{(\mathbf{x})}\right] = s_{(\mathbf{x})}^2 = \sigma_1^2 \left[ 1 - \mathbf{r}^T \mathbf{R}^{-1} \mathbf{r} + \varphi^T \left( \mathbf{F}^T \mathbf{R}^{-1} \mathbf{F} \right)^{-1} \varphi \right] \qquad \left( \varphi = \mathbf{F}^T \mathbf{R}^{-1} \mathbf{r} - \mathbf{L} \right) \tag{2.13}$$

We have further extended this VF formulation to include derivative information. This can be achieved in the same manner as described in Ref.[13]. The derivative-enhanced Kriging formulation is based on a linear combination of observed functions and their derivative components of gradient, Hessian and Hessian vector product. The following additional covariance terms are considered in the derivative-enhanced model:

$$\begin{aligned} Cov\left[ Z_{l_1(\mathbf{x}_i)}, \frac{\partial Z_{l_2(\mathbf{x}_j)}}{\partial x^k} \right] &= \sigma_{l_1} \sigma_{l_2} \frac{\partial R_{l_1 l_2}\left(\mathbf{x}_i, \mathbf{x}_j\right)}{\partial x^k} \\ Cov\left[ Z_{l_1(\mathbf{x}_i)}, \sum_{k_2=1}^{m} \frac{\partial^2 Z_{l_2(\mathbf{x}_i)}}{\partial x^{k_1} \partial x^{k_2}} v_{k_2} \right] &= \sigma_{l_1} \sigma_{l_2} \sum_{k_2=1}^{m} \frac{\partial^2 R_{l_1 l_2}\left(\mathbf{x}_i, \mathbf{x}_j\right)}{\partial x^{k_1} \partial x^{k_2}} v_{k_2} \end{aligned} \tag{2.14}$$

$$Cov\left[ Z_{l_1(\mathbf{x}_i)}, \frac{\partial^2 Z_{l_2(\mathbf{x}_i)}}{\partial x^{k_1} \partial x^{k_2}} \right] = \sigma_{l_1} \sigma_{l_2} \frac{\partial^2 R_{l_1 l_2}\left(\mathbf{x}_i, \mathbf{x}_j\right)}{\partial x^{k_1} \partial x^{k_2}}$$

$$\cdots \tag{2.14contd}$$

$$Cov\left[ \frac{\partial^2 Z_{l_1(\mathbf{x}_i)}}{\partial x^{k_1} \partial x^{k_2}}, \frac{\partial^2 Z_{l_2(\mathbf{x}_j)}}{\partial x^{k_3} \partial x^{k_4}} \right] = \sigma_{l_1} \sigma_{l_2} \frac{\partial^4 R_{l_1 l_2}\left(\mathbf{x}_i, \mathbf{x}_j\right)}{\partial x^{k_1} \partial x^{k_2} \partial x^{k_3} \partial x^{k_4}}$$

where $v_{k_2}$ is a known vector component which is defined to obtain a Hessian vector product.

**C. Correlation Functions**

The correlation matrix $\mathbf{R}$ and vector $\mathbf{r}$ are specified by a spatial correlation function and its derivatives. General correlation functions depend only on the distance between two locations as follows:

$$R\left(\mathbf{x}_i, \mathbf{x}_j\right) = \prod_{k=1}^{m} scf\left(\theta_k, d_k^{ij}\right) \qquad \left(d_k^{ij} = \left|x_i^k - x_j^k\right|\right) \tag{2.15}$$

where $\theta_k$ is the hyper parameter for $k$-th design variable. Gaussian or cubic spline functions are the most common forms for the spatial correlation function [9,13]. The hyper parameter expresses the distance weight for both spatial correlation functions. Generally, cubic spline functions yield a better conditioned correlation matrix than Gaussian functions. However, the cubic spline functions cannot be used in our practical cases since the fourth-derivatives of the correlation function are required as indicated in Eq.(2.14). Therefore, the following RBF is used for the spatial correlation function in this research:

$$scf\left(\theta_k, d_k^{ij}\right) = \begin{cases} \frac{1}{3}\left(1 - \theta_k d_k^{ij}\right)^6 \left(35\theta_k^2 {d_k^{ij}}^2 + 18\theta_k d_k^{ij} + 3\right) & for \;\; d_k^{ij} \le 1/\theta_k \;\; \left(\theta_k > 0\right) \\ 0 & else \end{cases} \tag{2.16}$$

This function also yields a better conditioned correlation matrix than the Gaussian function. The derivatives up to fourth-order are calculated by employing the TAPENADE automatic differentiation tool.

Theoretically, different sets of hyper parameters $\theta^{l_1 l_2}$ can be used for the VF Kriging model. In this research, however, a common set of $\theta$ is used for all correlation functions for simplicity as:

$$\theta = \theta^{11} = \theta^{12} = \theta^{13} = \theta^{22} = \theta^{23} = \theta^{33} \tag{2.17}$$

The correlation function can then be defined as follows:

$$R = R_{11} = R_{22} = R_{33} \qquad R_{12} = R_{13} = R_{23} = \rho R \tag{2.18}$$

where $\rho$ is a relaxation factor (set to 0.9999) to avoid an ill-conditioned correlation matrix when high and low-fidelity sample locations are very close together.

**D. Model Fitting by Maximum Likelihood Estimation**

The VF Kriging model still includes the undetermined parameters $\theta$, $\sigma_1^2$, $\sigma_1/\sigma_2$ and $\sigma_1/\sigma_3$. These parameters can be estimated by maximizing the likelihood (joint probability) function of the given samples [12]. This empirical approach finds the parameters which are most consistent with the sample data. Optimal mean and variance are analytically determined as:

$$\tilde{\beta} = \left(\mathbf{F}^T \mathbf{R}^{-1} \mathbf{F}\right)^{-1} \mathbf{F}^T \mathbf{R}^{-1} \mathbf{Y} \tag{2.19}$$

$$\sigma_1^2 = \frac{1}{n_1 + n_2 + n_3} \left(\mathbf{Y} - \mathbf{F}\tilde{\beta}\right)^T \mathbf{R}^{-1} \left(\mathbf{Y} - \mathbf{F}\tilde{\beta}\right) \tag{2.20}$$

One can notice that the constant model of Eq.(2.19) is equal to that of Eq.(2.12). The model standard deviation factors are also analytically determined as:

$$\frac{\sigma_1}{\sigma_l} = \left(\mathbf{Y}_l^T \mathbf{R}^{-1} \mathbf{Y}_l\right)^{-1} \mathbf{Y}_l^T \mathbf{R}^{-1} \left(\mathbf{F}\tilde{\beta} - \mathbf{Y}_{-l}\right) \quad (l = 2,3) \tag{2.21}$$

where

$$\begin{aligned} \mathbf{Y}_l &= \begin{bmatrix} 0 & \mathbf{y}_l & 0 \end{bmatrix}^T \\ \mathbf{Y}_{-l} &= \begin{bmatrix} \mathbf{y}_1 & 0 & \mathbf{y}_k \sigma_1 / \sigma_k \end{bmatrix}^T \qquad (k \neq l) \end{aligned} \tag{2.22}$$

Since there is no analytical form solutions for $\theta$, these are estimated by a numerical optimization which corresponds to the maximization of a likelihood function. For given parameters $\theta$, $\tilde{\beta}$, $\sigma_1^2$, $\sigma_1/\sigma_2$ and $\sigma_1/\sigma_3$, the following log-likelihood function is calculated:

$$\ln\left(L(\theta)\right) = -\frac{1}{2}\left(n_1 + n_2 + n_3\right)\ln\left(\sigma_1^2\right) - \frac{1}{2}\ln\left(\left|\mathbf{R}\right|\right) \tag{2.23}$$

The most consistent $\theta$ is determined through the maximization of the log-likelihood function. One of the issues associated with the maximum likelihood estimation is the multimodality of the log-likelihood function. Therefore, some global optimization method, such as genetic algorithms (GA), is preferable to find the global optimal parameters. However, the computational cost for this optimization problem can be relatively expensive above a certain number of sample points. This is because the evaluation of a likelihood function requires the inversion of $\mathbf{R}$ and the calculation of its determinant, and a large number of function evaluations (matrix inversions) with different sets of $\theta$ is required for a global optimization. In our approach, a master-slave type MPI-parallelized GA [27,28] is used for the optimization. The correlation matrix inversion and the determinant evaluation are solved with a Cholesky decomposition method.

