AIAA JOURNAL
Vol. 50, No. 9, September 2012

Hierarchical Kriging Model for Variable-Fidelity Surrogate Modeling

Zhong-Hua Han*
National Key Laboratory of Science and Technology on Aerodynamic Design and Research, Northwestern Polytechnical University, 710072 Xi'an, People's Republic of China
and
Stefan Görtz†
DLR, German Aerospace Center, D-38108 Braunschweig, Germany

DOI: 10.2514/1.J051354

The efficiency of building a surrogate model for the output of a computer code can be dramatically improved via variable-fidelity surrogate modeling techniques. In this article, a hierarchical kriging model is proposed and used for variable-fidelity surrogate modeling problems. Here, hierarchical kriging refers to a surrogate model of a high-fidelity function that uses a kriging model of a sampled lower-fidelity function as a model trend. As a consequence, the variation in the lower-fidelity data is mapped to the high-fidelity data, and a more accurate surrogate model for the high-fidelity function is obtained. A self-contained derivation of the hierarchical kriging model is presented. The proposed method is demonstrated with an analytical example and used for modeling the aerodynamic data of an RAE 2822 airfoil and an industrial transport aircraft configuration. The numerical examples show that it is efficient, accurate, and robust. It is also observed that hierarchical kriging provides a more reasonable mean-squared-error estimation than traditional cokriging. It can be applied to the efficient aerodynamic analysis and shape optimization of aircraft or any other research areas where computer codes of varying fidelity are in use.

Nomenclature

c_l, c_d, c_m or C_L, C_D, C_M	=	lift, drag, and pitching moment coefficients
$\mathbf{F}$	=	regression matrix for kriging predictor
L	=	likelihood function
Ma	=	Mach number
m	=	number of dimensions
n, n_{lf}	=	number of high- and low-fidelity sample points
p	=	parameter for Gaussian correlation function
$\mathbf{R}$	=	correlation matrix of kriging model
R, R_{lf}	=	spatial correlation functions
Re	=	Reynolds number
$\mathbf{r}$	=	correlation vector of kriging model
$\mathbf{S}$, $\mathbf{S}_{\text{lf}}$	=	sampling sites for high- and low-fidelity functions
$\mathbf{V}_{\text{HK}}$	=	hierarchical kriging predictor vector
$\mathbf{w}$	=	kriging weights
$\mathbf{x}$, $\mathbf{x}'$	=	independent variables
$\mathbf{y}_S$, $\mathbf{y}_{S,\text{lf}}$	=	response values
Y, Y_{lf}	=	random functions
$Z(\cdot)$, $Z_{\text{lf}}(\cdot)$	=	Gaussian random processes
α	=	angle of attack
β_0, $\beta_{0,\text{lf}}$	=	coefficients of trend model for kriging predictor
δ	=	deflection angle of horizontal tail
$\boldsymbol{\theta}$	=	hyper-parameter vector for spatial correlation function
μ	=	Lagrange multiplier
ξ	=	weighted distance for spatial correlation function
σ^2	=	process variance

Subscripts

init	=	initial value
lf	=	low-fidelity value
k	=	index $\in [1, m]$
S	=	sampled data

Superscripts

(i)	=	index $\in [1, n]$, referring to ith sample point
$'$	=	new point
$\wedge$	=	approximated value
$\sim$	=	redefined value

I. Introduction

SURROGATE modeling plays an increasingly important role in different areas of aerospace science and engineering, such as aerodynamic data production, aerodynamic shape optimization, structural design, multidisciplinary optimization (MDO), and design of aircraft or spacecraft [1,2]. Surrogate modeling can be used to greatly improve the design efficiency when high-fidelity but time-consuming numerical models (e.g., computational fluid dynamics, or CFD, models and finite-element, or FE, structural models) are employed. In addition, it can be very helpful in finding global optima, filtering numerical noise, realizing parallel design optimization, and integrating simulation codes of different disciplines into a process chain for engineering design. Here, the term "surrogate model" has the same meaning as "response surface model," "metamodel," "approximation model," "emulator," etc. Among the different surrogate modeling techniques, such as quadratic polynomial models or radial basis functions, the kriging method [3,4] has received growing popularity due to its good capability of predicting multidimensional and highly nonlinear responses based on sampled data while providing a useful error estimation (indicating the uncertainty of the prediction). Despite the popularity of kriging method, further improvements of its efficiency and accuracy are still required before it can be widely used for realistic engineering design problems such as aircraft design. Taking the aerodynamic data for loads production

Received 11 May 2011; revision received 29 February 2012; accepted for publication 29 February 2012.

*Associate Professor, School of Aeronautics; hanzh@nwpu.edu.cn. Member AIAA.
†Team Lead, Aero-Loads Prediction Group, Institute of Aerodynamics and Flow Technology; Stefan.Goertz@dlr.de. Member AIAA.

of an aircraft as an example, building an accurate or, at least, usable kriging based on sampled aerodynamic data obtained with a Navier–Stokes (NS) CFD code is very demanding, due to the fact that a large number of simulations is required. This situation may become even worse with increasing problem size [5–7] (increasing number of independent variables and increasing range of these variables).

Variable-fidelity modeling (VFM) has proved effective for further improving the efficiency of building a surrogate model. It uses a set of numerical methods with varying degrees of fidelity and computational expense (e.g., potential theory, Euler equations, and NS equations) to approximate the unknown function of interest as a function of the independent variables while reducing the number of the expensive high-fidelity computations. VFM as discussed in this article has the same meaning as "multifidelity modeling," "variable-complexity modeling," "data fusion," or "data merging".

The most popular method currently used for VFM is a correction-based method. The correction is called "bridge function" or "scaling function". The correction can be multiplicative, additive, or hybrid multiplicative/additive. The multiplicative bridge function was used for variable-fidelity optimization [8] since 1993. It is used to "locally" scale the low-fidelity function to approximate the high-fidelity function. When it is combined with the trust region method [9–11], the optimization cost can be greatly reduced. The additive bridge function was then developed as a "global" correction and has become one of the most popular methods for variable-fidelity optimization [12,13] or for data fusion [14,15]. In most of the cases, it outperforms a multiplicative bridge function as it is generally more accurate and robust. To represent more complicated correlation between low- and high-fidelity functions, a hybrid multiplicative/additive bridge function method was proposed in [16,17] for variable-fidelity optimization. A method of taking the low-fidelity function as an additional independent variable of the surrogate modeling problem was developed in [18], which can be viewed as one of the hybrid multiplicative/additive bridge function methods. More recently, a more general method called generalized hybrid bridge function was proposed in [19] and applied in the context of aerodata for loads prediction of aircraft. In [19], this generalize hybrid bridge function was combined with gradient-enhanced kriging (GEK), with the gradients computed by adjoint method [20]).

Another type of VFM method is the so-called cokriging method. Because the underlying theory of cokriging is that of two-variable or multivariable kriging, one can consider it as a general extension of (one-variable) kriging to a modeling method that is assisted by auxiliary variables or secondary information. Cokriging was proposed in the geostatistics community [21] and is widely used as a space interpolation method in mining engineer and other research areas. It was extended to applications with computer experiments by Kennedy and O'Hagan [22] (called KOH autoregressive model) in 2000. KOH's autoregressive model can be viewed as a version of cokriging that takes the output of a lower-fidelity computer code as an auxiliary variable to enhance the prediction of the output of a high-fidelity computer code. The building of the covariance matrix used in KOH's autoregressive model is a multistage process (see [23–25] for the details of the implementation). First, a kriging for the low-fidelity data is built. Then, another kriging for the difference between the outputs of the high- and low-fidelity methods is built with a multiplicative scaling factor being tuned. Finally, the covariance matrix is constructed and the cokriging is fitted. More recently, an alternative cokriging method was proposed in [26], which is an extension of the geostatistical cokriging method to variable-fidelity surrogate modeling of the output of computer codes. Different from the multistage process of KOH's method, the construction of the covariance matrix and the tuning of the cokriging is a one-step procedure in this new method; just put all the observed data (both low- and high-fidelity) in one column vector and form an augmented correlation matrix, and then tune the model. The authors stated that the implementation of this method (see [26]) is as simple as conventional kriging, but with slightly increased computational effort (a similar method was proposed in [27]). The incorporation of gradient information in the cokriging was also discussed and demonstrated in [26].

Generally, the correction-based VFM approaches are simple and robust, but they are less accurate than cokriging-based VFM methods and do not provide an estimate of the mean-squared error (MSE) of the predicted value. The cokriging-based VFM approaches, on the other hand, are more accurate than the correction-based VFM methods and do provide an MSE estimate, but the model itself as well as its implementation is more complicated and less robust. In addition, the MSE estimation of the existing VFM methods is not suitable for realistic applications (e.g., for error-based sample point refinement or for estimating the expected improvement (EI) [28] in the context of variable-fidelity optimization).

This article is motivated by the aspiration to develop a method that is as simple (and robust) as correction-based VFM methods and as accurate as cokriging-based VFM methods. Furthermore, it is motivated by the need to improve the MSE estimation of the existing VFM methods for sample point refinement and EI-based efficient global optimization (EGO) algorithms. This method is called hierarchical kriging in this article. The article is organized as follows. Section II gives the formulation of the hierarchical kriging method, including the hierarchical kriging predictor and its MSE, the choice of the correlation models and model fitting. Section III presents some numerical examples, including an analytical example with one independent parameter, aerodynamic examples for an RAE 2822 airfoil with one independent parameter, and an industrial aircraft configuration with three independent variables.

II. Hierarchical Kriging Formulation

Kriging is a statistical interpolation method suggested by Krige [3] in 1951 and mathematically formulated by Matheron [29] in 1963. Kriging gained popularity in design and analysis of deterministic computer experiments after the research work of Sacks et al. [4]. In this article, we refer to hierarchical kriging as an extension of Sacks's kriging to a model in which the prediction of the output of a high-fidelity computer experiment (high-fidelity function) is assisted by the output of the lower-fidelity computer experiments (low-fidelity functions).

