



Sequential optimization using multi-level cokriging and extended expected improvement criterion

Yixin Liu^{1,2} · Shishi Chen³ · Fenggang Wang^{1,2} · Fenfen Xiong^{1,2}

Received: 25 September 2017 / Revised: 13 January 2018 / Accepted: 18 February 2018 / Published online: 20 March 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

To reduce the computational cost of metamodel based design optimization that directly relies on the computationally expensive simulation, the multi-fidelity cokriging method has gained increasing attention by fusing data from two or more models with different levels of fidelity. In this paper, an enhanced cokriging based sequential optimization method is proposed. Firstly, the impact of considering full correlation of data among all models on the hyper-parameter estimation during cokriging modeling is investigated by setting up a unified maximum likelihood function. Then, to reduce the computational cost, an extended expected improvement function is established to more reasonably identify the location and fidelity level of the next response evaluation based on the original expected improvement criterion. The results from comparative studies and one airfoil aerodynamic optimization application show that the proposed cokriging based sequential optimization method is more accurate in modeling and efficient in model evaluation than some existing popular approaches, demonstrating its effectiveness and relative merits.

Keywords Cokriging · Gaussian process · Hyper-parameter estimation · Sequential sampling

Nomenclature

EI	Expected improvement
EEI	Extend expected improvement
CFD	Computational fluid dynamics
GP	Gaussian process
RMSE	Root-mean-square mean
LHD	Latin Hypercube design
\mathbf{d}	Response data
n	Number of sample points
t	Fidelity level of models
$y^t(\mathbf{x})$	t^{th} -level-fidelity model
D_t	Input sites
Γ	Input space

$R_t(\mathbf{x}, \mathbf{x}')$	Correlation matrix
θ	Roughness parameters
β	Regression coefficients
ρ	Scaling factor
σ^2	Spatial covariance parameter
δ	Discrepancy function
C^i	Calculation cost of the metamodel
$\text{Corr}(\mathbf{x}, t)$	Model correlation
$\eta(\mathbf{x}, t)$	Sample density function
$CR(t)$	Ratio of cost

Responsible Editor: Joaquim R. R. A. Martins

✉ Fenfen Xiong
fenfenx@bit.edu.cn

- ¹ School of Aerospace Engineering, Beijing Institute of Technology, Beijing 100081, People's Republic of China
- ² Key Laboratory of Dynamics and Control of Flight Vehicle, Ministry of Education, Beijing 100081, People's Republic of China
- ³ Beijing Electro-Mechanical Engineering Institute, Beijing 100074, People's Republic of China

1 Introduction

The application of surrogate models or metamodels to predicting the response of a computationally expensive simulation function has grown in popularity for engineering design optimization over the past 15 years (Jin et al. 2001; Simpson et al. 2001; Yang et al. 2005; Shi et al. 2013). Sometimes, the evaluation on the complex system of interest is so expensive that straightforward application of the metamodel-based optimization might be too costly. Generally, a complicated physical process can be modeled using several methods with different levels of fidelity, or a computer code for a complex

problem can be run at different levels of fidelity. For example, the aerodynamic analysis in aerospace engineering can be implemented using the high-fidelity computational fluid dynamics (CFD) simulation or low-fidelity engineering handbook equations; further CFD can also be solved with fine (high-fidelity) or coarse (low-fidelity) mesh. Therefore, the multi-fidelity modeling approach that combines the data from two or more models/codes (uniformly referred to as models hereafter in this paper) with different levels of fidelity to efficiently construct a metamodel has gained increasing attention in optimization design (Qian and Wu 2008; Zheng et al. 2013; Park et al. 2016; Ng and Eldred 2012; Tuo et al. 2013; Chen et al. 2015; El-Beltagy and Wright 2001). The differences between various MFS frameworks have been investigated in detail with the aid of examples including algebraic functions and a borehole example by Park et al. (2017). As a well-known multi-fidelity modeling technique with good computational characteristics, the cokriging method combining two data sets respectively from high- and low-fidelity responses based on the Gaussian process theory has been widely studied and applied to aerodynamic analysis (Han et al. 2012; Laurenceau and Sagaut 2008) and aerodynamic optimization (Toal and Keane 2011; Huang et al. 2013), robust optimization of gas-turbine compressor blade (Keane 2015), and prediction of soil bulk density (Yang et al. 2016). To handle multi-level computer models with a hierarchy of fidelity that often exist in engineering problems, the original two-level cokriging method has been extended to multi-level (s -level) cokriging using a hierarchical updating treatment (Kennedy and O'Hagan 2000). However, it is often difficult to know beforehand how many samples are needed to construct a sufficiently accurate metamodel, and thus sequential optimization based on the bi- and multi-level cokriging methods have been introduced to further reduce the computational cost (Xiong et al. 2008; Huang et al. 2006; Gratiet and Cannamela 2015) by employing the sequential sampling strategies (Jin et al. 2002; Xiong et al. 2009; Zhu et al. 2015).