**E. Addition of New Sample Points by Expected Improvement**

Once a Kriging model is constructed with given information, design optimization or uncertainty analysis can be realized on the surrogate model at low computational cost. In its design optimization process, new sample point information is added and then the surrogate model is iteratively updated. It is a straightforward approach to find the new sample points locations by minimizing the estimated objective function on the surrogate model. However, this approach usually converges to a local optimal point depending on the locations of initial sample points, because it does not take into account the uncertainty of the surrogate model.

Expected Improvement (EI) [4] expresses a potential for improvement which considers both estimated function value and uncertainty in the surrogate model. At a point $\mathbf{x}$, the EI for minimization problems is given as follows:

$$EI_{(\mathbf{x})} = \begin{cases} \left(y_{\min} - \hat{y}_{(\mathbf{x})}\right)\Phi\left(\dfrac{y_{\min} - \hat{y}_{(\mathbf{x})}}{s_{(\mathbf{x})}}\right) + s_{(\mathbf{x})}\phi\left(\dfrac{y_{\min} - \hat{y}_{(\mathbf{x})}}{s_{(\mathbf{x})}}\right) & if \;\; s_{(\mathbf{x})} > 0 \\ 0 & if \;\; s_{(\mathbf{x})} = 0 \end{cases} \tag{2.24}$$

where $\Phi$ and $\phi$ are the normal cumulative distribution function and probability density function, respectively. $s_{(\mathbf{x})}$ is the standard error of surrogate model expressed by Eq.(2.13). $y_{\min}$ is the minimum (optimal) function value among all given sample points. The new sample point is added in the location where the EI is maximal. This EI-based iterative approach can increase the accuracy of Kriging model efficiently by maintaining a balance between global and local search criteria. In our approach, a master-slave type MPI-parallelized GA is used for the maximization of EI.

## III. Validation in Analytical Function Problems

In this section, the developed derivative-enhanced and/or VF surrogate model is used to fit analytical functions with various sets of sample points. The sets of sample points are generated by a Latin Hypercube Sampling (LHS) method [3].

### A. 1D Analytical Function

In this subsection, a 1D function is approximated by a high-fidelity (exact) function as well as two low-fidelity functions. The exact function is shown in Fig.1a. Two high-fidelity sample points are created and then single fidelity (SF) Kriging model (conventional Kriging model) is created in Fig.1b. The SF Kriging model is quite inaccurate because of the shortage of function information. Two low-fidelity functions are defined by adding/subtracting a constant to/from the exact function. Five low-fidelity sample points are created for both low-fidelity functions in limited design spaces as Fig.1c. In Fig.1c, two VF models results are also shown by using one of the sets of the low-fidelity samples. It can be understood that the accuracy of estimated function is increased at a region where there are low-fidelity samples. In Fig.1d, a VF model is constructed by using the high-fidelity samples as well as both sets of low-fidelity samples. It shows good agreement with the exact function in all design space. Thus, the VF Kriging model can be constructed by the absolute values of a high-fidelity function as well as the trends obtained by low-fidelity functions.

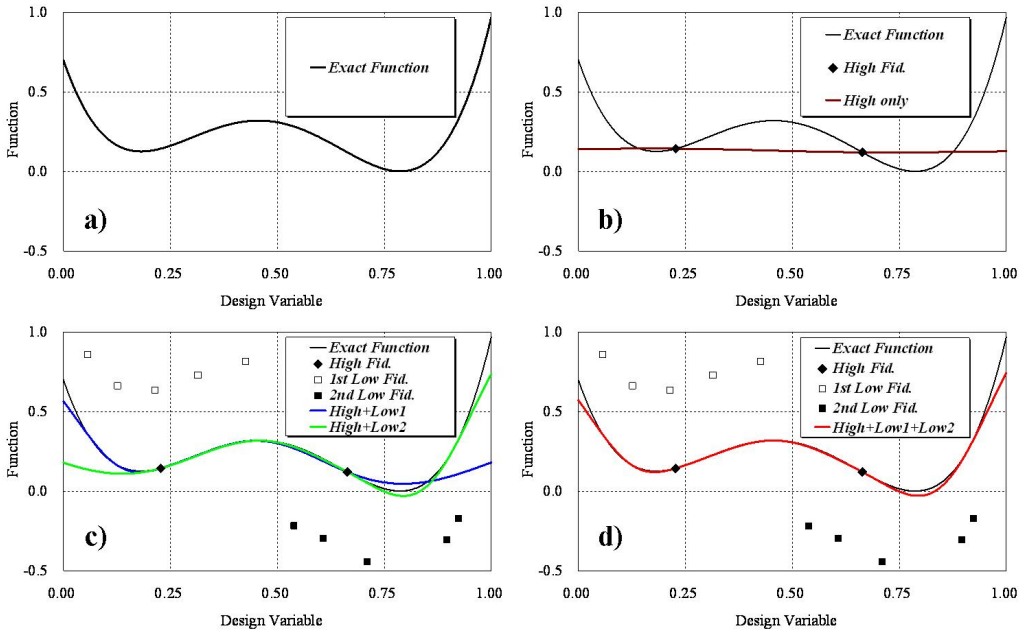


**Fig.1 Comparison in 1D Analytical Function**

**B. 2D Cosine Function**

Next, the developed surrogate model approaches are used to fit a 2D Cosine function by using various sets of sample points (from 0 to 50 for both high/low-fidelity samples) generated by LHS. The Cosine function is defined as:

$$Cosine_{(\mathbf{x})} = \cos\left(\sum_{k=1}^{m} x_k\right) \qquad \left(-2 \le x_k \le 2\right) \tag{3.1}$$

Once a surrogate model is constructed, the accuracy of the model is evaluated by using the root mean squared error (RMSE) between the exact function and approximate function values given by the surrogate model. The RMSE is given as follows:

$$RMSE = \frac{1}{M}\sqrt{\sum_{i=1}^{M}\left(\hat{y}_{(\mathbf{x}_i)} - y_{(\mathbf{x}_i)}^{exact}\right)^2} \tag{3.2}$$

where the coordinates $\mathbf{x}_i$ define an equally spaced Cartesian mesh which covers the entire design space. For the low-fidelity evaluation, the following various function models are defined:

$$\begin{aligned}
f_{Multi(\mathbf{x})}^{low} &= Cosine_{(\mathbf{x})} \times 0.1 \\
f_{Shift(\mathbf{x})}^{low} &= Cosine_{(\mathbf{x})} - 1.0 \\
f_{Xshift(\mathbf{x})}^{low} &= Cosine_{(\mathbf{x}+0.1)} \\
f_{Rndm(\mathbf{x})}^{low} &= Cosine_{(\mathbf{x})} + 0.1\left(ran - 0.5\right) \qquad \left(0 \le ran \le 1\right) \\
f_{Lin\ (\mathbf{x})}^{low} &= Cosine_{(\mathbf{x})} + 0.1\left(x_1 + x_2\right)
\end{aligned} \tag{3.3}$$

In Fig.2 the RMSE values are compared between the SF and VF Kriging approaches. For the VF cases, additional 50 low-fidelity sample points of $f_{Multi}^{low}$ function are also used to construct a Kriging model. The gradient and gradient/Hessian information are also used in the cases of *_FG* and *_FGH*, respectively. These derivatives are analytically calculated in this study. It is obvious that the model accuracy is increased by including the low-fidelity information as well as derivative information. As readers can guess, the most accurate result was given by the case in which function/gradient/Hessian information was used for both high/low-fidelity sample points. In Fig.3, the approximate functions are visualized for the SF cases of five high-fidelity sample points. Poor function prediction accuracy is achieved with the case of only five high-fidelity function information, while the accuracy is improved remarkably by including derivative information. In Fig.4, the approximate functions are visualized for the VF cases. In this case, function information is only used in both cases, and then the number of low-fidelity sample points is increased from 15 to 50. By including additional low-fidelity samples, the prediction accuracy is improved remarkably although the absolute values of the low-fidelity function are quite inaccurate.