There are different versions of kriging such as "simple kriging," "ordinary kriging," and "universal kriging". Among them, universal kriging refers to kriging with a general linear trend model. Usually the model trend is defined by low-order polynomials, typically linear or quadratic. Instead of using polynomials, in this article we use a secondary kriging model as the model trend of the primary kriging model of the function of interest. As the primary kriging model is built hierarchically, we call it hierarchical kriging (HK) accordingly in this article. Note that a similar method called kriging with external drift [30], sometimes as referred to as external drift kriging, has been available in the geostatistics community. The HK method can be regarded as an alternative method to the kriging with external drift method for the surrogate modeling problems in the engineering design community where deterministic computer experiments are carried out. Note that our HK method differs from kriging with external drift mainly due to the following facts.

1) Kriging with external drift assumes that the auxiliary function is sampled at the same locations as the primary function of interest and additionally sampled at all new locations to be interpolated, but there is no such assumption in our HK method.

2) Kriging with external drift uses so-called variograms to model the covariance between random functions at different sample sites, but our HK method uses "correlation functions" with correlation coefficients tuned by maximum likelihood estimation, which makes it well-suited for design and analysis of deterministic computer experiments.

The study on the theoretical link between HK and kriging with external drift is beyond the scope of the article. The readers are referred to [31] for details about the theory and implementation of kriging with external drift.

A. General Descriptions and Assumptions

Here, only two levels (or sets) of data for the same aerodynamic characteristic (e.g., C_L, C_D, or C_M) are considered. The method is

readily extended to three (or more) levels of fidelity (see the Appendix) and to include gradient information (see [19]). In addition, the output of the HK model is assumed to be scalar. This means that we seek HK models for C_L, C_D, and C_M, for example.

For an m-dimensional problem, suppose we are concerned with the prediction of an expensive-to-evaluate (and unknown) high-fidelity function $y\colon \mathbb{R}^m \to \mathbb{R}$ with the assistance of a cheaper-to-evaluate low-fidelity function $y_{\text{lf}}\colon \mathbb{R}^m \to \mathbb{R}$. Assume that the high-fidelity function y is sampled at sites

$$\mathbf{S} = [\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}]^T \in \mathbb{R}^{n\times m} \tag{1}$$

with the corresponding responses

$$\mathbf{y}_S = [y^{(1)}, \ldots, y^{(n)}]^T = [y(\mathbf{x}^{(1)}), \ldots, y(\mathbf{x}^{(n)})]^T \in \mathbb{R}^n \tag{2}$$

Here, n is the number of sampling sites for the high-fidelity computations. The pair $(\mathbf{S}, \mathbf{y}_S)$ denotes the high-fidelity sampled data sets in the vector space. Similarly, assume that the low-fidelity function is sampled at n_{lf} sites $\mathbf{S}_{\text{lf}}$ with corresponding responses $\mathbf{y}_{S,\text{lf}}$. Note that the high-fidelity sampling sites $\mathbf{S}$ do not need to be the subset of the low-fidelity sampling sites $\mathbf{S}_{\text{lf}}$.

With the aforementioned descriptions and assumptions, our objective here is to develop a surrogate model for predicting the output of a high-fidelity code for any untried site $\mathbf{x} \in \mathbb{R}^m$ (that is, to estimate $y(\mathbf{x})$) based on the sampled data sets $(\mathbf{S}, \mathbf{y}_S)$ and $(\mathbf{S}_{\text{lf}}, \mathbf{y}_{S,\text{lf}})$, in an attempt to achieve the desired accuracy with the least possible number of the high-fidelity computations.

B. Kriging of Low-Fidelity Function

To build a surrogate model for the high-fidelity function, we first build a surrogate model for the lower-fidelity function that will be used thereafter to assist the prediction. Assume a random process corresponding to the low-fidelity function y_{lf}:

$$Y_{\text{lf}}(\mathbf{x}) = \beta_{0,\text{lf}} + Z_{\text{lf}}(\mathbf{x}) \tag{3}$$

where $\beta_{0,\text{lf}}$ is an unknown constant, and $Z_{\text{lf}}(\mathbf{x})$ is a stationary random process. Then, we can follow [4] to build a kriging based on the sampled data set $(\mathbf{S}_{\text{lf}}, \mathbf{y}_{S,\text{lf}})$. After the kriging is fitted, the prediction of the low-fidelity function at any untried point $\mathbf{x}$ can be written as

$$\hat{y}_{\text{lf}}(\mathbf{x}) = \beta_{0,\text{lf}} + \mathbf{r}_{\text{lf}}^T(\mathbf{x})\mathbf{R}_{\text{lf}}^{-1}(\mathbf{y}_{S,\text{lf}} - \beta_{0,\text{lf}}\mathbf{1}) \tag{4}$$

where $\beta_{0,\text{lf}} = (\mathbf{1}^T\mathbf{R}_{\text{lf}}^{-1}\mathbf{1})^{-1}\mathbf{1}^T\mathbf{R}_{\text{lf}}^{-1}\mathbf{y}_{S,\text{lf}}$; $\mathbf{R}_{\text{lf}} \in \mathbb{R}^{n_{\text{lf}}\times n_{\text{lf}}}$ is the correlation matrix representing the correlation between the observed points; $\mathbf{1} \in \mathbb{R}^{n_{\text{lf}}}$ is a column vector filled with ones; and $\mathbf{r}_{\text{lf}} \in \mathbb{R}^{n_{\text{lf}}}$ is the correlation vector representing the correlation between the untried point and the observed points. The reader is referred to [32–34] for the details of building such a kriging.

C. Hierarchical Kriging of High-Fidelity Function

Different from conventional kriging, here we assume that the stationary random process corresponding to the high-fidelity function is of the form

$$Y(\mathbf{x}) = \beta_0\hat{y}_{\text{lf}}(\mathbf{x}) + Z(\mathbf{x}) \tag{5}$$

with the approximated low-fidelity function $\hat{y}_{\text{lf}}(\mathbf{x})$ (see Sec. II.B) scaled by an unknown constant factor β_0 serving as the global trend function (or mean) and the stationary random processes $Z(\cdot)$ having zero mean and a covariance of

$$\text{Cov}[Z(\mathbf{x}), Z(\mathbf{x}')] = \sigma^2 R(\mathbf{x}, \mathbf{x}') \tag{6}$$

Here, σ^2 is the process variance of $Z(\cdot)$, and $R(\mathbf{x}, \mathbf{x}')$ is the spatial correlation function, which only depends on the Euclidean distance between two sites, $\mathbf{x}$ and $\mathbf{x}'$. Assuming that the high-fidelity function can be approximated by a linear combination of the observed high-fidelity data $\mathbf{y}_S$, the HK predictor of $y(\mathbf{x})$ at an untried $\mathbf{x}$ is formally defined as

$$\hat{y}(\mathbf{x}) = \mathbf{w}^T\mathbf{y}_S \tag{7}$$

where $\mathbf{w} = [w^{(1)}, \ldots, w^{(n)}]^T$ is a vector of weight coefficients associated with the sampled high-fidelity data. Then, we replace $\mathbf{y}_S = [y^{(1)}, \ldots, y^{(n)}]^T$ with the corresponding random quantities $\mathbf{Y}_S = [Y^{(1)}, \ldots, Y^{(n)}]^T$. We also treat $\hat{y}(\mathbf{x})$ as random and try to minimize its MSE:

$$\text{MSE}[\hat{y}(\mathbf{x})] = E[(\mathbf{w}^T\mathbf{Y}_S - Y(\mathbf{x}))^2] \tag{8}$$

subject to the unbiasedness constraint:

$$E\left[\sum_{i=1}^{n} w^{(i)}Y(\mathbf{x}^{(i)})\right] = E[Y(\mathbf{x})] \tag{9}$$

Solving this constrained minimization problem, $\mathbf{w} = [w^{(1)}, \ldots, w^{(n)}]^T$ in Eq. (7) can be found by solving the following linear equations:

$$\begin{cases} \sum_{j=1}^{n} w^{(j)}\text{Cov}[Z(\mathbf{x}^{(i)}), Z(\mathbf{x}^{(j)})] + \frac{\mu}{2}\hat{y}_{\text{lf}}(\mathbf{x}^{(i)}) \\ \quad = \text{Cov}[Z(\mathbf{x}^{(i)}), Z(\mathbf{x})], \qquad i = 1, \ldots, n \\ \sum_{i=1}^{n} w^{(i)}\hat{y}_{\text{lf}}(\mathbf{x}^{(i)}) = \hat{y}_{\text{lf}}(\mathbf{x}) \end{cases} \tag{10}$$

where μ is the Lagrange multiplier used to change the constrained minimization of the MSE into an unconstrained one, which in turn is necessary to fulfill the unbiasedness condition. Substituting Eq. (6) into Eq. (10) leads to the following linear equations:

$$\begin{cases} \sum_{j=1}^{n} w^{(j)}R(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) + \frac{\mu}{2\sigma^2}\hat{y}_{\text{lf}}(\mathbf{x}^{(i)}) = R(\mathbf{x}^{(i)}, \mathbf{x}), \quad i = 1, \ldots, n \\ \sum_{i=1}^{n} w^{(i)}\hat{y}_{\text{lf}}(\mathbf{x}^{(i)}) = \hat{y}_{\text{lf}}(\mathbf{x}) \end{cases} \tag{11}$$

or in matrix form:

$$\begin{bmatrix} \mathbf{R} & \mathbf{F} \\ \mathbf{F}^T & 0 \end{bmatrix}\begin{bmatrix} \mathbf{w} \\ \tilde{\mu} \end{bmatrix} = \begin{bmatrix} \mathbf{r} \\ \hat{y}_{\text{lf}}(\mathbf{x}) \end{bmatrix} \tag{12}$$

where

$$\begin{aligned} &\mathbf{F} = [\hat{y}_{\text{lf}}(\mathbf{x}^{(1)}), \ldots, \hat{y}_{\text{lf}}(\mathbf{x}^{(n)})]^T, && \tilde{\mu} = \mu/(2\sigma^2) \\ &\mathbf{R} := (R(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}))_{i,j} \in \mathbb{R}^{n\times n}, && \mathbf{r} := (R(\mathbf{x}^{(i)}, \mathbf{x}))_i \in \mathbb{R}^n \end{aligned} \tag{13}$$