For the multi-level cokriging based sequential optimization, one key procedure is to estimate the unknown hyper-parameters during cokriging modeling, which evidently has impact on the accuracy of metamodel as well as optimization. In the well-known Bayesian discrepancy-based autoregressive multi-fidelity model using GP proposed by Kennedy and O'Hagan, the assumption of a Markov property for the covariance structure of the observed data is made, and these hyper-parameters are divided into s groups and then estimated group by group using the maximum likelihood estimation method considering the correlation of the observed data from two models with adjacent fidelities (Kennedy and O'Hagan 2000). Later, based on the autoregressive model proposed by Kennedy and O'Hagan, Forrester et al. went into more detail about the hyper-parameter estimation of cokriging from the sampling point of view (Forrester et al. 2007). To reduce the

numerical complexity and computational expense in estimating the model hyper-parameters of the method proposed by Kennedy and O'Hagan, Le Gratiet adopted the "Jeffreys priors" and Bayesian method to obtain the closed-form estimations of more model hyper-parameters firstly and estimate the rest using the maximum likelihood estimation method group by group at the sacrifice of the prediction accuracy, in which only the correlation among the observed data from the same model was exploited (Le Gratiet 2013). Overall, although the existing multi-fidelity cokriging modeling approaches utilize the data from all models for response prediction, most of those methods only consider the correlation of the data from two models with adjacent fidelities or from the same model in estimating the unknown hyper-parameters using the maximum likelihood estimation method. Since the lower-fidelity models generally yield a similar trend of the highest-fidelity response, all models with different fidelities actually are correlated to each other to some extent. If the correlation can be fully utilized for hyper-parameter estimation, the accuracy of cokriging metamodel may be improved. Therefore, a unified maximum likelihood function is established accordingly for hyper-parameter estimation in an all-in-one manner during cokriging modeling by utilizing the full correlation of data among all models in this work.

The other important issue associated with the cokriging based sequential optimization is the sequential sampling strategy. As a more efficient one, the objective-oriented approach that brings the design objective into account has been widely studied. The most popular objective-oriented sequential sampling criterion was developed by Jones et al. in their efficient global optimization (EGO) algorithm (Jones et al. 1998), which employs an expected improvement (EI) function to balance the trade-off between the global exploration for reducing the metamodel uncertainty and the local exploitation for searching the optimum. Further, Sasena investigated and compared different objective-oriented sequential sampling criteria to shown their relative merits (Sasena 2002). Xiong et al. (2008) developed a bi-fidelity optimization approach, in which the statistical lower bounding criterion is employed for sequential sampling. However, these methods merely adopt the high-fidelity simulations to construct or update the metamodels. In some region of design space, if the accuracy improvement of the updated metamodel using samples respectively from higher- and lower-fidelity models is relatively close, the lower-fidelity samples are preferred since it is computationally cheaper. Thus, Huang et al. (2006) developed a kriging based sequential optimization scheme using multiple-fidelity responses, in which an integrated objective-oriented sequential sampling criterion was proposed to determine both the location and fidelity level of the subsequent new evaluation to reduce the computational cost. However, during our study, it is found that the samples generated by this criterion

may cluster within a certain region, which clearly will cause a waste of samples and thus reduce the efficiency of optimization. Therefore, an extended sequential sampling strategy is established by introducing a sample density function into the one proposed by Huang et al., which can further reduce the computational cost through more reasonably identifying the location and the fidelity level of the new evaluation.

The contributions of this work mainly lie in: 1) based on the autoregressive model proposed by Kennedy and O’Hagan, establishing a unified maximum likelihood function for hyper-parameter estimation to investigate the impact of utilizing full correlation of all observed data on the accuracy of cokriging metamodel during modeling; 2) developing an extended objective-oriented sequential sampling strategy with the consideration of the sample cluster issue to reduce the computational cost of the cokriging based sequential optimization. The remainder of the paper is organized as follows. In Section 2, the basic ideas of the cokriging technique for hierarchical multi-fidelity modeling is briefly reviewed. In Section 3, the proposed cokriging based sequential optimization method is presented in detail along with the hyper-parameter estimation method and the objective-oriented sampling strategy. In Section 4, comparative studies are conducted to verify the effectiveness of the proposed method. In Section 5, an airfoil shape multi-fidelity optimization problem is solved to further demonstrate the effectiveness of the proposed method. Conclusions are drawn in Section 6.

2 Review on hierarchical multi-level cokriging modeling

The classic mathematical formula of hierarchical multi-level cokriging metamodel proposed by Kennedy and O’Hagan (2000) based on Gaussian process (GP) considering s levels of models $y^1(\mathbf{x}), \dots, y^s(\mathbf{x})$ is represented as

$$y^t(\mathbf{x}) = \rho_{t-1}y^{t-1}(\mathbf{x}) + \delta^t(\mathbf{x}), t = 2, \dots, s \tag{1}$$

where $\mathbf{x} = [x_1, x_2, \dots, x_p] \in \mathbb{R}^p$ is a p -dimensional set of input variables, and $y^t(\mathbf{x}): \mathbb{R}^p \rightarrow \mathbb{R}$ ($t = 1, \dots, s$) denotes the model response at the t^{th} level of fidelity.

A larger t corresponds to a higher level of fidelity, and thus $y^1(\mathbf{x})$ is the lowest-fidelity model and $y^s(\mathbf{x})$ is the highest one. The parameter ρ_{t-1} represents the scaling factor, capturing the strength of the relationship between $y^t(\mathbf{x})$ and $y^{t-1}(\mathbf{x})$; $\delta^t(\mathbf{x})$ represents the discrepancy between $y^t(\mathbf{x})$ and $\rho_{t-1}y^{t-1}(\mathbf{x})$, and it is considered to be independent of $y^{t-1}(\mathbf{x}), \dots, y^1(\mathbf{x})$. The dependence assumed in our work refers to the dependence of the priori of δ and $y^{t-1}(\mathbf{x}), \dots, y^1(\mathbf{x})$ before the collection of data. The additivity of GPs should be utilized during the modeling of the proposed cokriging method, of which the

premise is that the two GPs added together should be independent to each other.