In Fig.5, the RMSEs are compared between the various low-fidelity models. All sample points have function information only, and the number of low-fidelity sample points is fixed to 50. It is seen that all VF approaches are helpful for smaller numbers of high-fidelity samples (note that $n_1 << n_2$ represents the most general situation in practical applications). The VF cases with $f_{Xshift}^{low}$ and $f_{Rndm}^{low}$ perform more poorly than the SF model with larger numbers of high-fidelity samples while the VF cases with $f_{Multi}^{low}$ or $f_{Shift}^{low}$ show better performance. On the other hand, the absolute function values of $f_{Xshift}^{low}$ and $f_{Rndm}^{low}$ are much closer to the high-fidelity function than $f_{Multi}^{low}$ or $f_{Shift}^{low}$. In the developed VF formulation, therefore, the important aspect of the low-fidelity function is not the absolute values, but the trends of the function. Since $f_{Multi}^{low}$ or $f_{Shift}^{low}$ are defined by multiplying or adding a constant value, the trends of high-fidelity function are well preserved in the low-fidelity models.

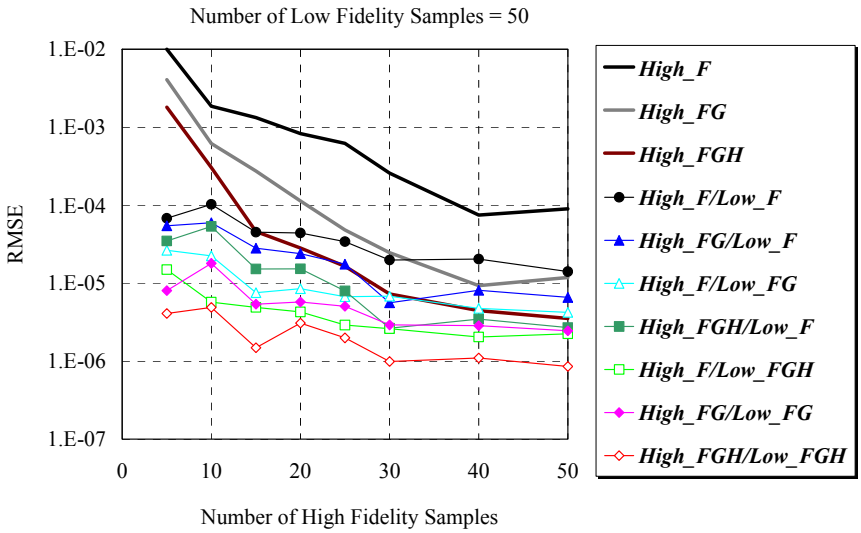


**Fig.2 Comparison of RMSEs between SF and VF Kriging Approaches**

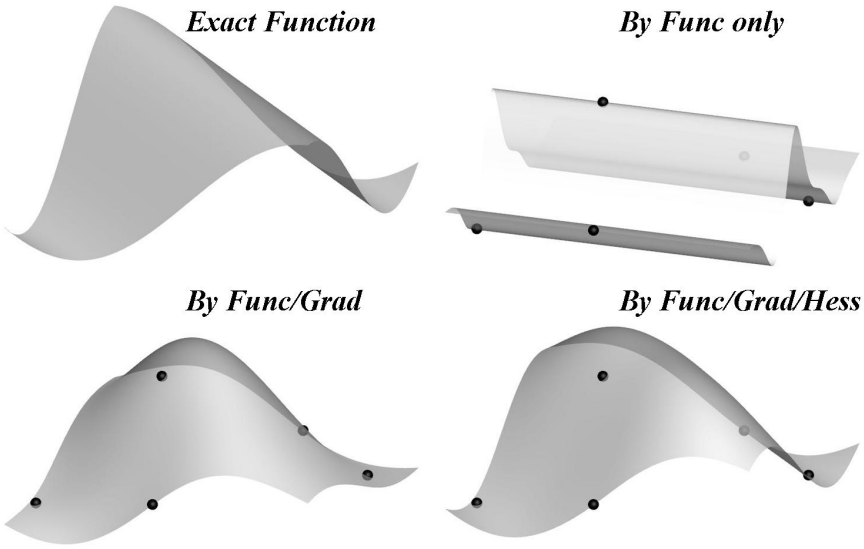


**Fig.3 2D Cosine Function fitted by SF Kriging Model (only 5 high-fidelity samples), Black points: high-fidelity samples**

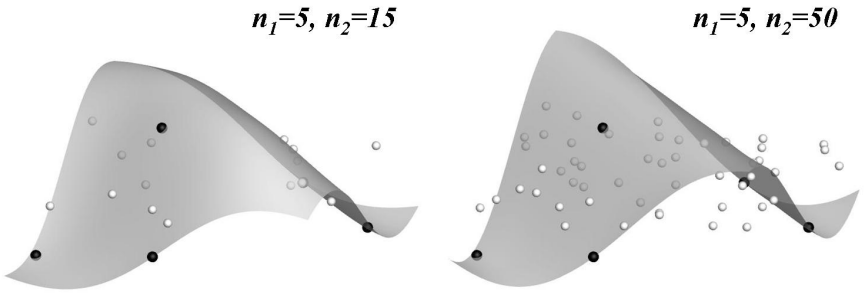


**Fig.4 2D Cosine Function fitted by VF Kriging Model (only function information), Black points: high-fidelity samples, White points: low-fidelity samples**

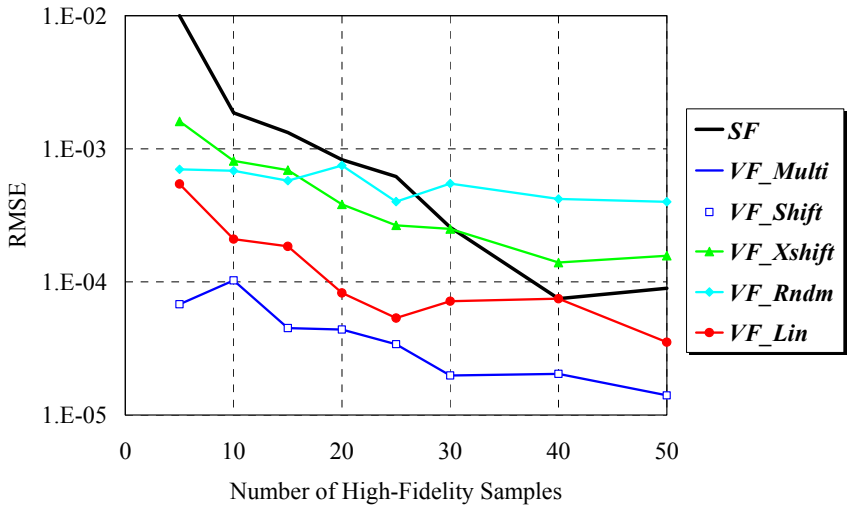


**Fig.5 Comparison of RMSEs between Various Low-Fidelity Models**

## IV. Application to Aerodynamic Data Modeling

### A. 2D NACA0012 Airfoil

In this section the developed VF Kriging approaches are used to model an airfoil two-dimensional aerodynamic database. We consider the steady inviscid flow around a NACA0012 airfoil as a flow example which is described in more detail in Mani and Mavriplis [29,30]. The governing Euler equations of the flow problem are discretized by a finite-volume approach and are solved with second-order spatial accuracy. In this study, two parameters are considered; Mach number and angle of attack. Their ranges are specified as $0.5 \le M_\infty \le 1.5$ and $0^\circ \le \alpha \le 5^\circ$. The exact surface or validation data is obtained through a series of 21x21=441 flow computations covering the entire parameter space at equally spaced increments. Two fidelity levels are defined by the difference of computational mesh resolution. The computational mesh for the high-fidelity function contains approximately 20,000 triangular elements while that for the low-fidelity function contains only 1,700 elements. Their mesh distributions are shown in Fig.6. Taking the computational cost of a high-fidelity function evaluation as one unit, that of a low-fidelity function evaluation is about one over thirtieth.

The exact hypersurfaces of lift and drag coefficients are shown in Fig.7. It can be observed that their behaviors are complicated at the transonic Mach numbers. In Fig.8, estimated hypersurfaces by Kriging models are shown. In this figure, the number of low-fidelity samples is increased from 0 to 50 while that of high-fidelity samples is fixed to 5. These sample points are chosen by LHS. It can be understood that the low-fidelity information is helpful to construct accurate surrogate models for these practical non-linear aerodynamic functional outputs. The accuracy of the surrogate models is evaluated by using the mean error (ME) between the exact and estimated function values. The ME is given as follows:

$$ME = \frac{1}{M}\sum_{i=1}^{M}\left|\hat{y}_{(\mathbf{x}_i)} - y_{(\mathbf{x}_i)}^{exact}\right| \tag{4.1}$$

In Fig.9 the ME values are compared between the SF and VF Kriging approaches. Again, the accuracy improvement at smaller numbers of high-fidelity sample points is observed in the VF results.