Once the weight coefficients $\mathbf{w} = [w^{(1)}, \ldots, w^{(n)}]^T$ are obtained, the HK predictor for any untried $\mathbf{x}$ is given by

$$\hat{y}(\mathbf{x}) = \begin{bmatrix} \mathbf{r}(\mathbf{x}) \\ \hat{y}_{\text{lf}}(\mathbf{x}) \end{bmatrix}^T\begin{bmatrix} \mathbf{R} & \mathbf{F} \\ \mathbf{F}^T & 0 \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{y}_S \\ 0 \end{bmatrix} \tag{14}$$

By inverting the partitioned matrix, the HK predictor can be written in a form as

$$\hat{y}(\mathbf{x}) = \beta_0\hat{y}_{\text{lf}}(\mathbf{x}) + \mathbf{r}^T(\mathbf{x})\underbrace{\mathbf{R}^{-1}(\mathbf{y}_S - \beta_0\mathbf{F})}_{=:\mathbf{v}_{\text{HK}}} \tag{15}$$

where $\beta_0 = (\mathbf{F}^T\mathbf{R}^{-1}\mathbf{F})^{-1}\mathbf{F}^T\mathbf{R}^{-1}\mathbf{y}_S$ is a scaling factor, which indicates how much the low- and high-fidelity functions are correlated to each other. Note that the vector $\mathbf{V}_{\text{HK}} \in \mathbb{R}^n$ only depends on the observed data, and it can be calculated at the model-fitting stage of HK. Once $\mathbf{V}_{\text{HK}}$ is obtained, the prediction of the unknown y at any untried $\mathbf{x}$ only requires recalculating $\mathbf{r}^T(\mathbf{x})$ and $\hat{y}_{\text{lf}}(\mathbf{x})$. Note that $\hat{y}_{\text{lf}}(\mathbf{x})$ is the low-fidelity kriging given by Eq. (4).

The MSE of the HK prediction is readily proven to be

$$\begin{aligned} \text{MSE}\{\hat{y}(\mathbf{x})\} = \sigma^2\{&1.0 - \mathbf{r}^T\mathbf{R}^{-1}\mathbf{r} + [\mathbf{r}^T\mathbf{R}^{-1}\mathbf{F} \\ &- \hat{y}_{\text{lf}}(\mathbf{x})](\mathbf{F}^T\mathbf{R}^{-1}\mathbf{F})^{-1}[\mathbf{r}^T\mathbf{R}^{-1}\mathbf{F} - \hat{y}_{\text{lf}}(\mathbf{x})]^T\} \end{aligned} \tag{16}$$

From the previous formula, one can find out that the behavior of the approximated low-fidelity function $\hat{y}_{\text{lf}}(\mathbf{x})$ is explicitly included in the MSE of $\hat{y}(\mathbf{x})$, which allows for a better MSE estimation than with traditional kriging. This is one of the unique features of our HK modeling method. Compared to the traditional

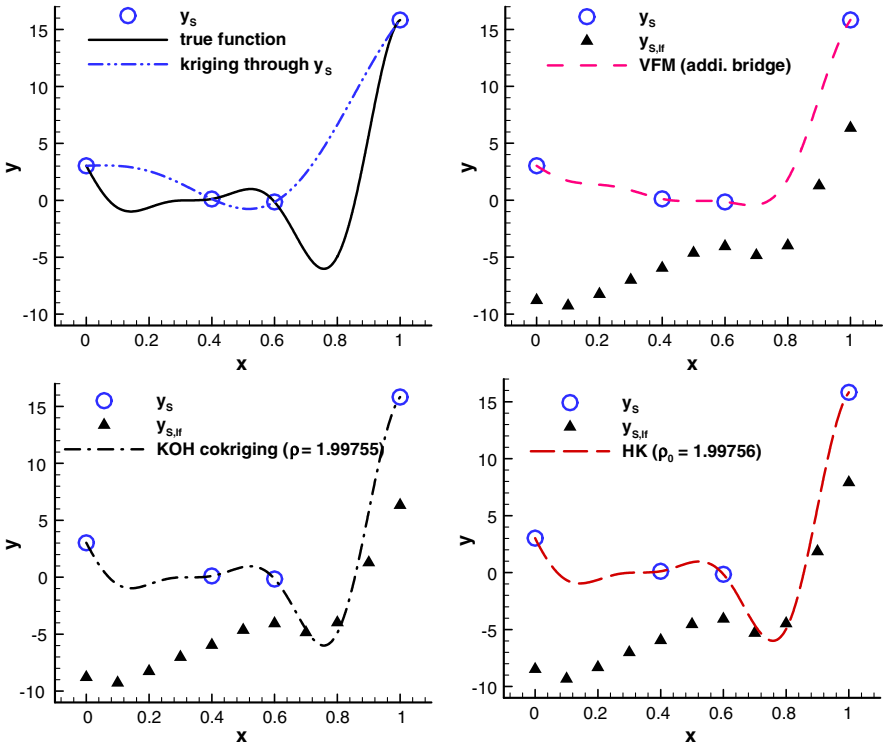
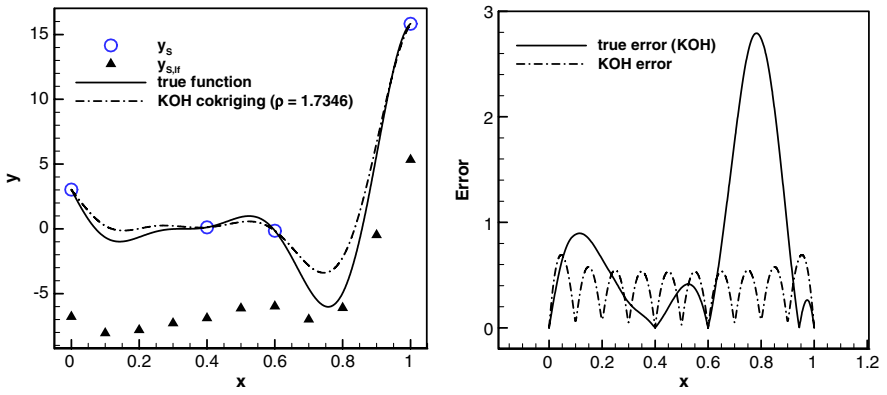


Fig. 1 Comparison of different approximation methods for an analytical example taken from [23].

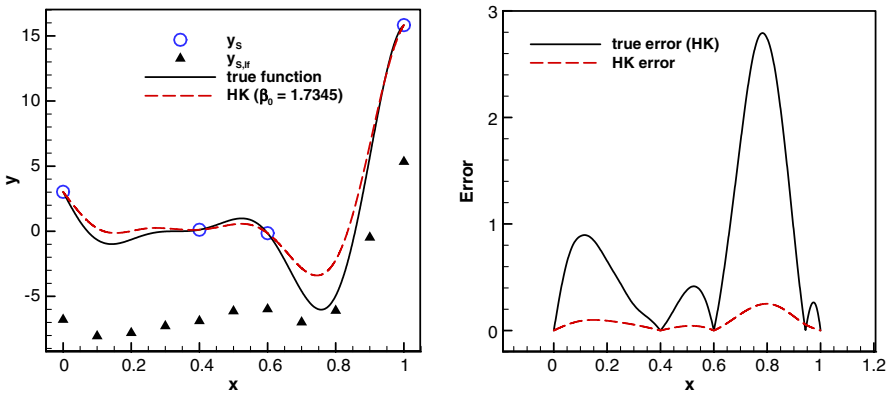
cokriging model [22], the HK model does not require modeling the "cross correlation" between the low- and high-fidelity functions, resulting in reduced model complexity and smaller size of the correlation matrix.

D. Correlation Models

The construction of the correlation matrix **R** and the correlation vector **r** requires the calculation of the correlation function. The correlation function for random parameters at two sites, **x**, **x**′,



a) KOH method



b) Hierarchical kriging method

Fig. 2 Comparison of KOH and HK methods for an analytical example taken from [23] but with modified low-fidelity function: a) KOH cokriging, and b) HK.

is assumed to only depend on the spatial distance, namely $R(\mathbf{x}, \mathbf{x}') = R(\mathbf{x} - \mathbf{x}')$. Here, we focus on a family of correlation models that are of the form

$$R(\mathbf{x}, \mathbf{x}') = \prod_{k=1}^{m} R_k(\theta_k, x_k - x_k') \tag{17}$$

where $\theta = (\theta_1, \ldots, \theta_m) \in \mathbb{R}^m$ are the distance weights (key hyperparameters to be tuned). The most popular form of this correlation function is the Gaussian exponential function. It is of the form

$$\begin{aligned} R_k(\theta_k, x_k - x_k') &= \exp(-\theta_k |x_k - x_k'|^{p_k}) \\ R(\boldsymbol{\theta}, \mathbf{x}, \mathbf{x}') &= \prod_{k=1}^{m} \exp(-\theta_k |x_k - x_k'|^{p_k}), \qquad 1 < p_k \leq 2 \end{aligned} \tag{18}$$

Note that the Gaussian exponential function is not second-order differentiable when $p_k < 2$. In this article, a cubic spline correlation function [35] is employed as it is nearly as good as the Gaussian exponential function but always second-order differentiable, which is a desired feature when gradient information is to be incorporated. It is of the form

$$R_k(\theta_k, x_k - x_k') = \begin{cases} 1 - 15\xi_k^2 + 30\xi_k^3 & \text{for } 0 \leq \xi_k \leq 0.2 \\ 1.25(1 - \xi_k)^3 & \text{for } 0.2 < \xi_k < 1 \\ 0 & \text{for } \xi_k \geq 1 \end{cases},$$

$$\text{where } \xi_k = \theta_k |x_k - x_k'| \tag{19}$$

E. Hierarchical Kriging Fitting by Maximum Likelihood Estimation

Assuming that the sampled data are distributed according to a Gaussian process, the responses at sampling sites are considered to be correlated random functions with the corresponding likelihood function given by

$$L(\beta_0, \sigma^2, \boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi\sigma^2)^n |\mathbf{R}|}} \exp\left(-\frac{1}{2} \frac{(\mathbf{y}_S - \beta_0 \mathbf{F})^T \mathbf{R}^{-1} (\mathbf{y}_S - \beta_0 \mathbf{F})^T}{\sigma^2} \right) \tag{20}$$