It is assumed that the model output response is observed without random measurement error. For the t^{th} level ($t = 1, \dots, s$) model, it is supposed that a set of response observations $\mathbf{d}_t = [y^t(\mathbf{x}_1^t), \dots, y^t(\mathbf{x}_{n_t}^t)]^T$ at input sites $\mathbf{D}_t = [(\mathbf{x}_1^t)^T, (\mathbf{x}_2^t)^T, \dots, (\mathbf{x}_{n_t}^t)^T]^T$ have been collected. Let $\mathbf{d} = [\mathbf{d}_1^T, \dots, \mathbf{d}_s^T]^T$ denotes all of the collected response data from all models at the input space $\Gamma = [\mathbf{D}_1; \mathbf{D}_2; \dots; \mathbf{D}_s]$. Furthermore, the correlation matrix between response observations taken at points in \mathbf{D}_k and \mathbf{D}_l ($\mathbf{D}_k \subseteq \Gamma, \mathbf{D}_l \subseteq \Gamma$) is denoted by $R_t(\mathbf{D}_k, \mathbf{D}_l)$; if $k = l$, $R_t(\mathbf{D}_k, \mathbf{D}_k)$ is denoted as $R_t(\mathbf{D}_k)$ ($\mathbf{D}_k \subseteq \Gamma$). One popular choice for $R_t(\mathbf{x}, \mathbf{x}')$ in the computer experiment literature is the Gaussian correlation function

$$R_t(\mathbf{x}, \mathbf{x}') = r_t(\mathbf{x} - \mathbf{x}', \boldsymbol{\theta}^t) = \exp\left\{-\sum_{k=1}^p \theta_k^t (x_k - x'_k)^2\right\} \tag{2}$$

where $\boldsymbol{\theta}^t = [\theta_1^t, \theta_2^t, \dots, \theta_p^t]^T$ is a vector of spatial correlation parameters (aka scale or roughness parameters) with respect to each dimension of the input variable \mathbf{x} . It is used to represent the rate at which the correlation between $y^t(\mathbf{x})$ and $y^t(\mathbf{x}')$ decays to zero as \mathbf{x} and \mathbf{x}' diverge.

In the cokriging modeling, the GP theory is applied to build metamodels for the functional response of interest. Based on various assumptions, a GP model in a classical form can be built to represent the highest-fidelity output response $y^s(\mathbf{x})$.

$$y^s(\mathbf{x}) \sim \mathcal{GP}\left(m(\mathbf{x}), V(\mathbf{x}, \mathbf{x}')\right) \tag{3}$$

where $m(\mathbf{x})$ is the *mean function* usually expressed as $h(\mathbf{x})^T \boldsymbol{\beta}$, with $h(\mathbf{x})^T$ denoting a column vector of pre-specified polynomial functions (*i.e.*, constant, linear, quadratic, etc.), and $\boldsymbol{\beta}$ a column vector of the to-be-determined regression coefficients; $V(\mathbf{x}, \mathbf{x}') = \sigma^2 R(\mathbf{x}, \mathbf{x}') = \sigma^2 r(\mathbf{x} - \mathbf{x}', \boldsymbol{\theta})$ is the *covariance function*, representing the spatial covariance between any two inputs \mathbf{x} and \mathbf{x}' of the process.

The unknown parameters $\boldsymbol{\phi} = \{\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}, \boldsymbol{\rho}\}$, referred to as hyper-parameters, characterizes the GP model, where the regression coefficients are $\boldsymbol{\beta} = (\beta_1, \dots, \beta_s)^T$, the scaling factors $\boldsymbol{\rho} = (\rho_1, \dots, \rho_{s-1})^T$, $\sigma^2 = (\sigma_1^2, \dots, \sigma_s^2)^T$ and the spatial correlation parameters $\boldsymbol{\theta} = (\theta_1, \dots, \theta_s)^T$. Once $\boldsymbol{\phi}$ are estimated, the s -level cokriging model is established, of which the response prediction and the corresponding mean square error (MSE) at any new input sites \mathbf{x}_p ($\mathbf{x}_p = [\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_{n_p}^T]^T$) can be predicted. Since the same structure of cokriging model is employed for the proposed FC method as that of the one proposed by Kennedy and O’Hagan, the formulations of the

involved matrixes and vectors (such as $h(\mathbf{x})^T$ and $V(\mathbf{x}, \mathbf{x}')$) as well as the response prediction and MSE above are the same as that of FC. To introduce the proposed method more smoothly and clearly, these formulations are presented in Subsection 3.1.

3 The proposed cokriging-based sequential optimization procedure

Although the existing multi-fidelity cokriging modeling approaches utilize the data from all models for response prediction, most of those methods only consider the correlation of the data from two models with adjacent fidelities or from the same model in estimating the unknown hyper-parameters. Therefore, a method for hyper-parameters estimation is proposed by utilizing full covariance of all the data, which is denoted as FC. Meanwhile, to more reasonably identifying the location as well as the fidelity level of the new evaluation, an extended EI (denoted as EEI) criterion is developed to further reduce the computational cost of sequential optimization.

The outline of the proposed sequential optimization approach employing FC in conjunction with EEI is provided in Fig. 1 and described as below. It follows the general procedure of the existing kriging based sequential optimization proposed by Huang et al. (2006), while two improvements are made (see the two shaded boxes). The step-by-step description of the proposed cokriging based sequential optimization approach is given as follows.