Uncertainty analysis is also executed at the center location of $M_\infty = 0.8$ and $\alpha = 2.5^\circ$. The results are compared between full Non-Linear Monte-Carlo (NLMC) and IMC simulations. In this analysis, uncertainties are given to the two parameters of $M_\infty$ and $\alpha$ depending on a normal distribution. The mean of the normal distribution is fixed to the center location while the standard deviation of all parameters ($\sigma_D$) is specified by

designers. The number of function calls for a Monte-Carlo simulation is fixed to 1000 in this study, and then the mean/variance of aerodynamic functions are calculated. The function calls are directly solved by non-linear CFD calculations in the NLMC simulation. It is executed for seven different values of $\sigma_D$, that means 7000 CFD function calls are required for this analysis. In the IMC simulation, on the other hand, the function calls are solved by a surrogate model, which can dramatically reduce the computational cost for the uncertainty analysis. In Fig.10, the 1000 points' locations are visualized on exact hypersurfaces for $\sigma_D$ of 0.1. In Figs.11 and 12, the variations in the mean of lift/drag coefficients with respect to the variation of $\sigma_D$ are compared for SF and VF Kriging model cases. Accurate uncertainty analyses are realized by the increase in number of high-fidelity sample points as well as number of low-fidelity sample points. Thus, the VF Kriging model is promising for accurate uncertainty analysis with low computational cost. Since the CFD functional outputs on this design space are relatively noisy, the CFD derivative information is not used to construct surrogate models.

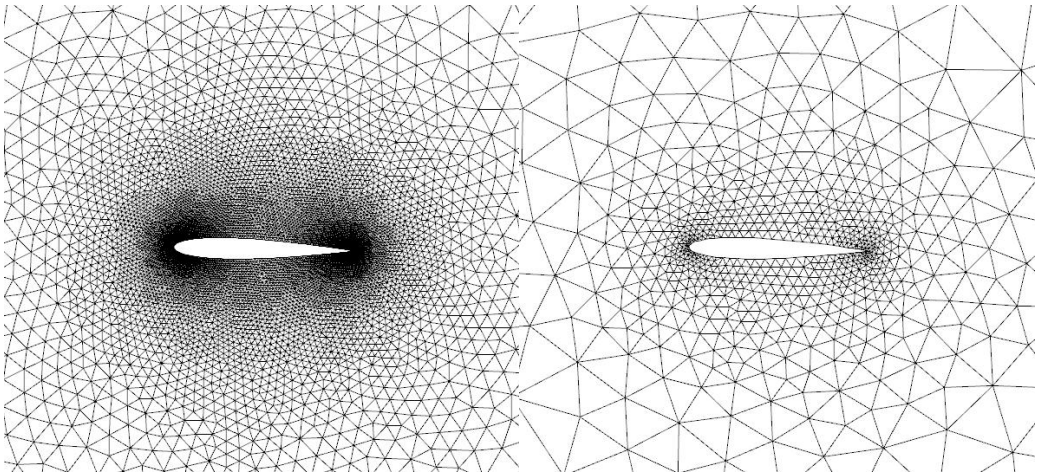

**Fig.6 Computational Meshes for High-fidelity (Left) and Low-fidelity (Right) Evaluations**

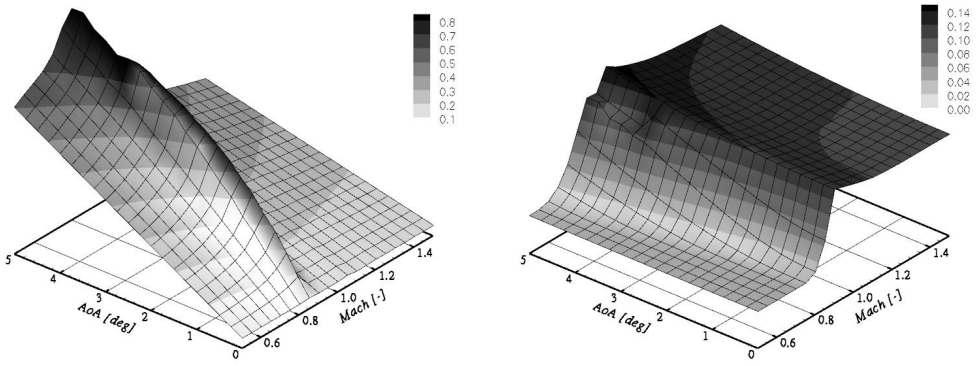


**Fig.7 Exact Hypersurfaces of Lift (Left) and Drag (Right) Coefficients**

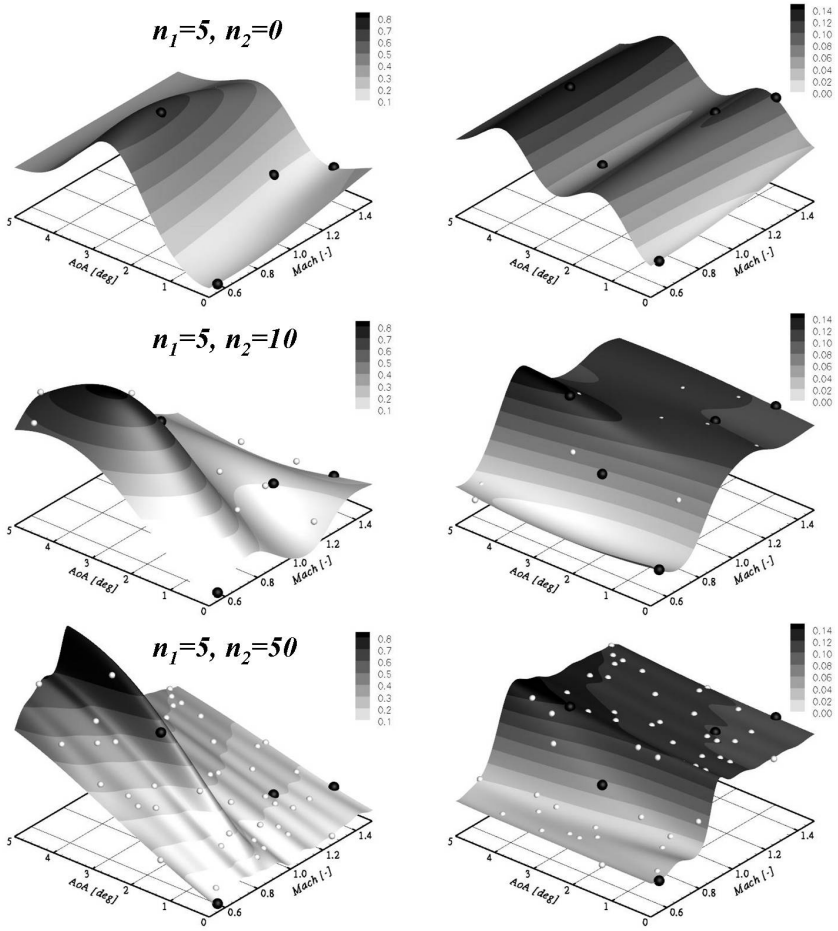


**Fig.8 Estimated Hypersurfaces of Lift (Left) and Drag (Right) Coefficients**
**Black points: high-fidelity samples, White points: low-fidelity samples**