Optimal estimates of the scaling factor and the process variance,

$$\begin{aligned} \beta_0(\boldsymbol{\theta}) &= (\mathbf{F}^T \mathbf{R}^{-1} \mathbf{F})^{-1} \mathbf{F}^T \mathbf{R}^{-1} \mathbf{y}_S \\ \sigma^2(\boldsymbol{\theta}, \beta_0) &= \frac{1}{n} (\mathbf{y}_S - \beta_0 \mathbf{F})^T \mathbf{R}^{-1} (\mathbf{y}_S - \beta_0 \mathbf{F}) \end{aligned} \tag{21}$$

are obtained analytically yet depend on the unknown hyperparameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$. Note that the scaling factor β_0 is implicitly tuned, whereas in KOH's cokriging method it is explicitly tuned. Substituting it into the associated Eq. (20) and taking the logarithm, we are left with maximizing

$$\ell_n[L(\boldsymbol{\theta})] = -n \ell_n \sigma^2(\boldsymbol{\theta}) - \ell_n |\mathbf{R}(\boldsymbol{\theta})| \tag{22}$$

As there is no closed form solution for $\boldsymbol{\theta}$, it has to be found by numerical optimization, which is of the form

$$\boldsymbol{\theta} = \arg\max_{\theta} \{\ell_n[L(\boldsymbol{\theta})]\} \tag{23}$$

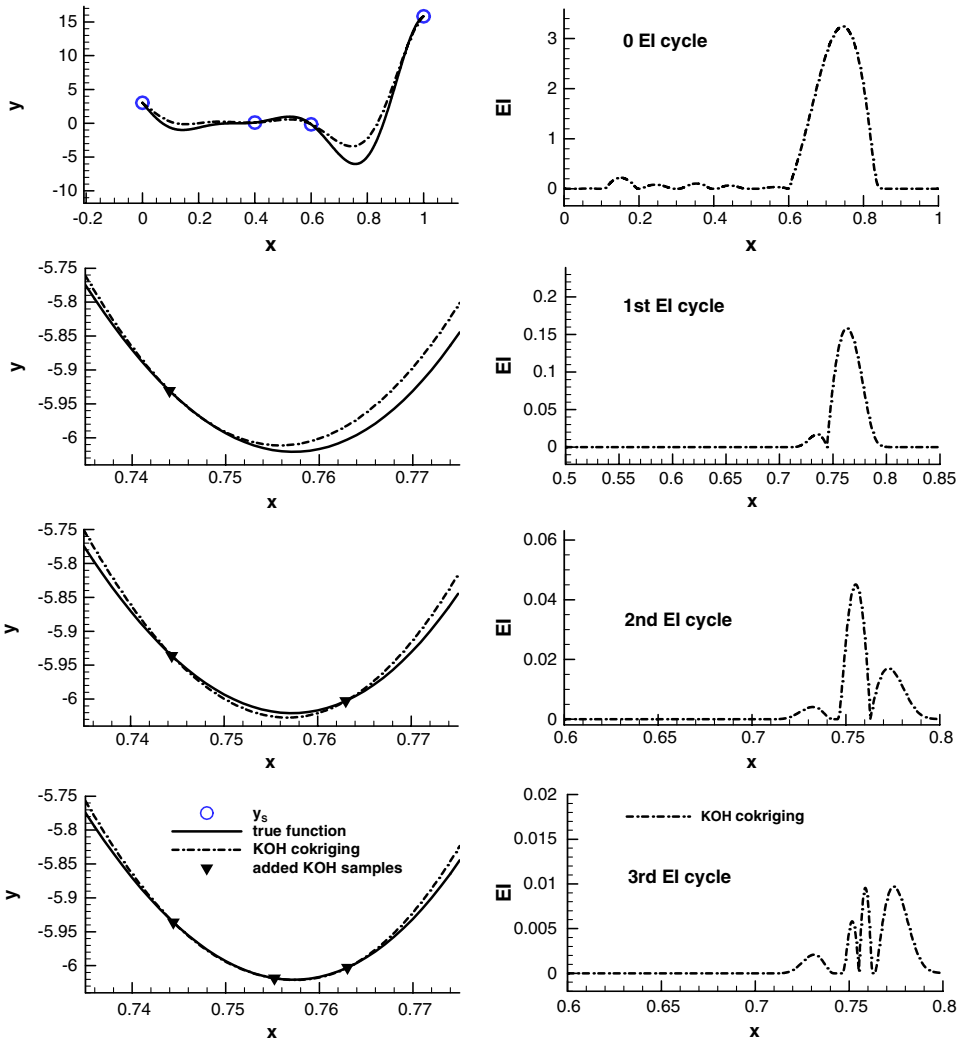


Fig. 3 Refinement of KOH model by expected improvement for an analytical example.

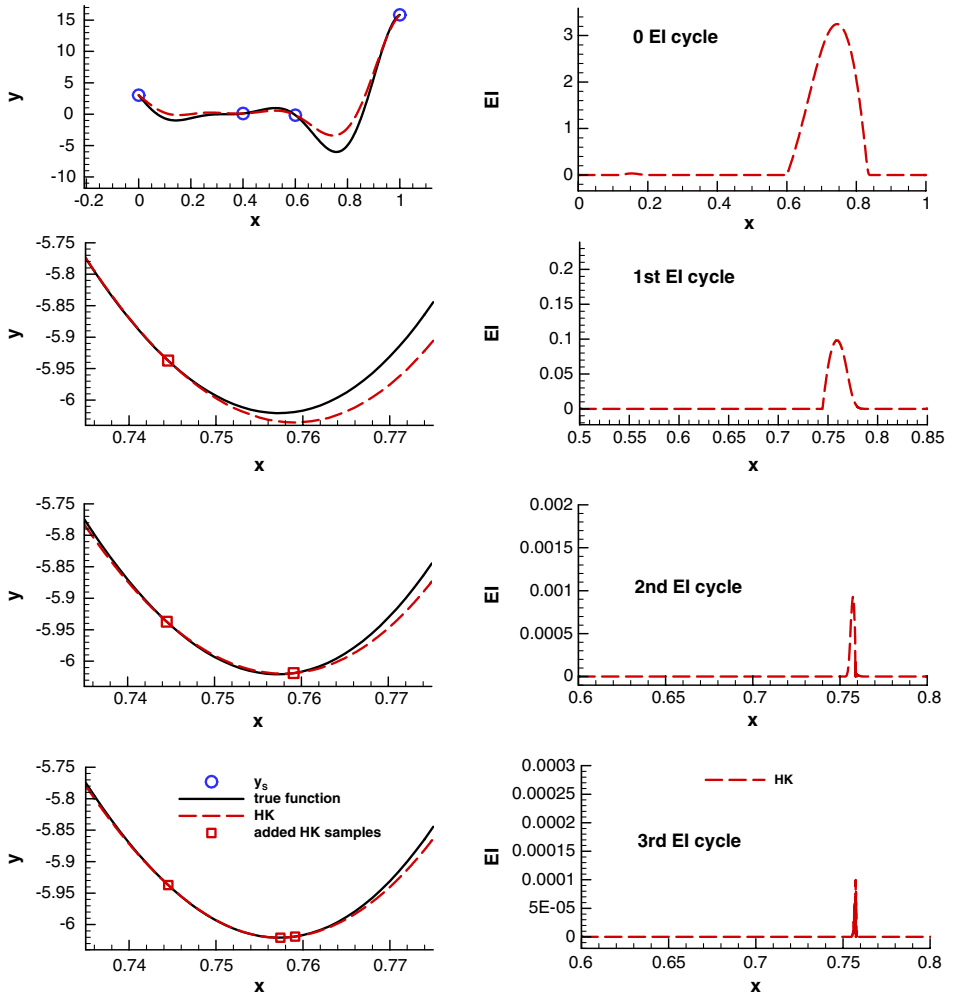


Fig. 4 Refinement of HK model by expected improvement for an analytical example.

Following the approach presented in [36], we solve this optimization problem approximately by using a genetic algorithm to avoid the problem of local search methods when the design space is multimodal. In addition, the range of hyper-parameters is limited to $[10^{-2}\boldsymbol{\theta}_{\text{ini}}, 10^{3}\boldsymbol{\theta}_{\text{ini}}]$, where $\boldsymbol{\theta}_{\text{ini}}$ is the initial guess. The readers are referred to [34] for more hyper-parameter tuning strategies.

III. Numerical Examples

A. Verification of Hierarchical Kriging for One-Dimensional Analytical Problem

To verify the HK method proposed in this article, an analytical test case by Forrester et al. [23] is employed here. The analytical function $y = (6x-2)^2 \sin(12x-4)$, $x \in [0,1]$ represents the high-fidelity function and the low-fidelity function given by $y_{\text{lf}} = 0.5y + 10(x-0.5) - 5$. The x locations of the low- and high-fidelity samples are $\mathbf{S}_{\text{lf}} = \{0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$ and $\mathbf{S} = \{0.0, 0.4, 0.6, 1\}$, respectively. Note that the high-fidelity sample sites are chosen manually. Two high-fidelity samples are placed at the extremes of the parameter space to avoid extrapolation, and two within; purposely, the inner samples are not placed anywhere close to the global minimum of the function.

Figure 1 shows the comparison of different approximation methods with the true function (solid line). The open circle symbols represent the high-fidelity samples, whereas the filled triangles correspond to the low-fidelity samples. The kriging results using the four high-fidelity samples alone are denoted by the dashed-double-dotted line. The results using the VFM method with an additive bridge function are also shown and denoted by a dashed line. The results obtained with our HK (long-dashed line) and KOH's method (dashed-dotted line) notably outperform those computed with standard kriging or the VFM method with an additive bridge function. It is also observed that our HK method is as accurate as KOH's method and that both predictions are in very good agreement with the true function. Note that, in this case, the implicitly tuned scaling factor of our HK method ($\beta_0 = 1.99756$) is very close to the explicitly tuned scaling factor of KOH's method ($\rho = 1.99755$), which in turn validates our HK method.