- Step 1:** Generate certain initial input samples for each model with varying fidelity.
- Step 2:** Calculate the response data of the multi-hierarchical lower-fidelity models and the single high-fidelity model.
- Step 3:** Based on the current input and output data, construct the cokriging model using the FC method.
- Step 4:** Obtain the global optimum based on the cokriging model.

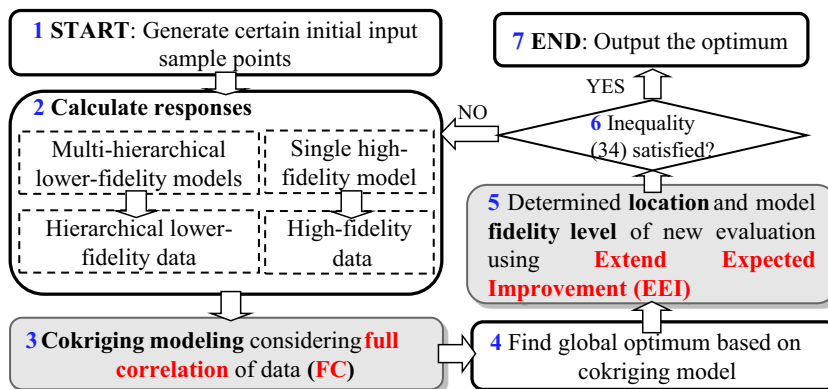
- Step 5:** Determine the input location and the model fidelity level of the next new evaluation by the EEI criterion.
- Step 6:** If the convergence criterion (inequality (34)) is satisfied, go to Step 7, stop the procedure and output the current optimal solutions; if not, go to Step 2, *i.e.* calculate the response data of model with the selected fidelity level at the new input location in Step 5, which is then added to the old data set to update the cokriging model.

3.1 Parameter estimation considering full correlation of data

The full correlation of all the observed data is utilized in this work to obtain the hyper-parameters $\Phi = \{\beta, \sigma^2, \theta, \rho\}$ of the multilevel cokriging model, which is then compared to the methods proposed by Kennedy and O’Hagan (denoted as KOH) (Kennedy and O’Hagan 2000) and Gratiet (denoted as BM) (Le Gratiet 2013). With KOH, the Markov property is assumed and Φ are reckoned numerically by the maximum likelihood estimation (MLE) method in s separate steps, where only the correlation of the observed data between two models with adjacent fidelity (*i.e.*, $y^f(\mathbf{x})$ and $y^{f-1}(\mathbf{x})$) is utilized. For BM, some parameters $\{\rho, \beta, \sigma^2\}$ are estimated based on the “Jeffreys priors” and then θ is obtained by MLE numerically in s separate steps, in which only the correlation of the observed data within the same model (*i.e.*, $y^f(\mathbf{x})$) is explored. However, for FC, the correlation of the observed data among all the models ($y^1(\mathbf{x}), \dots, y^s(\mathbf{x})$) is fully explored to estimate all hyper-parameters altogether by maximizing one unified likelihood function. For the detailed introduction of KOH and BM, readers can refer to (Kennedy and O’Hagan 2000; Le Gratiet 2013), respectively.

With FC, the high-fidelity model $y^s(\mathbf{x})$ can be expressed as below based on the general formulation of cokriging model in (1).

Fig. 1 Flowchart of the proposed cokriging based sequential optimization approach



$$y^s(\mathbf{x}) = \left(\prod_{i=1}^{s-1} \rho_i\right) y^1(\mathbf{x}) + \left(\prod_{i=2}^{s-1} \rho_i\right) \delta^2(\mathbf{x}) + \dots + \rho_{s-1} \delta^{s-1}(\mathbf{x}) + \delta^s(\mathbf{x}) \tag{4}$$

It is assumed that $y^1(\mathbf{x}), \delta^2(\mathbf{x}), \dots, \delta^s(\mathbf{x})$ can be respectively modeled as a GP as below.

$$y^1(\cdot) \sim \mathcal{GP}\left(h^1(\cdot)^T \beta_1, V_d^1\right) \\ \delta^t(\cdot) \sim \mathcal{GP}\left(h^t(\cdot)^T \beta_t, V_d^t\right), t = 2, \dots, s \tag{5}$$

Then, additively the model $y^s(\mathbf{x})$ can be expressed by a GP based on (4) and (5).

$$y^s(\cdot) \sim \mathcal{GP}\left(h^1(\cdot)^T \beta_1 \left(\prod_{i=1}^{s-1} \rho_i\right) + h^2(\cdot)^T \beta_2 \left(\prod_{i=2}^{s-1} \rho_i\right) + \dots + h^s(\cdot)^T \beta_s, \left(\prod_{i=1}^{s-1} \rho_i\right)^2 V_d^1 + \left(\prod_{i=2}^{s-1} \rho_i\right)^2 V_d^2 + \dots + V_d^s\right) \tag{6}$$

Based on the assumption of GP, all the collected data follow a multivariate normal distribution, *i.e.*

$$\mathbf{d} \sim \mathcal{N}(\mathbf{H}\boldsymbol{\beta}, \mathbf{V}_d) \tag{7}$$

where the involved matrixes and vectors are formulated as below.