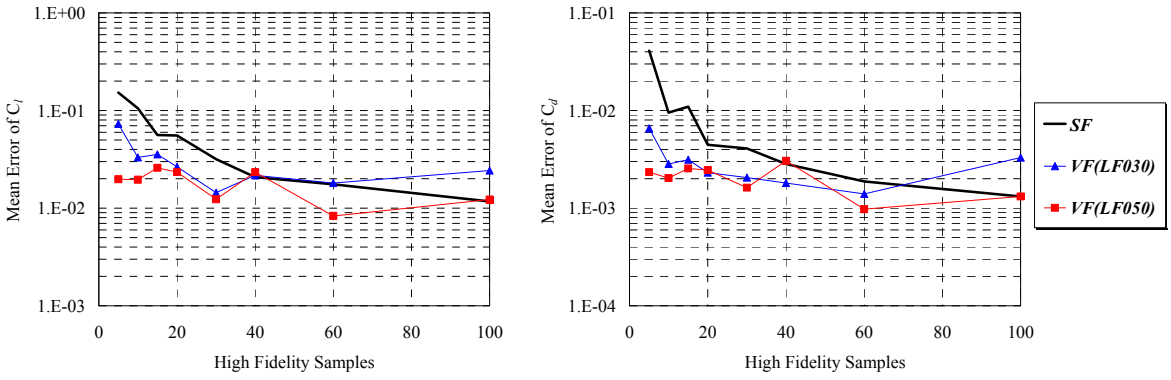


**Fig.9 Comparison of Mean Error of Lift and Drag Coefficients**

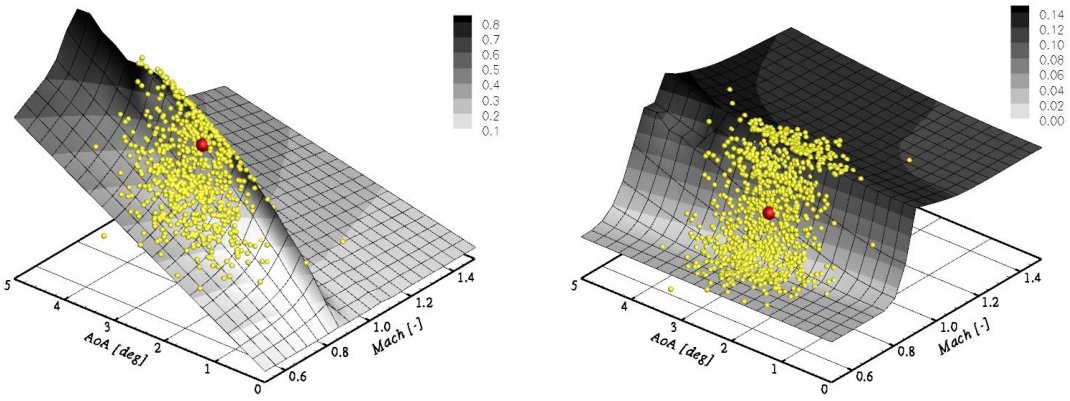


**Fig.10 Locations of 1000 Points for a Monte-Carlo Simulation ($\sigma_D = 0.1$), Left: on Lift, Right: on Drag Hypersurfaces**

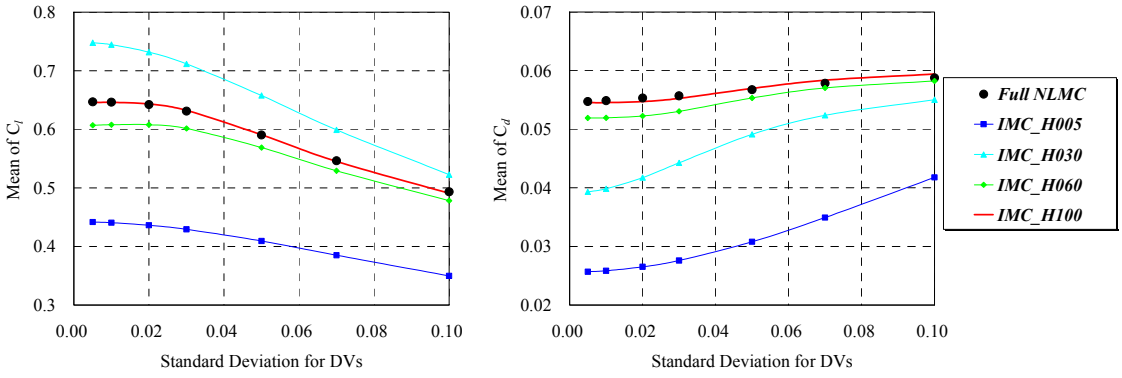


**Fig.11 Uncertainty Analysis by SF Kriging Models ($5 \le n_1 \le 100$, $n_2 = 0$), Left: Mean of Lift, Right: that of Drag**

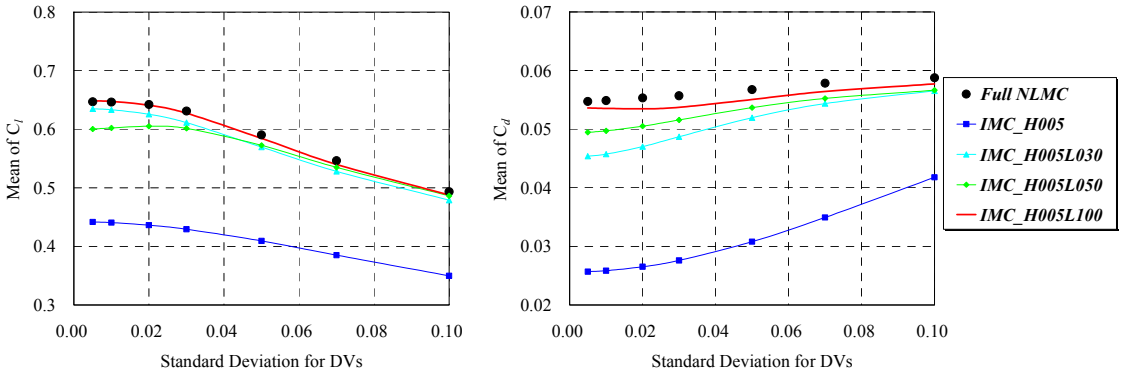


**Fig.12 Uncertainty Analysis by VF Kriging Models ($n_1 = 5$, $0 \le n_2 \le 100$), Left: Mean of Lift, Right: that of Drag**

**B. DPW-IV Case**

In this section, the VF Kriging approach is used to model aerodynamic coefficients of the common research model (CRM) aircraft configuration, which was one of the major subjects of the fourth AIAA drag prediction workshop (DPW) held in 2009 [31]. The CRM configuration is composed by wing, body and horizontal tail as shown in Fig.13. In this study, RANS computational results are used as high-fidelity information while inviscid Euler results are used as low-fidelity information. The number of computational mesh points for RANS computations is about 3.7 millions (tetrahedral/prism/pyramid elements) [32] and the standard Spalart-Allmaras model [33] is utilized. On the other hand, the number of mesh points for Euler computations is about 1.9 millions (tetrahedral elements). These flow computations are performed via the NSU3D unstructured mesh multigrid solver [34]. The computational cost of a RANS computation is roughly 50 times larger than that of an inviscid computation.

Again, Mach number and angle of attack are considered as two input parameters. Their ranges are specified as $0.7 \le M_\infty \le 0.87$ and $0^\circ \le \alpha \le 4^\circ$. The high/low-fidelity hypersurfaces are obtained through a series of 7x7=49 flow computations covering the entire parameter space. These hypersurfaces are shown in Fig.14. In this study, the number of high-fidelity samples is fixed to five points and these are set on the center and four corner locations. Then, all low-fidelity sample points are utilized for the VF case. The estimated hypersurfaces by SF/VF Kriging models are shown in Fig.15. Since the shapes of exact hypersurfaces are simple, the differences between SF and VF models are relatively small. In Fig.16, estimated drag polar curves at $M_\infty = 0.85$ are shown. The VF model gives better agreement with the RANS results although there is no high-fidelity sample point at this Mach number.