Next, the example is slightly modified to feature a more complicated correlation between the low- and high-fidelity

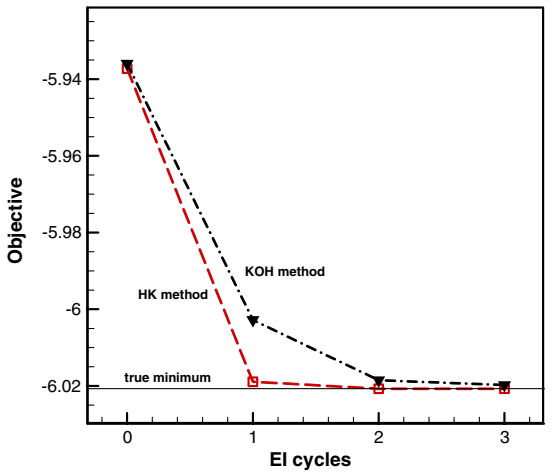
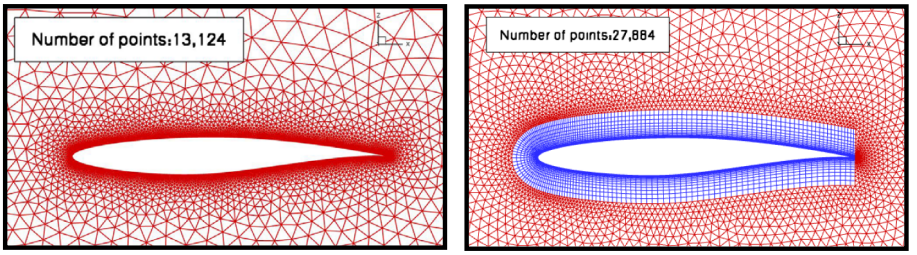


Fig. 5 Convergence history of objective function using expected improvement function based on predicted KOH cokriging error and HK error for an analytical example.

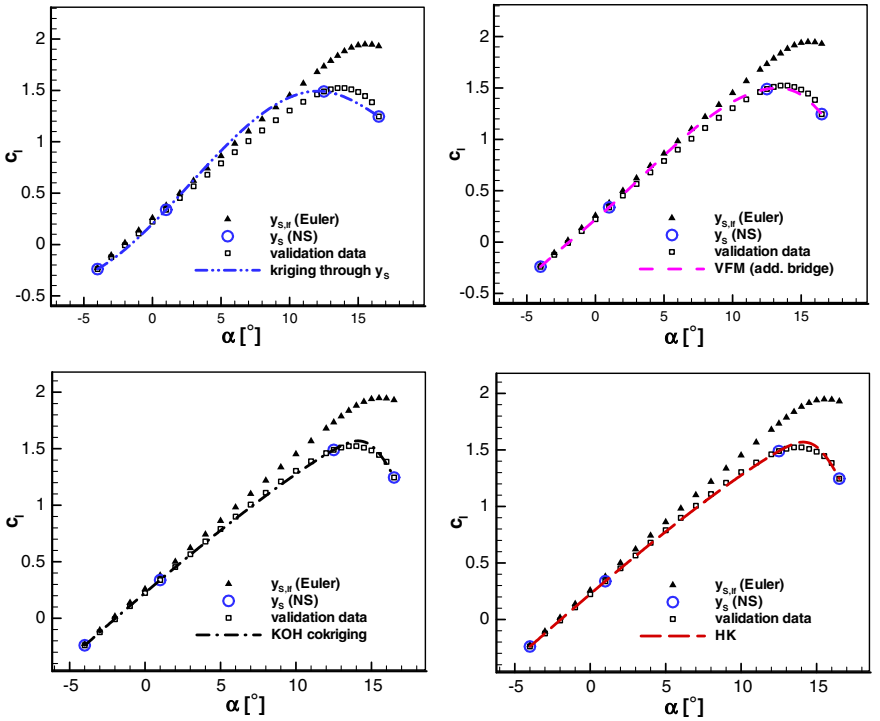


a) Grid for inviscid flow computations –low fidelity (26 samples)

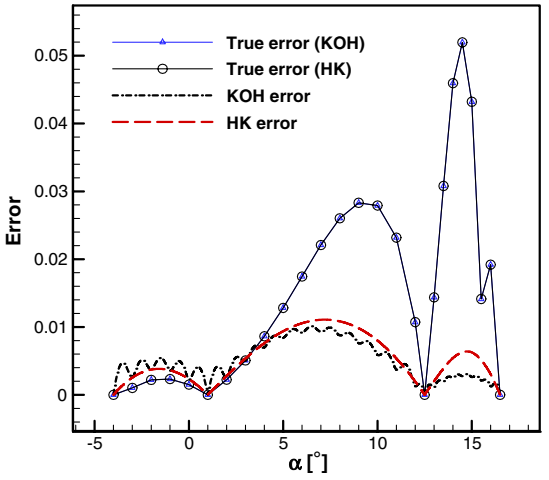
b) Grid for viscous flow computations – high fidelity (4 samples)

Fig. 6 Computational grids for the low- and high-fidelity computations of the flow around an RAE 2822 airfoil.

functions, and a further comparison of HK with KOH's methods is made. The high-fidelity function as well as all the sample sites are fixed, but the low-fidelity function is now given by $y_{\mathrm{lf}} = 0.4y + 10(0.4x^4 + 0.1x^2 + 0.2x + 0.2) - 10$. Figure 2a (left) and Fig. 2b (left) show the comparison of the function values predicted by KOH's and HK methods with the true function, respectively, which reveals that both methods make almost identical predictions, respectively. Figure 2a (right) and Fig. 2b (right) show the comparison of the RMSE predicted by KOH's and HK methods with the true error. Both KOH's and HK methods underestimate the absolute value of the



a)



b)

Fig. 7 Comparison of different approximation methods for the lift coefficient of the RAE 2822 airfoil: a) function values, and b) rms error.

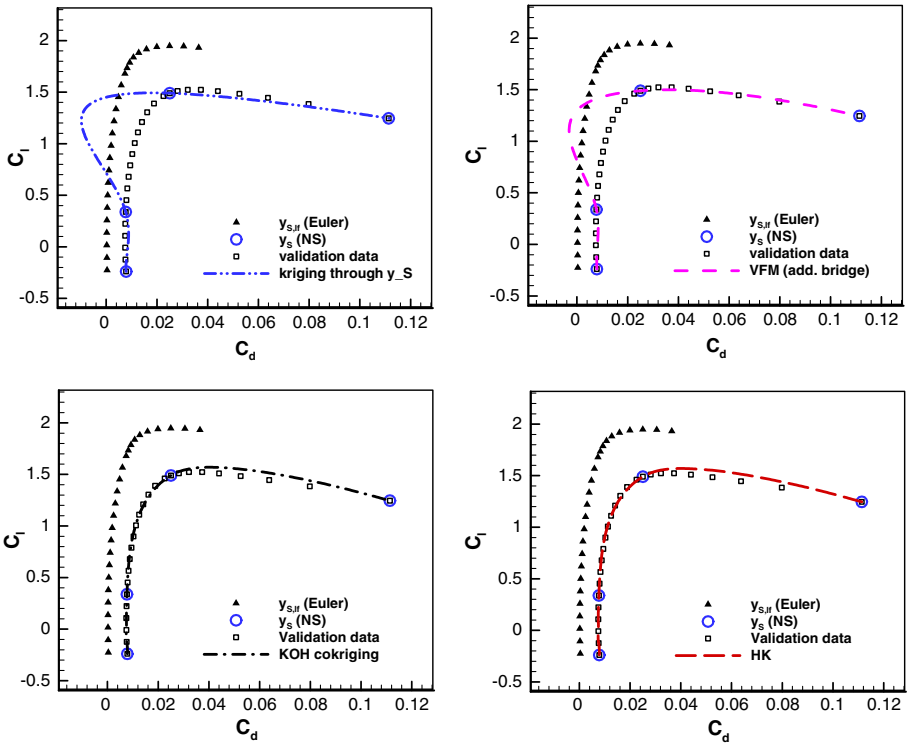


Fig. 8 Comparison of different approximation methods for the drag polar of the RAE 2822 airfoil.

RMSE. However, we find that the RMSE predicted by our HK method (long-dashed line) is more consistent with the trend of the corresponding true error $|\hat{y}(\mathbf{x}) - y(\mathbf{x})|$ (solid line), whereas the RMSE predicted by KOH's method (dashed-dotted line) misses the trend. The reason that HK provides a more reasonable RMSE estimation can be explained by the fact that the behavior of the low-fidelity function is explicitly included RMSE formulation [see Eq. (16)], whereas there is no such property for KOH's method. As the low-fidelity function captures the trend of the high-fidelity function, a better prediction of the RMSE can be obtained.

To further demonstrate the advantages of our HK method for RMSE estimation, the expected improvement (EI) function [18,28] is employed to refine the surrogate models and to improve the prediction of the minimum of the high-fidelity function (objective function). The refinement processes of KOH and HK models are sketched in Figs. 3 and 4, respectively; the convergence history of the objective function is sketched in Fig. 5. The results confirm that HK has advantage over KOH's method when it comes to adaptive sampling.

B. Demonstration of Hierarchical Kriging for RAE 2822 Airfoil with One Independent Variable

Next, we consider an RAE 2822 airfoil and use HK to generate VFM models for the lift and drag coefficients as functions of one independent variable: angle of attack. The Mach number and Reynolds number are taken as 0.2 and 6.5×10^6, respectively. The angle of attack α is varied between -4 and $+16.5$ deg. The computational grids for the Euler and NS computations with the DLR TAU code [37–39] are shown in Fig. 6. We sample the parameter space with 26 points for the Euler computations and four points ($a \in \{-4, 1, 12.5, 16.5 \text{ deg}\}$) for the NS computations. Note that the high-fidelity sample points are chosen manually with two samples placed at the extremes of the parameter space to avoid extrapolation and two within. An additional 26 NS computations are performed to get validation data. In the following figures, the high- and low-fidelity samples are denoted by circles and triangles, respectively, while the validation data is represented by square symbols.