The matrix \mathbf{H} of size $(n_1 + \dots + n_s) \times s$ is

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_1(\mathbf{D}_1) & 0 & \dots & 0 \\ \rho_1 \mathbf{H}_1(\mathbf{D}_2) & \mathbf{H}_2(\mathbf{D}_2) & 0 & 0 \\ \rho_1 \rho_2 \mathbf{H}_1(\mathbf{D}_3) & \rho_2 \mathbf{H}_2(\mathbf{D}_3) & \mathbf{H}_3(\mathbf{D}_3) & \vdots \\ \vdots & \vdots & \vdots & 0 \\ \left(\prod_{i=1}^{s-1} \rho_i\right) \mathbf{H}_1(\mathbf{D}_s) & \left(\prod_{i=2}^{s-1} \rho_i\right) \mathbf{H}_2(\mathbf{D}_s) & \dots & \mathbf{H}_s(\mathbf{D}_s) \end{pmatrix} \tag{8}$$

with the vector $\mathbf{H}_t(\mathbf{D}_t) = [h^t(\mathbf{x}_1^t), h^t(\mathbf{x}_2^t), \dots, h^t(\mathbf{x}_{n_t}^t)]^T$, where $h^t(\mathbf{x})^T$ denotes a column vector of pre-specified polynomial functions ($h^t(\mathbf{x})^T = 1$ is adopted in this paper).

The vector $\boldsymbol{\beta}$ of size $s \times 1$ is given as

$$\boldsymbol{\beta} = (\beta_1, \dots, \beta_s)^T \tag{9}$$

The matrix \mathbf{V}_d of size $(n_1 + \dots + n_s) \times (n_1 + \dots + n_s)$ is given as

$$\mathbf{V}_d = \begin{pmatrix} V_{1,1} & \dots & V_{1,s} \\ \vdots & \ddots & \vdots \\ V_{s,1} & \dots & V_{s,s} \end{pmatrix}, \tag{10}$$

where the t^{th} diagonal block ($n_t \times n_t$) is defined as

$$V_{t,t} = \sigma_t^2 R_t(\mathbf{D}_t) + \sigma_{t-1}^2 \rho_{t-1}^2 R_{t-1}(\mathbf{D}_t) + \dots + \sigma_1^2 \left(\prod_{i=1}^{t-1} \rho_i^2\right) R_1(\mathbf{D}_t), \tag{11}$$

and the off-diagonal block of size $n_t \times n_{t'}$ is given by

$$V_{t,t'} = \left(\prod_{i=t}^{t'-1} \rho_i\right) \left(\sigma_t^2 R_t(\mathbf{D}_t, \mathbf{D}_{t'}) + \dots + \sigma_1^2 \left(\prod_{i=1}^{t-1} \rho_i^2\right) R_1(\mathbf{D}_t, \mathbf{D}_{t'})\right), \tag{12} \\ 1 \leq t < t' \leq s.$$

Then the MLE method is employed to estimate the hyper-parameters $\boldsymbol{\Phi} = \{\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}, \boldsymbol{\rho}\}$. Given the collected response data set \mathbf{d} , the unified likelihood function of the hyper-parameters $\boldsymbol{\Phi}$ is constructed as

$$\mathcal{L}(\boldsymbol{\Phi}|\mathbf{d}) \propto |\mathbf{V}_d|^{-1/2} |\mathbf{W}|^{1/2} \exp\left\{-\frac{1}{2}(\mathbf{d}-\mathbf{H}\boldsymbol{\beta})^T \mathbf{V}_d^{-1}(\mathbf{d}-\mathbf{H}\boldsymbol{\beta})\right\} \tag{13}$$

where $\mathbf{W} = (\mathbf{H}^T \mathbf{V}_d^{-1} \mathbf{H})^{-1}$.

Solving the maximum unified likelihood function in (13) is equivalent to maximizing its logarithm (*i.e.*, log-likelihood function) as below.

$$\max_{\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}, \boldsymbol{\rho}} \mathcal{L}(\boldsymbol{\Phi}|\mathbf{d}) = \max_{\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}, \boldsymbol{\rho}} \log[\mathcal{L}(\boldsymbol{\Phi}|\mathbf{d})] \\ = \max_{\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}, \boldsymbol{\rho}} \ell(\boldsymbol{\Phi}|\mathbf{d}) \tag{14}$$

where $\ell(\boldsymbol{\Phi}|\mathbf{d}) \propto \frac{1}{2} \log|\mathbf{W}| - \frac{1}{2} \log|\mathbf{V}_d| - \frac{1}{2}(\mathbf{d}-\mathbf{H}\boldsymbol{\beta})^T \mathbf{V}_d^{-1}(\mathbf{d}-\mathbf{H}\boldsymbol{\beta})$.

According to the first order optimality condition

$$\left. \frac{\partial \ell(\boldsymbol{\Phi}|\mathbf{d})}{\partial \boldsymbol{\beta}} \right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = 0, \tag{15} \\ (\mathbf{d}-\mathbf{H}\hat{\boldsymbol{\beta}})^T \mathbf{V}_d^{-1} \mathbf{H} = \mathbf{d}^T \mathbf{V}_d^{-1} \mathbf{H} - \hat{\boldsymbol{\beta}}^T \mathbf{H}^T \mathbf{V}_d^{-1} \mathbf{H} = 0$$

the estimation in the closed-form expression $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$ can be obtained.

$$\hat{\boldsymbol{\beta}} = \mathbf{W} \mathbf{H} \mathbf{V}_d^{-1} \mathbf{d} \tag{16}$$

which is expressed in a function of σ^2 and $\boldsymbol{\theta}$.