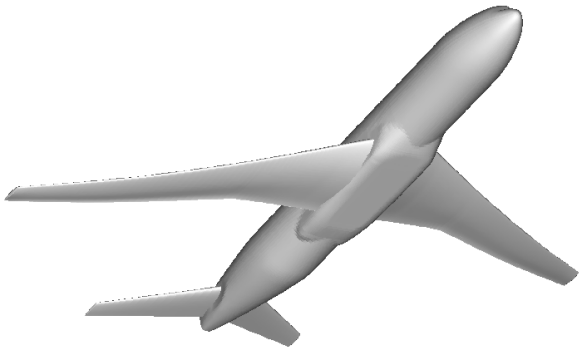

**Fig.13 Common Research Model Aircraft Configuration**

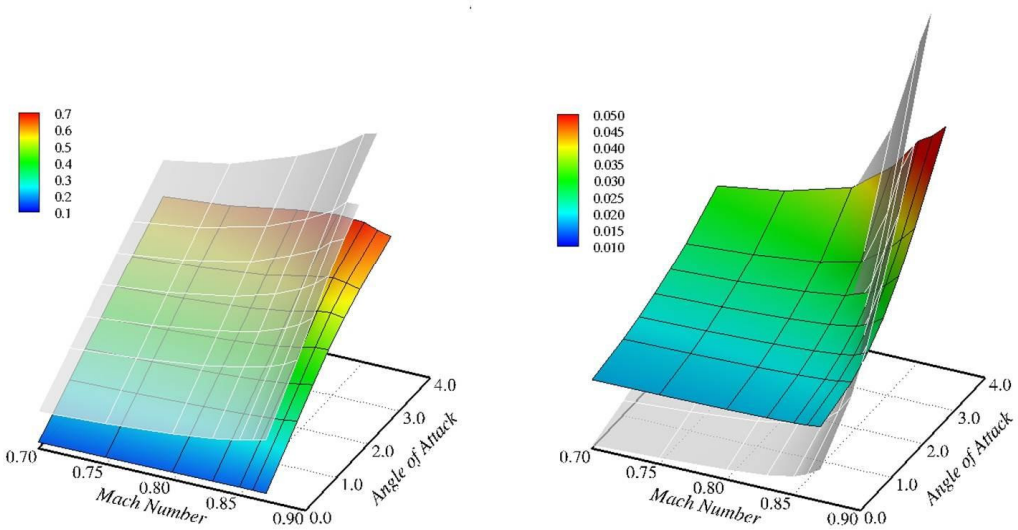


**Fig.14 Hypersurfaces by RANS (Colored) and Inviscid (Monochrome) Computations, Left: Lift, Right: Drag Coefficient**

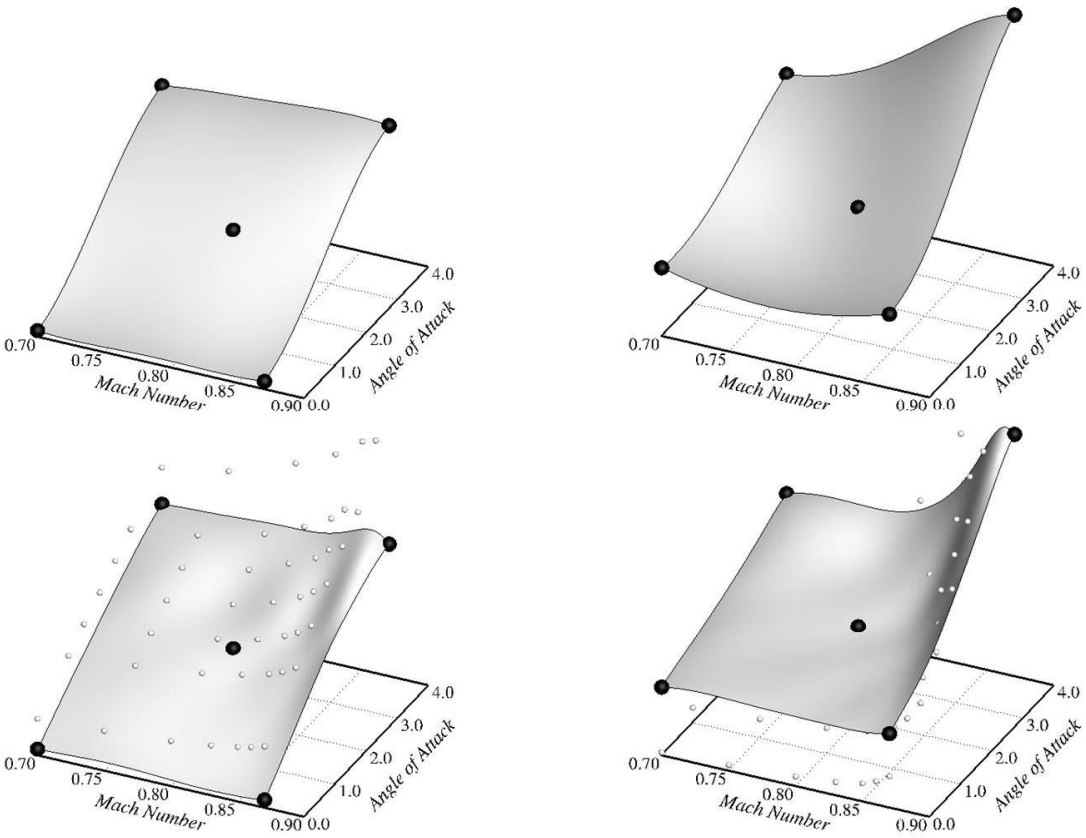


**Fig.15 Estimated Hypersurfaces by SF (Upper) and VF (Lower) Kriging Models, Left: Lift, Right: Drag Coefficient**

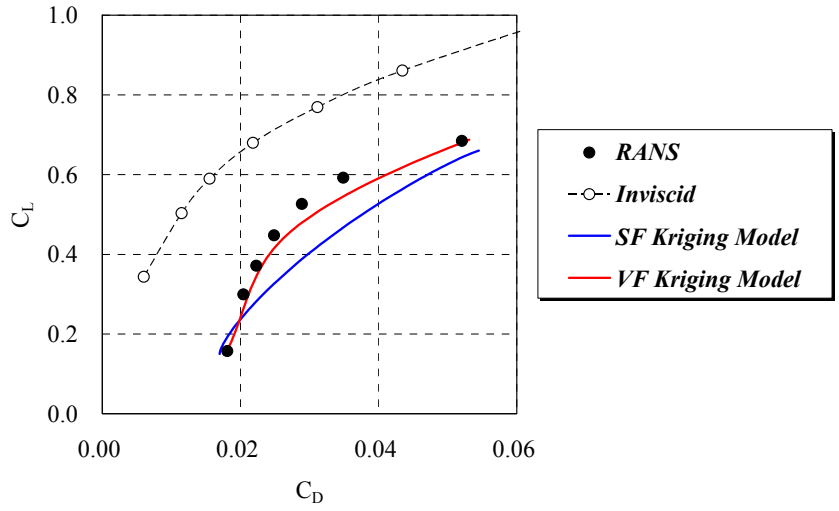


**Fig.16 Drag Polar of CRM Configuration @ M=0.85**

## V. Application to Drag Minimization Problem of 2D Airfoil

In this section, the developed derivative-enhanced VF Kriging approaches are used for a 2D airfoil shape optimization.

### A. Problem Definition

We consider again the steady inviscid flow around a NACA0012 airfoil. The PARSEC parameterization method [35] is used to represent airfoil shapes. As shown in Fig.17, the geometry is described in this method by 11 characteristic parameters of airfoil, such as leading edge radius, positions/curvatures of upper/lower crest and so on. Since the vertical location and thickness of trailing edge are fixed to zero in this study, the number of design variables is nine. The free-stream Mach number is $M_\infty$=0.755 with $\alpha$=1.25 degrees. Two fidelity levels are again defined by the difference of computational mesh resolution, as explained in the previous section (Fig.6). The required deformation and movement of the computational meshes are performed via a linear tension spring analogy [36]. This flow condition gives $C_l$=0.268 and $C_d$=0.00521 for the NACA0012 airfoil by the high-fidelity evaluation while $C_l$=0.262 and $C_d$=0.00522 by the low-fidelity evaluation. Taking the computational cost of a high-fidelity function evaluation as one unit, that of a low-fidelity function evaluation is about one over tenth in this flow condition.

The objective function of this study is defined as follows:

$$F = \frac{1}{2}\left(C_l - C_l^*\right)^2 + \frac{100}{2}\left(C_d - C_d^*\right)^2 \tag{5.1}$$

where a star denotes a target lift or drag coefficient and the factor of one hundred is introduced since the drag coefficient is about an order of magnitude smaller than the lift coefficient in this case. The target lift and drag coefficients are respectively set to 0.675 and 0.0, that is, we attempt a lift constrained drag minimization problem. In addition, a quadratic penalty term is added to the objective function if the sectional area of the new design is less than 90% of the NACA0012 airfoil. The derivatives of this objective function can be evaluated efficiently by using a discrete adjoint approach [29,30]. The computational cost for the adjoint gradient evaluation is almost comparable with that of a CFD calculation.

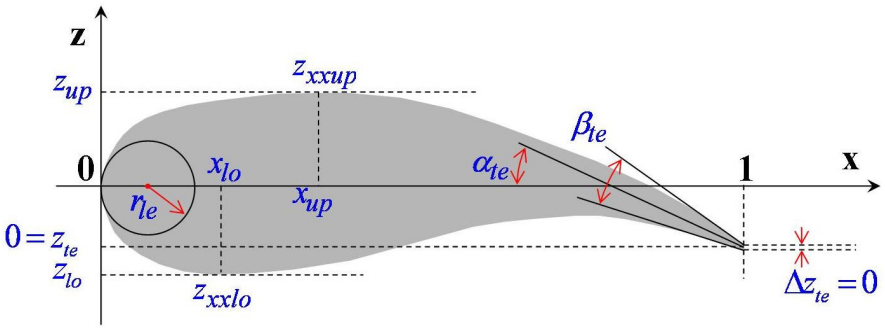


**Fig.17 Airfoil Shape Parameterization by PARSEC**

### B. Optimization Results

For this design optimization problem, 16 and 128 initial sample points have been respectively chosen by the LHS method for the high- and low-fidelity function evaluations. The adjoint gradient has not been evaluated for the initial samples. When constructing a surrogate model, new sample points are added which are chosen by maximizing the EI value. For the cases utilizing derivative-enhanced surrogate models, the adjoint gradient evaluations are invoked only when a new design has better performance than all previous designs. The flowchart of optimization process is shown in Fig.18.