Figure 7 shows a comparison of the different methods for surrogate modeling of the lift coefficient as a function of the angle of attack, respectively. It is shown in Figs. 7a and 7b that the HK method (long-dashed line) is nearly as accurate as KOH's method (dashed-dotted line), while both outperform the VFM with an additive bridge function (dashed line) and the kriging of the four high-fidelity samples alone (dashed-double-dotted line). Figure 7b shows a comparison of the corresponding RMSE as a function of the angle of attack. It is found that the RMSE predicted by HK is smoother and more reasonable than that predicted by KOH's method (more consistent with the true error $|\hat{c}_l(\alpha) - c_l(\alpha)|$). The comparison of different methods for approximating the drag polar is shown in Fig. 8. The same conclusions as for lift versus angle of attack also apply here. If the pitching moment coefficient is of interest, the trend provided by the low-fidelity model (Euler) is inadequate at high angles of attack, as shown in Fig. 9. In such a case, less of a benefit is observed by considering the low-fidelity data. For this rather hard case, a scaling factor of -0.603 is found by the explicit tuning strategy of KOH's method. In contrast, a scaling factor of -0.824 is found by the implicit tuning strategy of the HK model [see Eq. (21)], which leads to a slightly better, more consistent prediction if compared with the validation data. Sample-point refinement based

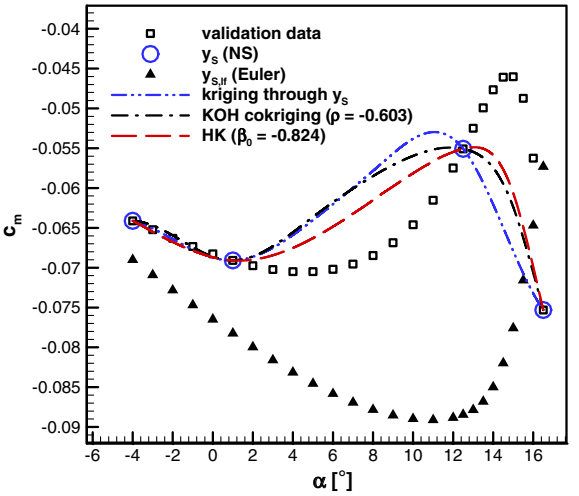


Fig. 9 Comparison of KOH's and HK methods for the moment coefficient of the RAE 2822 airfoil.

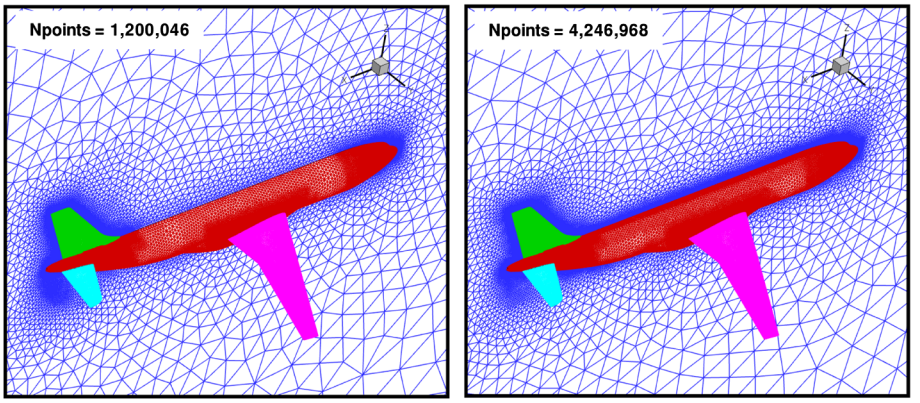


Fig. 10 Computational grids for the Euler and NS computations; left: Euler grid; right: NS grid.

on the predicted error is required to further improve the prediction in such a case.

C. Modeling of the Aerodata of a Three-Dimensional Industrial Aircraft Configuration

Here, we apply the HK method to the efficient prediction of the aerodynamic raw data of an industrial configuration, the DLR F12, with three independent variables: angle of attack α, Mach number Ma, and the deflection angle of the horizontal tail δ. Note that only the integrated aerodynamic data such as C_L, C_D, and C_M are considered.

The computational grids used for the Euler and NS computations with the DLR TAU code [37–39] are shown in Fig. 10. The Euler and NS computations are performed in parallel on 24 domains. An Euler computation on the grid with 1 million grid points roughly takes about 24 min, whereas a single NS computation takes about 3 h wall clock time on the grid with around 4 million grid points. Note that the Spalart–Allmaras one-equation turbulence model is used for the NS computations, assuming a Reynolds number of 6.5 million.

The sample points are chosen using the quasi–Monte Carlo method [40]. The independent variables as well as their range are listed in Table 1. A total of 94 sample points are chosen for the Euler calculations, and 80 sample points are chosen for the NS computations (acting as test points). The parameter space and the 94 sample points are shown in Fig. 11. All the Euler samples are used for building the low-fidelity kriging that aims to capture the global trend of the unknown high-fidelity aerodynamic function, while a varying number of NS samples is used to build an HK model for the high-fidelity aerodynamic function with the kriging of the low-fidelity function being used as trend models according to Eq. (15). As the number of NS samples is increased, the accuracy of the HK models can be evaluated on the test points obtained by the NS computations.

Figure 12 shows the results for the drag coefficient. Both the maximum and average errors are studied as a function of the number of NS samples used to build the surrogate model. The error at the test points is computed by comparing the value predicted by the different surrogate models with the value from the NS computations. The test points are the high-fidelity samples that were not used to build the surrogate model. The line with open circle symbols is for conventional kriging that uses the NS sample points alone, while the line with delta symbols and the line with square symbols are for KOH and HK, respectively. Both KOH and HK combine the Euler and NS samples, with the Euler data capturing the global trend and the NS data correcting the absolute values. For a given number of

Table 1 Independent Variables

Independent variables	Symbol	Range
Angle of attack	α	[−4, 6 deg]
Mach number	Ma	[0.6, 0.9]
Deflection angle of horizontal tail	δ	[−6, 6 deg]

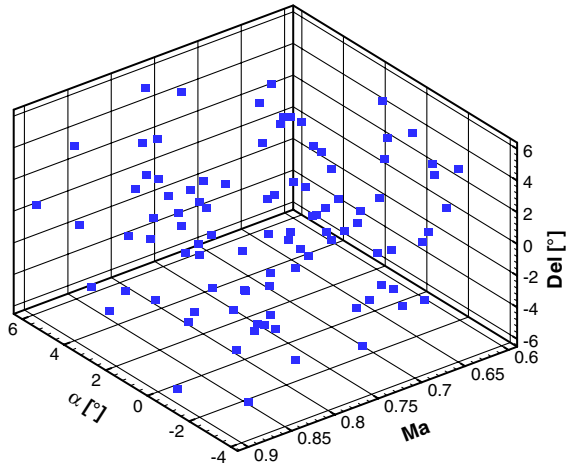


Fig. 11 Parameter space and sample points (94 samples for three variables chosen using the quasi–Monte Carlo method [40]).

high-fidelity samples, the maximum and average errors of the KOH and HK models are lower than that of the kriging based on the same number of NS samples alone. In turn, the number of the NS samples required for obtaining a model with a given maximum or average error can be reduced. As an Euler computation is typically much cheaper than a NS computation, the overall efficiency of producing the aerodynamic raw data can be improved. It is also observed that in this test case HK outperforms the cokriging method by KOH, in terms of both maximum and mean errors. The maximum error is increasing in the case of KOH when the number of high-fidelity samples is between 25 and 60. The reason probably is that the explicitly tuned scaling factor of using KOH is closer to zero, and thus the result is closer to that of kriging through NS samples alone.

As both KOH's and HK methods provide an RMSE estimate, further comparison is conducted by adaptive sampling based on the estimated RMSE. Note that the new NS sample points are always chosen from the remaining of all the 80 NS sample points. In addition, for real applications, refinement by adding a single sample point per refinement cycle is inefficient; refinement by adding multiple sample points per cycle [41–43] is thus a subject of future research. The effect of adaptive sampling on the accuracy of the KOH and HK models is shown in Fig. 13. Based on the model relying on 10 high-fidelity samples, additional samples are chosen adaptively by maximizing the estimated RMSEs (model error). It is shown that, for a given number of samples, the maximum error of both the KOH and HK models can be reduced compared to the results without adaptive sampling (see Fig. 12). It is observed again that HK outperforms KOH cokriging. The reason may be explained by the fact that the RMSE estimation of HK considers the behavior of the low-fidelity function [see Eq. (16)] but KOH cokriging does not.

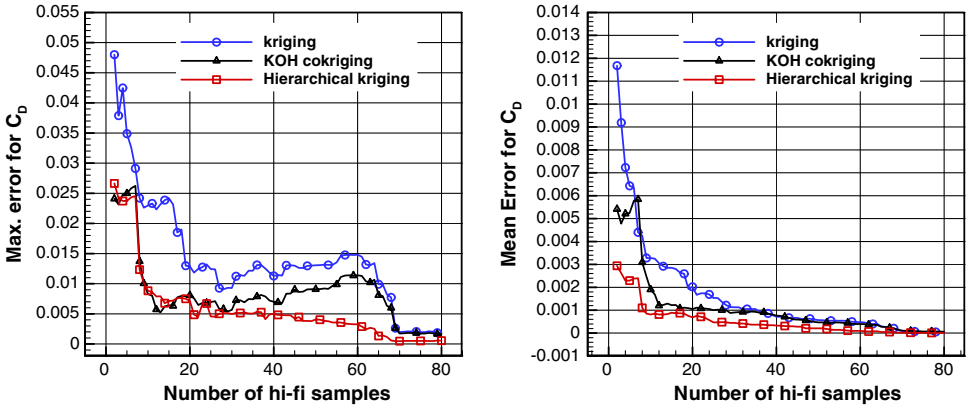


Fig. 12 Comparison of kriging, KOH and HK in terms of maximum and average error in drag coefficient as a function of number of high-fidelity samples.

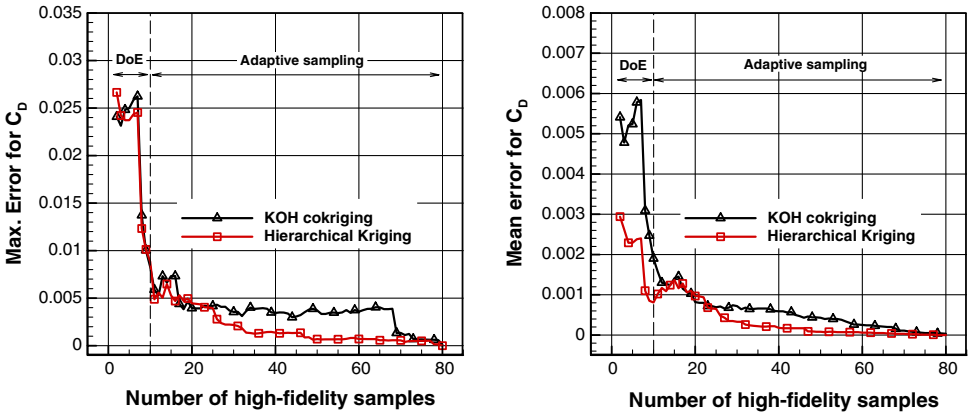


Fig. 13 Effect of adaptive sampling on the accuracy of KOH and HK models in terms of maximum and average error in drag coefficient as a function of number of high-fidelity samples.