The rest of hyper-parameters σ^2 , $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$ can be estimated by solving the maximum unified log-likelihood function below using numerical optimization algorithm, *e.g.*, a genetic algorithm or simulated annealing.

$$\max_{\sigma^2, \boldsymbol{\theta}, \boldsymbol{\rho}} \frac{1}{2} \log|\mathbf{W}| - \frac{1}{2} \log|\mathbf{V}_d| - \frac{1}{2}(\mathbf{d}-\mathbf{H}\hat{\boldsymbol{\beta}})^T \mathbf{V}_d^{-1}(\mathbf{d}-\mathbf{H}\hat{\boldsymbol{\beta}}) \tag{17}$$

$\boldsymbol{\beta}$ can then be estimated by using (16). It should be noticed that the hyper-parameters are estimated altogether in the proposed method by solving the unified likelihood function, while they are estimated group by group in s separate steps in KOH and BM methods. Meanwhile, the covariance among all the models is utilized. Once all the hyper-parameters are estimated, the

response prediction $y(\mathbf{x}_p)$ and its corresponding mean square error (MSE) at any new input sites \mathbf{x}_p ($\mathbf{x}_p = [\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_{n_p}^T]^T$) can be predicted. With the GP model, the collected data \mathbf{d} together with the to-be-predicted responses $\mathbf{y}(\mathbf{x}_p) = [y(\mathbf{x}_1), \dots, y(\mathbf{x}_{n_p})]^T$ follow a multivariate normal distribution.

$$\begin{bmatrix} \mathbf{d} \\ \mathbf{y}(\mathbf{x}_p) \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mathbf{H} \\ \mathbf{H}_p \end{bmatrix} \boldsymbol{\beta}, \begin{bmatrix} \mathbf{V}_d & \mathbf{T}_p^T \\ \mathbf{T}_p & \mathbf{V}_p \end{bmatrix} \right) \quad (18)$$

where \mathbf{H} and \mathbf{V}_d are the polynomial function and covariance matrix for the observed data \mathbf{d} , which have been formulated in (8) and (10), respectively; \mathbf{H}_p and \mathbf{V}_p are the polynomial function and covariance matrix for the to-be-predicted responses $\mathbf{y}(\mathbf{x}_p)$; \mathbf{T}_p denotes the covariance matrix between $\mathbf{y}(\mathbf{x}_p)$ and \mathbf{d} . Details regarding the matrices \mathbf{H}_p , \mathbf{V}_p , and \mathbf{T}_p will be given in (21)–(25) below. Then the final prediction of $\mathbf{y}(\mathbf{x}_p)$ and its MSE can be computed by

$$\hat{\mathbf{y}}(\mathbf{x}_p) = \mathbf{H}_p \hat{\boldsymbol{\beta}} + \mathbf{T}_p \mathbf{V}_d^{-1} (\mathbf{d} - \mathbf{H} \hat{\boldsymbol{\beta}}), \quad (19)$$

$$\begin{aligned} \text{MSE}[\hat{\mathbf{y}}(\mathbf{x}_p)] &= \mathbf{V}_p - \mathbf{T}_p \mathbf{V}_d^{-1} \mathbf{T}_p^T \\ &\quad + \left(\mathbf{H}_p^T - \mathbf{H}^T \mathbf{V}_d^{-1} \mathbf{T}_p^T \right)^T \mathbf{W} \left(\mathbf{H}_p^T - \mathbf{H}^T \mathbf{V}_d^{-1} \mathbf{T}_p^T \right), \end{aligned} \quad (20)$$

where

$$\begin{aligned} \mathbf{H}_p &= \left(\left(\prod_{i=1}^{s-1} \rho_i \right) h^1(\mathbf{x})^T, \left(\prod_{i=2}^{s-1} \rho_i \right) h^2(\mathbf{x})^T, \dots, \rho_{s-1}(x) h^{s-1}(\mathbf{x})^T, h^s(\mathbf{x})^T \right)^T \quad (21) \\ &\quad \times (\text{size } 1 \times s) \end{aligned}$$

$$\mathbf{T}_p = \left(t_1(\mathbf{x}_p, \mathbf{D}_1)^T, \dots, t_s(\mathbf{x}_p, \mathbf{D}_s)^T \right) (\text{size } n_p \times s) \quad (22)$$

$$t_1(\mathbf{x}_p, \mathbf{D}_1)^T = \left(\prod_{i=1}^{s-1} \rho_i \right) \sigma_1^2 R_1(\mathbf{x}_p, \mathbf{D}_1) \quad (23)$$

$$\begin{aligned} t_t(\mathbf{x}_p, \mathbf{D}_t)^T &= \rho_{t-1} t_t(\mathbf{x}_p, \mathbf{D}_t)^T + \left(\prod_{i=t}^s \rho_i \right) \sigma_t^2 R_t(\mathbf{x}_p, \mathbf{D}_t), \quad (24) \\ t &= 2, \dots, s \end{aligned}$$

$$\begin{aligned} \mathbf{V}_p &= \mathbf{V}_d^s(\mathbf{x}_p, \mathbf{x}_p) + \rho_{s-1}^2 \mathbf{V}_d^{s-1}(\mathbf{x}_p, \mathbf{x}_p) + \rho_{s-1}^2 \rho_{s-2}^2 \mathbf{V}_d^{s-2}(\mathbf{x}_p, \mathbf{x}_p) + \dots \\ &\quad + \left(\prod_{t=2}^{s-1} \rho_t \right)^2 \mathbf{V}_d^2(\mathbf{x}_p, \mathbf{x}_p) (\text{size } n_p \times n_p) \end{aligned} \quad (25)$$