In Table.2, the strategies of design optimization cases are summarized, and the optimization histories are shown in Fig.19. The performance of optimal airfoil given in the case of *SF_Low* is also analyzed by the high-fidelity

evaluation, and then its objective function value is also included in Fig.19 as *HF Eval for Opt_SF_Low*. It can be seen that the high-fidelity evaluation gives the worst performance among all optimal designs. It can be also seen that two derivative-enhanced cases find better designs than function-based Kriging cases. For more detailed analysis, another optimization histories are shown in Fig.20, in which x-axis is changed to a computational cost factor (CCF). It is defined as follows by considering the computational costs for high/low-fidelity function and high-fidelity adjoint gradient evaluations:

$$CCF = 2 \times n_1^{FG} + 1 \times n_1^{F} + 0.1 \times n_2^{F} \tag{5.2}$$

Faster convergence towards the global optimal direction, i.e. the improvement of search efficiency, is confirmed in the derivative-enhanced and/or VF surrogate model approaches by comparing CCF. The pressure distributions around initial and optimal airfoils obtained by *SF_High* and *VF_Grad* are shown in Fig.21. The optimal airfoil shape of *SF_High* is a kind of supercritical airfoils achieving the reduction of shock wave on its upper surface. Although the optimal airfoil shape of *VF_Grad* is not a kind of supercritical airfoils, its performance overcomes that of *SF_High* due to more isotropic compression on the upper surface.

**Table.2 Strategies of Design Optimization Cases**

| Case | Initial Samples | Function Evaluation for Additional Samples | Adjoint Gradient |
|---|---|---|---|
| SF_High | 16 high-fid. | high-fidelity | - |
| SF_High_Grad | 16 high-fid. | high-fidelity | invoke for new optimals |
| SF_Low | 128 low-fid. | low-fidelity | - |
| VF | 128 low-fid. | high-fidelity | - |
| VF_Grad | 128 low-fid. | high-fidelity | invoke for new optimals |

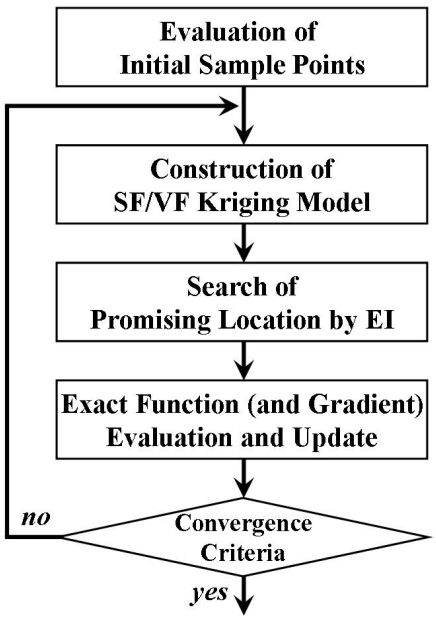


**Fig.18 Flowchart of Optimization Process**

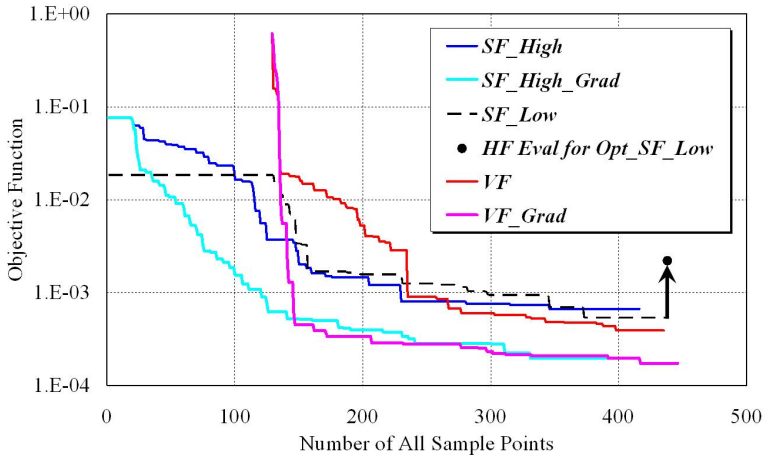


**Fig.19 Optimization History by Number of All Sample Points**

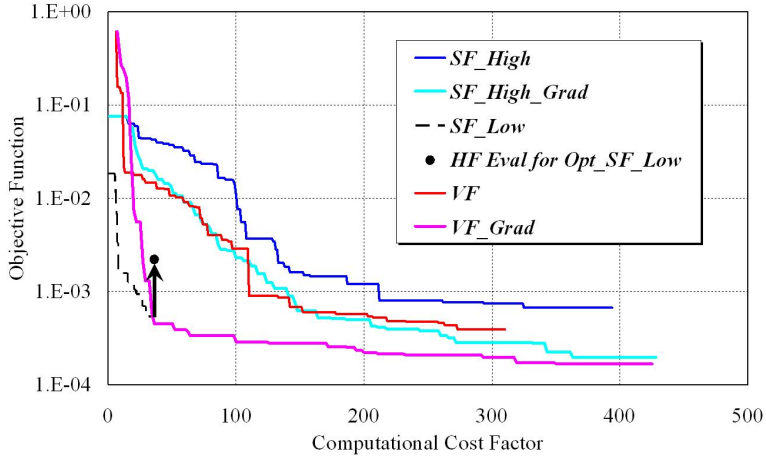


**Fig.20 Optimization History by Computational Cost Factor**

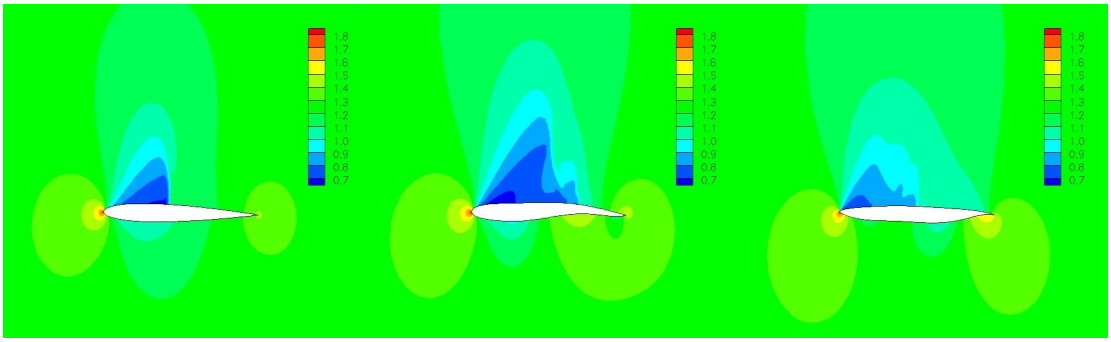

**Fig.21 Pressure Distributions around Airfoils,**
**From L to R: NACA0012, Optimal in *SF_High*, Optimal in *VF_Grad***

## VI. Conclusion

In this paper, derivative-enhanced variable fidelity surrogate models have been developed based on a direct Kriging formulation. In the developed VF approach, the trends of low-fidelity function values as well as a small number of high-fidelity function's absolute values are simultaneously utilized to construct an accurate surrogate model. The derivative information of arbitrary fidelity function is also used to construct a more accurate Kriging model. Firstly, analytical function fitting problems were investigated. The accuracy of surrogate models was sufficiently increased by adding derivative information and/or low-fidelity information in the Kriging formulation.

Modeling of aerodynamic data and drag minimization problems were also investigated. In these problems, different fidelity levels were defined by coarser/finer computational meshes or RANS/inviscid physical models. It was confirmed that the VF approach was promising for estimating high-fidelity aerodynamic functions accurately. Efficient uncertainty analysis based on the surrogate model, which is referred to as inexpensive Monte-Carlo simulation, was also illustrated. The drag minimization problem was solved by combining the developed surrogate models with an efficient adjoint CFD gradient evaluation method. Faster reduction of the objective function value in the derivative-enhanced VF Kriging model approach was demonstrated, and this is expected to be more beneficial in higher dimensions and/or for more complex problems.