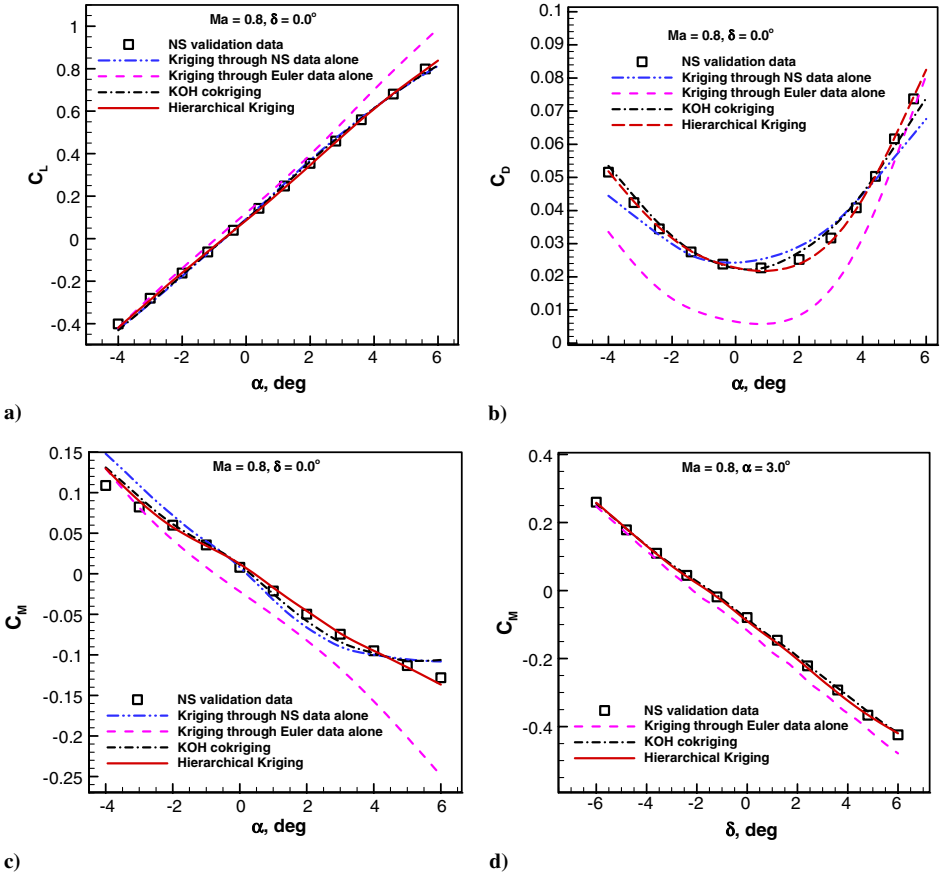


Fig. 14 Comparison of kriging, KOH cokriging and HK predictions with validation data for lift, drag, and pitching moment coefficients (25 NS samples and 94 Euler samples).

Note that another benefit of HK is that its correlation matrix is smaller than that of a KOH cokriging, and thus the building of the surrogate model itself is more efficient.

The results are more closely examined by making cuts through the resulting response surfaces for the lift, drag, and pitching moment coefficients, built from 25 NS samples and 94 Euler samples chosen by the design of experiments method. The results for a fixed Mach number of $Ma = 0.8$ and a deflection angle of the horizontal tail of $\delta = 0$ deg are shown in Figs. 14a–14c. The results for a fixed Mach number of $Ma = 0.8$ and an angle of attack of $\alpha = 3$ deg are shown in Fig. 14d. It is shown that both the KOH and HK models give accurate results compared with the NS validation data and that HK slightly outperforms KOH cokriging.

IV. Conclusions

A hierarchical kriging for variable-fidelity surrogate modeling of high-fidelity functions based on a set of computer codes with varying degrees of fidelity and computational expense was developed. The hierarchical kriging was validated against an analytical problem and demonstrated for aerodynamic examples for an RAE 2822 airfoil and a three-dimensional industrial aircraft configuration. Some conclusions can be drawn as follows.

1) Hierarchical kriging is an extension of conventional kriging to a model that directly takes the lower-fidelity aerodynamic function as a model trend of the kriging for the high-fidelity function of interest. As a consequence, the trend of the lower-fidelity function is mapped to the sampled high-fidelity data, resulting in a more accurate surrogate model for the high-fidelity function of interest.

2) Hierarchical kriging is as simple (and robust) as the correction-based variable-fidelity modeling methods and as accurate as cokriging. The correlation matrix of hierarchical kriging is smaller than that of cokriging. Moreover, it provides a more reasonable estimate of the mean squared error than any of the existing kriging and cokriging methods. This is due to the fact that the behavior of the lower-fidelity function is explicitly incorporated into the formula for the mean squared error. The improved mean-squared-error estimate is more suitable for efficient global optimization (EGO) via the expected improvement function, for example.

3) A scaling factor accounting for the influence of the low-fidelity function on the prediction of the high-fidelity function is implicitly tuned during the model fitting stage of hierarchical kriging. It is identical to the explicitly tuned scaling factor of Kennedy and O'Hagan's cokriging model.

Besides the application to aerodynamic problems, the hierarchical kriging is also well-suited for use in any other research areas where computer codes of varying fidelity are in use. Hierarchical kriging will be applied to EGO as future research.

Appendix: Hierarchical Kriging for Multiple Levels of Fidelity

Assume that l sets of data corresponding to l levels of fidelity data obtained by running computer codes of varying fidelity and computational expense are available, with the data of the highest level of fidelity denoted by level 1. The aerodynamic functions corresponding to these sets of data are defined as

$$y_i = f_i(\mathbf{x}), \qquad i = 1, \ldots, l \tag{A1}$$

where y_i: $\mathbb{R}^m \to \mathbb{R}$. The corresponding sampled data are defined as

$$\mathbf{y}_{S,i} = [y_i^{(1)}, \ldots, y_i^{(n_i)}]^T, \qquad i = 1, \ldots, l \tag{A2}$$

where n_i denotes the number of sample points for ith level of data (typically $n_1 \ll n_2 \ll \ldots \ll n_l$).

Note that here we are concerned with the prediction of the aerodynamic function of the highest fidelity (level 1) based on all the sample data. To this end, we propose to implement HK as follows (see Fig. A1 for the schematics of building procedure).

1) First, we build a kriging $\hat{y}_l$ based on the sampled data of the lowest fidelity (see Sec. II.B).

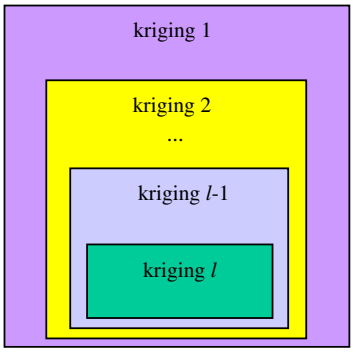


Fig. A1 Schematics of building hierarchical kriging for multiple levels of fidelity.

2) Then, we build a kriging $\hat{y}_{l-1}$ for level $l-1$, with the following assumption for the corresponding random function:

$$Y_{l-1}(\mathbf{x}) = \beta_{0,l}\hat{y}_l + Z_{l-1}(\mathbf{x}) \tag{A3}$$

Here, the kriging $\hat{y}_l$ serves as the model trend of $\hat{y}_{l-1}$. Note that the building procedure of $\hat{y}_{l-1}$ is same as that described in Sec. II.

3) Subsequently, we can build the kriging $\hat{y}_i$ for ith level with the lower-fidelity kriging $\hat{y}_{i+1}$ serving as its model trend.

4) The aforementioned procedure is repeated until the kriging $\hat{y}_1$ for the aerodynamic function of the highest fidelity is built. The corresponding random function is defined as

$$Y_1(\mathbf{x}) = \beta_{0,2}\hat{y}_2 + Z_1(\mathbf{x}) \tag{A4}$$

Once $\hat{y}_2$ is built, the same procedure as described in Sec. II also applies here for building $\hat{y}_1$.

Acknowledgments

This research was sponsored by the European Regional Development Fund, Economic Development Fund of the Federal German State of Lower Saxony; contract/grant W3-80026826. The first author was also supported by the National Natural Science Foundation of China under grant 10902088 and the Aeronautical Science Foundation of China under grant 2011ZA53008. The authors are grateful for the comments of the reviewers who helped to improve on the original manuscript.

References

[1] Queipo, N. V., Haftka, R. T., Shyy, W., Goel, T., Vaidyanathan, R., and Tucher, P. K., "Surrogate-Based Analysis and Optimization," *Progress in Aerospace Sciences*, Vol. 41, 2005, pp. 1–28. doi:10.1016/j.paerosci.2005.02.001

[2] Simpson, T. W., Toropov, V., Balabanov, V., and Viana, F. A. C., "Design and Analysis of Computer Experiments in Multidisciplinary Design Optimization: A Review of How Far We Have Come—Or Not," 12th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, Victoria, B.C., Canada, AIAA Paper 2008-5802, Sept. 2008.

[3] Krige, D. G., "A Statistical Approach to Some Basic Mine Valuations Problems on the Witwatersrand," *Journal of the Chemical, Metallurgical and Mining Engineering Society of South Africa*, Vol. 52, No. 6, 1951, pp. 119–139.

[4] Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P., "Design and Analysis of Computer Experiments," *Statistical Science*, Vol. 4, 1989, pp. 409–423. doi:10.1214/ss/1177012413

[5] Shan, S., and Wang, G. G., "Survey of Modeling and Optimization Strategies for High-Dimensional Design Problems," 12th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, Victoria, B.C., Canada, AIAA Paper 2008-5842, Sept. 2008.