Meanwhile, it should be pointed out that in KOH and BM, the input sets are assumed to be nested, *i.e.*, $\mathbf{D}_s \subseteq \dots \subseteq \mathbf{D}_1$, which is not necessary but allows to have a closed-form expression for the parameter posterior distribution. In KOH, it is necessary to calculate $\delta_t(x) = \mathbf{d}_{t+1} - \rho_t y^t(\mathbf{D}_{t+1})$ to define the

likelihood function in estimating the parameters $(\boldsymbol{\theta}_t, \sigma_t^2, \rho_{t-1})$. Meanwhile, for BM, $G_{t+1} = [y^t(\mathbf{D}_{t+1}) \ \mathbf{H}_{t+1}(\mathbf{D}_{t+1})]$ should be defined in the “Jeffreys priors” to obtain the estimations of parameters $(\rho_{t-1}, \boldsymbol{\beta}_t, \sigma_t^2, \boldsymbol{\theta}_t)$. Clearly, $y^t(\mathbf{D}_{t+1})$ exists only if the sample points for y^{t+1} and y^t are nested, *i.e.*, $\mathbf{D}_{t+1} \subseteq \mathbf{D}_t$. If not, an approximated GP model for each lower-fidelity model should be constructed firstly, based on which the multi-level cokriging model can also be constructed. However, for the FC, it is not restricted by this assumption, and thus it is more flexible.

In addition, in the FC method, the maximum likelihood function contains a matrix \mathbf{V}_d , of which the size $(n_1 + \dots + n_s) \times (n_1 + \dots + n_s)$ is determined by the number of models for fusion and the number of sample points from each model. If the size of \mathbf{V}_d is large, the computational cost in solving the maximum likelihood function will be clearly increased. Meanwhile, during the process of MLE, some garbage results may be generated. In this case, some additional repeated optimization can be conducted to help find the global optimum. For BM and KOH, since all items in their maximum likelihood function are scalars or vectors, the computational cost is clearly smaller than FC. However, compared with the computational cost of model response evaluation (in practice, it often takes a very long time to run one model evaluation), such increase in the computational time of model fusion can be negligible. The proposed FC method aims to increase the accuracy of the final cokriging model, which is still of great significance to some engineering problems.

3.2 The extended sequential sampling strategy

Several criteria have been investigated to decide a new infilling sample site for the objective-oriented sequential sampling. The statistical lower bounding (SLB) criterion and the expected improvement (EI) criterion are the two most widely used approaches in the literature. Based on the results of Jones et al.’s work (Jones et al. 1998), an extended EI (EEI) criterion is defined to determine both the location of the new input sample and the fidelity level of the new evaluation.

$$\text{EEI}(\mathbf{x}, t) = \text{EI}(\mathbf{x}) \times |\text{Corr}(\mathbf{x}, t)| \times \text{CR}(t) \times \eta(\mathbf{x}, t) \quad (26)$$

where $\text{EI}(\mathbf{x})$ is the original EI function, and the other three items will be described as below.

(1) Model correlation $\text{Corr}(\mathbf{x}, t)$

$\text{Corr}(\mathbf{x}, t)$ represents the posterior correlation coefficient between the stochastic predicted responses $y_{po}^t(\mathbf{x})$ from the t -level fidelity model and $y_{po}^s(\mathbf{x})$ from the s -level (highest) fidelity one at the input site \mathbf{x} , and can be computed by

$$\begin{aligned} \text{Corr}(\mathbf{x}, t) &= \text{Corr}\left[y_{po}^t(\mathbf{x}), y_{po}^s(\mathbf{x})\right] \\ &= \frac{\text{Cov}\left[y_{po}^t(\mathbf{x}), y_{po}^s(\mathbf{x})\right]}{\sqrt{\text{Var}\left[y_{po}^t(\mathbf{x})\right] \times \text{Var}\left[y_{po}^s(\mathbf{x})\right]}}. \end{aligned} \tag{27}$$

In (27), $\text{Cov}\left[y_{po}^t(\mathbf{x}), y_{po}^s(\mathbf{x})\right]$ denotes the posterior covariance at the input site \mathbf{x} between the stochastic predicted responses $y_{po}^t(\mathbf{x})$ and $y_{po}^s(\mathbf{x})$, and it can be obtained by

$$\begin{aligned} \text{Cov}\left[y^t(\mathbf{x}), y^s(\mathbf{x})\right] &= \text{Cov}\left[y^t(\mathbf{x}), y^t(\mathbf{x})\rho + \delta(\mathbf{x})\right] \\ &= \text{Cov}\left[y^t(\mathbf{x}), y^t(\mathbf{x})\rho\right] + \underbrace{\text{Cov}\left[y^t(\mathbf{x}), \delta(\mathbf{x})\right]}_{=0} = (\sigma^t)^2 \rho R_t(\mathbf{x}, \mathbf{x}) \end{aligned} \tag{29}$$

$$\begin{aligned} \text{Cov}\left[y_{po}^t(\mathbf{x}), y_{po}^s(\mathbf{x})\right] &= \text{Cov}\left[y^t(\mathbf{x}), y^s(\mathbf{x})\right] - \mathbf{T}_p^t \mathbf{V}_d^{-1} \left(\mathbf{T}_p^s\right)^T \\ &\quad + \left(\left(\mathbf{H}_p^t\right)^T - \mathbf{H}^T \mathbf{V}_d^{-1} \left(\mathbf{T}_p^t\right)^T\right) \mathbf{w} \left(\left(\mathbf{H}_p^s\right)^T - \mathbf{H}^T \mathbf{V}_d^{-1} \left(\mathbf{T}_p^s\right)^T\right)^T, \end{aligned} \tag{28}$$