Thus, the developed derivative-enhanced VF Kriging approach is promising for efficient design optimization, aerodynamic database construction and uncertainty analysis problems. The developed approach can improve design efficiency dramatically since the adjoint derivatives and/or low-fidelity function values can be calculated with low computational cost.

## Acknowledgments

We are very grateful to Dr. Karthik Mani and Dr. Markus P. Rumpfkeil for making their flow and adjoint solver available to us.

## References

[1]Peter, J., and Marcelet, M., "Comparison of Surrogate Models for Turbomachinery Design," *WSEAS Transactions on Fluid Mechanics*, Issue.1, Vol.3, 2008.

[2]Cressie, N., "The Origins of Kriging," *Mathematical Geology*, Vol.22, No.3, pp.239-252, 1990.

[3]Koehler, J. R., and Owen, A. B., "Computer Experiments," In Ghosh, S., Rao, C.R., (Eds.), *Handbook of Statistics*, pp.261-308, 1996.

[4]Jones, D. R., Schonlau, M., and Welch, W. J., "Efficient Global Optimization of Expensive Black-Box Functions," *Journal of Global Optimization*, Vol.13, pp.455–492, 1998.

[5]Simpson, T. W., Korte, J. J., Mauery, T. M., and Mistree, F., "Comparison of Response Surface and Kriging Models for Multidisciplinary Design Optimization," AIAA-98-4755, 1998.

[6]Chung, H. S., and Alonso, J. J., "Using Gradients to Construct Cokriging Approximation Models for High-Dimensional Design Optimization Problems," AIAA-2002–0317, 2002.

[7]Martin, J. D., and Simpson, T. W., "Use of Kriging Models to Approximate Deterministic Computer Models," *AIAA Journal*, Vol.43, No.4, pp.853-863, 2005.

[8]Jeong, S., Murayama, M., and Yamamoto, K., "Efficient Optimization Design Method Using Kriging Model," *Journal of Aircraft*, Vol.42, No.2, pp.413-420, 2005.

[9]Laurenceau, J., and Sagaut, P., "Building Efficient Response Surfaces of Aerodynamic Functions with Kriging and Cokriging," *AIAA Journal*, Vol.46, No.2, pp.498-507, 2008.

[10]Laurenceau, J., and Meaux, M., "Comparison of Gradient and Response Surface Based Optimization Frameworks Using Adjoint Method," AIAA-2008-1889, 2008.

[11]Han, Z. H., Görtz, S., and Zimmermann, R., "On Improving Efficiency and Accuracy of Variable-Fidelity Surrogate Modeling in Aero-data for Loads Context," CEAS 2009 European Air and Space Conference, 2009.

[12]Han, Z. H., Zimmermann, R., and Görtz, S., "A New Cokriging Method for Variable-Fidelity Surrogate Modeling of Aerodynamic Data," AIAA-2010-1225, 2010.

[13]Yamazaki, W., Rumpfkeil, M. P., and Mavriplis, D. J., "Design Optimization Utilizing Gradient/Hessian Enhanced Surrogate Model," AIAA-2010-4363, 2010.

[14]Yamazaki, W., Mouton, S., and Carrier, G., "Geometry Parameterization and Computational Mesh Deformation by Physics-Based Direct Manipulation Approaches," *AIAA Journal*, Vol.48, No.8, pp.1817-1832, 2010.

[15]Jameson, A., "Efficient Aerodynamic Shape Optimization," AIAA Paper 2004-4369, 2004.

[16]Mavriplis, D. J., "A Discrete Adjoint-Based Approach for Optimization Problems on Three-Dimensional Unstructured Meshes," AIAA Paper 2006-0050, 2006.

[17]Kim, H. J., Sasaki, D., Obayashi, S., and Nakahashi, K., "Aerodynamic Optimization of Supersonic Transport Wing Using Unstructured Adjoint Method," *AIAA Journal*, Vol.39, No.6, pp.1011-1020, 2001.

[18]Nielsen, E. J. and Anderson, W. K., "Recent Improvements in Aerodynamic Optimization on Unstructured Meshes," *AIAA Journal*, Vol.40, No.6, pp.1155–1163, 2002.
[19]Giles, M. B., Duta, M. C., Muller, J. D., and Pierce, N. A., "Algorithm Developments for Discrete Adjoint Methods," *AIAA Journal*, Vol.41, No.2, pp.198-205, 2003.
[20]Nielsen, E. J., and Park, M. A., "Using an Adjoint Approach to Eliminate Mesh Sensitivities in Computational Design," *AIAA Journal*, Vol.44, No.5, pp.948-953, 2006.
[21]Peter, J., and Mayeur, M., "Improving Accuracy and Robustness of a Discrete Direct Differentiation Method and Discrete Adjoint Method for Aerodynamic Shape Optimization," Proceedings of the ECCOMAS CFD Conference, 2006.
[22]Rumpfkeil, M. P., and Mavriplis, D. J., "Efficient Hessian Calculations using Automatic Differentiation and the Adjoint Method," AIAA-2010-1268, 2010.
[23]Rumpfkeil, M. P., Yamazaki, W., and Mavriplis, D. J., "Uncertainty Analysis Utilizing Gradient and Hessian Information," 6th International Conference on Computational Fluid Dynamics (ICCFD6), St. Petersburg, Russia, 2010.
[24]Rumpfkeil, M. P., Yamazaki, W., and Mavriplis, D. J., "A Dynamic Sampling Method for Kriging and Cokriging Surrogate Models," 49th AIAA Aerospace Sciences Meeting, 2011 (submitted for publication).
[25]Hascoët, L., "TAPENADE: A Tool for Automatic Differentiation of Programs," Proceedings of the ECCOMAS Conference, Jyvaskyla, Finland, 2004.
[26]Alexandrov, N. M., Lewis, R. M., Gumbert, C. R. Green, L. L., and Newman, P. A., "Approximation and Model Management in Aerodynamic Optimization with Variable-Fidelity Models," *Journal of Aircraft*, Vol.38, No.6, pp.1093-1101, 2001.
[27]Fonseca, C. M., and Fleming, P. J., "Genetic Algorithms for Multiobjective Optimization: Formulation, Discussion and Generalization," Proceedings of the 5th International Conference on Genetic Algorithms, Morgan Kaufmann Publishers, Inc., San Mateo, pp.416-423, 1993.
[28]Kim, H. J., and Liou, M.-S., "New Multi-Objective Genetic Algorithms for Diversity and Convergence Enhancement," AIAA Paper 2009-1168, 2009.
[29]Mani, K., and Mavriplis, D. J., "An Unsteady Discrete Adjoint Formulation for Two-Dimensional Flow Problems with Deforming Meshes," AIAA-2007-60, 2007.
[30]Mani, K., and Mavriplis, D. J., "Unsteady Discrete Adjoint Formulation for Two-Dimensional Flow Problems with Deforming Meshes," *AIAA Journal*, Vol.46, No.6, pp.1351-1364, 2008.
[31]Vassberg, J. C. et al, "Summary of the Fourth AIAA CFD Drag Prediction Workshop," AIAA-2010-4547, 2010.
[32]Mavriplis, D. J., and Long, M., "NSU3D Results for the Fourth AIAA Drag Prediction Workshop," AIAA-2010-4550, 2010.
[33]Spalart, P. R., and Allmaras, S. R., "A One-equation Turbulence Model for Aerodynamic Flows," *La Recherche Aerospatiale*, Vol.1, pp.5-21, 1994.
[34]Mavriplis, D. J., "Aerodynamic Drag Prediction Using Unstructured Mesh Solvers," *CFD-Based Drag Prediction and Reduction, eds. H. Deconoinck, K. Sermus and C. van Dam*, VKI Lecture Series 2003-02, von Karman Institute for Fluid Dynamics, 2003.
[35]Sobieczky, H., "Parametric Airfoils and Wings," *Notes on Numerical Fluid Mechanics*, Vol.68, pp.71-88, 1998.
[36]Batina, J. T., "Unsteady Euler Airfoil Solutions Using Unstructured Dynamic Meshes," *AIAA Journal*, Vol.28, No.8, pp.1381-1388, 1990.