[6] Wang, G. G., and Shan, S., "Review of Metamodeling Techniques in Support of Engineering Design Optimization," *Journal of Mechanical Design*, Vol. 129, No. 4, 2007, pp. 370–380. doi:10.1115/1.2429697

[7] Koch, P. N., Simpson, T. W., Allen, J. K., and Mistree, F., "Statistical

Approximations for Multidisciplinary Design Optimization: The Problem of the Size," *Journal of Aircraft*, Vol. 36, No. 1, 1999, pp. 275–286.
doi:10.2514/2.2435

[8] Chang, K. J., Haftka, R. T., Giles, G. L, and Kao, P.-J., "Sensitivity-Based Scaling for Approximating Structural Response," *Journal of Aircraft*, Vol. 30, No. 2, 1993, pp. 283–288.
doi:10.2514/3.48278

[9] Alexandrov, N. M., Lewis, R. M., Gumbert, C. R., Green, L. L., and Newman, P. A., "Optimization with Variable-Fidelity Models Applied to Wing Design," 38th Aerospace Sciences Meeting and Exhibit, Reno, NV, AIAA Paper 2000-0841, Jan. 2000.

[10] Alexandrov, N. M., Nielsen, E. J., Lewis, R. M., and Anderson, W. K., "First-Order Model Management with Variable-Fidelity Physics Applied to Multi-Element Airfoil Optimization," 8th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, Long Beach, CA, AIAA Paper 2000-4886, Sept. 2000.

[11] Alexandrov, N. M., Dennis, J. E., Jr., Lewis, R. M., and Torczon, V., "A Trust-Region Framework for Managing the Use of Approximation Models in Optimization," *Structural Optimization*, Vol. 15, No. 1, 1998, pp. 16–23.

[12] Choi, S., Alonso, J. J., Kroo, I. M., and Wintzer, M., "Multi-Fidelity Design Optimization of Low-Boom Supersonic Business Jets," 45th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, Palm Springs, CA, AIAA Paper 2004-1530, April 2004.

[13] Choi, S., Alonso, J. J., Kim, S., Kroo, I., and Wintzer, M., "Two-Level Multi-Fidelity Design Optimization Studies for Supersonic Jets," 43rd Aerospace Sciences Meeting and Exhibit, Reno, NV, AIAA Paper 2005-53, Jan. 2005.

[14] Tang, C. Y., Gee, K., and Lawrence, S. L., "Generation of Aerodynamic Data Using a Design of Experiment and Data Fusion Approach," 43rd Aerospace Sciences Meeting and Exhibit, Reno, NV, AIAA Paper 2005-1137, Jan. 2005.

[15] Han, Z.-H., Görtz, S., and Hain, R., "A Variable-Fidelity Modeling Method for Aero-Loads Prediction," *New Results in Numerical and Experimental Fluid Mechanics, 7. Notes on Numerical Fluid Mechanics and Multidisciplinary Design*, Vol. 112, edited by A. Dillmann, G. Heller, M. Klaas, H.-P. Kreplin, W. Nitsche, W. Schröder, Springer, New York, 2010, pp. 17–25.

[16] Gano, S. E., Renaud, J. E., and Sanders, B., "Variable Fidelity Optimization Using a Kriging Based Scaling Function," 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, Albany, NY, AIAA Paper 2004-4460, Aug. 2004.

[17] Gano, S. E., Renaud, J. E., Martin, J. D., and Simpson, T. W., "Update Strategies for Kriging Models for Use in Variable Fidelity Optimization," 46th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, Austin, TX, AIAA Paper 2005-2057, April 2005.

[18] Ghoreyshi, M., Badcock, K. J., and Woodgate, M. A., "Accelerating the Numerical Generation of Aerodynamic Models for Flight Simulation," *Journal of Aircraft*, Vol. 46, No. 3, 2009, pp. 972–980.
doi:10.2514/1.39626

[19] Han, Z.-H., Görtz, S., and Zimmermann, R., "Improving Variable-Fidelity Surrogate Modeling via Gradient-Enhanced Kriging and a Generalized Hybrid Bridge Function," *Aerospace Science and Technology*, Jan. 2012. doi:10.1016/j.ast.2012.01.006.

[20] Jameson, A., "Optimum Aerodynamic Design Using CFD and Control Theory," 12th AIAA Computational Fluid Dynamics Conference, San Diego, CA, AIAA Paper 1995-1729, June 1995.

[21] Journel, A. G., and Huijbregts, J. C., *Mining Geostatistics*, Academic Press, New York, 1978, p. 600.

[22] Kennedy, M. C., and O'Hagan, A., "Predicting the Output from a Complex Computer Code When Fast Approximations Are Available," *Biometrika*, Vol. 87, No. 1, 2000, pp. 1–13.
doi:10.1093/biomet/87.1.1

[23] Forrester, A. I. J., Sóbester, A., and Keane, A. J., "Multi-Fidelity Optimization via Surrogate Modeling," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Vol. 463, No. 2088, 2007, pp. 3251–3269.
doi:10.1098/rspa.2007.1900

[24] Forrester, A. I. J., and Keane, A. J., "Recent Advances in Surrogate-Based Optimization," *Progress in Aerospace Sciences*, Vol. 45, No. 1–3, Jan.–April 2009, pp. 50–79.
doi:10.1016/j.paerosci.2008.11.001

[25] Kuya, Y., Takeda, K., Zhang, X., and Forrester, A. I. J., "Multifidelity Surrogate Modeling of Experimental and Computational Aerodynamic Data Sets," *AIAA Journal*, Vol. 49, No. 2, Feb. 2011, pp. 289–298.
doi:10.2514/1.J050384

[26] Han, Z.-H., Zimmermann, R., and Görtz, S., "Alternative Cokriging Model for Variable-Fidelity Surrogate Modeling," *AIAA Journal*, Vol. 50, No. 5, 2012, pp. 1205–1210.
doi:10.2514/1.J051243

[27] Zimmermann, R., and Han, Z.-H., "Simplified Cross-Correlation Estimation for Multi-Fidelity Surrogate Cokriging Models," *Advances and Applications in Mathematical Sciences*, Vol. 7, No. 2, 2010, pp. 181–201.

[28] Jones, D. R., Schonlau, M., and Welch, W. J., "Efficient Global Optimization of Expensive Black-Box Functions," *Journal of Global Optimization*, Vol. 13, 1998, pp. 455–492.
doi:10.1023/A:1008306431147

[29] Matheron, G. M., "Principles of geostatistics," *Economic Geology*, Vol. 58, No. 8, 1963, pp. 1246–1266.
doi:10.2113/gsecongeo.58.8.1246

[30] Rivoirard, J., "On the Structural Link Between Variables in Kriging with External Drift," *Mathematical geology*, Vol. 34, No. 7, 2002, pp. 797–807.
doi:10.1023/A:1020972510120

[31] Hengl, T., Geuvelink, G. B. M., and Stein, A., "Comparison of Kriging with External Drift and Regression-Kriging," International Institute for Geo-Information Science and Earth Observation, TN, 2003, p. 18.

[32] Simpson, T. W., Mauery, T. M., Korte, J. J., and Mistree, F., "Kriging Models for Global Approximation in Simulation-Based Multidisciplinary Design Optimization," *AIAA Journal*, Vol. 39, No. 12, 2001, pp. 2233–2241.
doi:10.2514/2.1234

[33] Martin, J. D., and Simpson, T. W., "Use of Kriging Models to Approximate Deterministic Computer Models," *AIAA Journal*, Vol. 43, No. 4, 2005, pp. 853–863.
doi:10.2514/1.8650

[34] Toal, D. J. J., Bressloff, N. W., and Kean, A. J., "Kriging Hyperparameter Tuning Strategies," *AIAA Journal*, Vol. 46, No. 5, 2008, pp. 1240–1252.
doi:10.2514/1.34822

[35] Lophaven, S. N., Nielsen, H. B., and Søndergaard, J., "DACE—A Matlab Kriging Toolbox (Version 2.0)," Technical University of Denmark, Informatics and Mathematical Modelling, TR IMM-REP-2002-12, Copenhagen, Denmark, 2002, pp. 1–34.

[36] Deb, K., "An Efficient Constraint Handling Method for Genetic Algorithms," *Computer Methods in Applied Mechanics and Engineering*, Vol. 186, 2000, pp. 311–338.
doi:10.1016/S0045-7825(99)00389-8

[37] Gerhold, T., Hannemann, V., and Schwamborn, D., "On the Validation of the DLR-TAU Code," *New Results in Numerical and Experimental Fluid Mechanics*, edited by W. Nitsche, H.-J. Heinemann, and R. Hilbig, Notes on Numerical Fluid Mechanics, Vol. 72, Springer–Verlag, Berlin, 1999, pp. 426—433.

[38] Kroll, N., and Fassbender, J. K. (Eds.), "MEGAFLOW—Numerical Flow Simulation for Aircraft Design," *Notes on Numerical Fluid Mechanics and Multidisciplinary Design*, Springer–Verlag, Berlin, Vol. 89, 2005.

[39] Schwamborn, D., Gerhold, T., and Heinrich, R., "The DLR TAU-Code: Recent Applications in Research and Industry," *Proceedings of the European Conference on Computational Fluid Dynamics*, edited by P. Wesseling, E. Oñate, and J. Périaux, Delft Univ. of Technology, Delft, The Netherlands, 2006.

[40] Lemieux, C., "Monte Carlo and Quasi–Monte Carlo Sampling," *Springer Series in Statistics*, Springer, New York, 2009, p. 373.

[41] Viana, F. A. C., Haftka, R. T., and Watson, L. T., "Why Not Run the Efficient Global Optimization Algorithm with Multiple Surrogates," 51th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Orlando, FL, AIAA Paper 2010-3090, April 2010.

[42] Ponweiser, W., Wagner, T., and Vincze, M., "Clustered Multiple Generalized Expected Improvement: A Novel Infill Sampling Criterion for Surrogate Models," *IEEE Congress on Evolutionary Computation*, IEEE Publ., Piscataway, NJ, June 2008, pp. 3514–3521.

[43] Ginsbourger, D., Le Riche, R., and Carraro, L., "Kriging Is Well-Suited to Parallelize Optimization," *Computational Intelligence in Expensive Optimization Problems*, edited by Y. Tenne, and C. Goh, Springer–Verlag, Berlin, Vol. 2, 2010, pp. 131–162.

R. Haftka
Associate Editor