where

In the above equation, ρ denotes the weight coefficient between $y^t(\mathbf{x})$ and $y^s(\mathbf{x})$, which is a scalar calculated as $\rho = \rho_t \rho_{t+1} \dots \rho_s$. Meanwhile, in (27), $\text{Var}\left[y_{po}^s(\mathbf{x})\right] = \text{Cov}\left[y_{po}^s(\mathbf{x}), y_{po}^s(\mathbf{x})\right]$ and $\text{Var}\left[y_{po}^t(\mathbf{x})\right] = \text{Cov}\left[y_{po}^t(\mathbf{x}), y_{po}^t(\mathbf{x})\right]$ can also be calculated by (28).

If $t = s$, for any \mathbf{x} , $\text{Corr}(\mathbf{x}, t) = 1$, which stands for the correlation of the stochastic predicted responses from the highest-fidelity model itself. If $t \neq s$, for any \mathbf{x} , $|\text{Corr}(\mathbf{x}, t)| \in [0, 1]$, which describes the ratio between the accuracy improvement for the predicted response in adding the lower-fidelity response data $y^t(\mathbf{x})$ and highest-fidelity one $y^s(\mathbf{x})$. It should be noticed that for the input set that already has the corresponding model response data, $\text{Var}\left[y_{po}^t(\mathbf{x})\right] \times \text{Var}\left[y_{po}^s(\mathbf{x})\right]$ and $\text{Cov}\left[y^t(\mathbf{x}), y^s(\mathbf{x})\right]$ are all equal to 0. In this case, $\text{Corr}(\mathbf{x}, t)$ is defined as 0 in our study.

(2) Ratio of cost $\text{CR}(t)$

$\text{CR}(t)$ represents the computational cost ratio of a single response evaluation between the highest-fidelity model $y^s(\mathbf{x})$ and the t -level one $y^t(\mathbf{x})$, *i.e.*

$$\text{CR}(t) = \frac{C^s}{C^t} \tag{30}$$

where C^s and C^t respectively represents the computational cost of a single response evaluation for the highest-fidelity model $y^s(\mathbf{x})$ and the t -th level lower-fidelity model $y^t(\mathbf{x})$.

If $t = s$, $\text{CR}(t) = 1$; if $t \neq s$, $\text{CR}(t) \geq 1$, which means that it is more efficient to employ the lower-fidelity model $y^t(\mathbf{x})$ to evaluate the response. The inclusion of $\text{CR}(t)$ in (26) aims to select the computationally cheapest model as much as possible that is then called to evaluate the response at the identified new input location, considering the trade-off from other two aspects (model correction and sample density). That is to say, if similar accuracy improvement and sample density property can be obtained, the model with the smallest cost will be selected to reduce the computational cost as much as possible.

(3) Sample density function $\eta(\mathbf{x}, t)$

$\eta(\mathbf{x}, t)$ describes the density of the input samples from the same model, which is used to avoid the clustering of samples. The spatial correlation of the GP is employed to quantify the distance between the input samples, and thus $\eta(\mathbf{x}, t)$ is defined as

$$\eta(\mathbf{x}, t) = \prod_{i=1}^{N^t} [1 - R(\mathbf{x}, \mathbf{x}^i)] \tag{31}$$

where N^t represents the number of samples from t -th level fidelity model, while \mathbf{x}^i are the locations of input samples; $R(\mathbf{x}, \mathbf{x}^i)$ is the spatial correlation function of GP, which is the same as that used in the cokriging modeling above.

Moreover, if higher-fidelity response data exist at a certain input location, adding new lower-fidelity response data at this input location generally contributes minimally to the improvement of the accuracy of the response prediction. Therefore, based on (31), the sample density factor function $\eta(\mathbf{x}, t)$ is redefined as below to avoid unnecessary waste of samples.

$$\eta(\mathbf{x}, t) = \prod_{N^t}^{N^s} \prod_{i=1}^{N^t} [1 - R(\mathbf{x}, \mathbf{x}^i)], \forall t \leq s \tag{32}$$

Clearly, each item in (26) plays an important role. As the basis of the EEI criterion, the original $\text{EI}(\mathbf{x})$ quantifies the improvement of response prediction accuracy at any input location. If $\text{Corr}(\mathbf{x}, t)$ is removed, the model with the lowest fidelity will always be selected for response evaluation since it has the smallest cost. If $\text{CR}(t)$ is removed, the high-fidelity model will always be selected for response evaluation since it can obtain the maximum improvement of response prediction accuracy. If $\eta(\mathbf{x}, t)$ is removed, samples may get clustered, decreasing the optimization efficiency. In addition, if only a single model is used for response prediction, EEI will be automatically degenerated to the original EI function.

Note that apart from the EEI criterion defined in (26), other formulations certainly can be adopted to balance among the model correlation, computational cost, and the sample density

Table 1 Function descriptions

No.	Functions
1	$y^3(x) = \sin(x) + 0.2x + (x-0.5)^2/16 + 0.5$ $y^2(x) = (x-0.5)(x-4)(x-9)/20 + 2$ $y^1(x) = \sin(x) + 0.2x + 0.5$ $x \in [0, 10]$
2	$y^3(x) = 2\sin(\pi x/5)$ $y^2(x) = x(x-5)(x-12)/30$ $y^1(x) = (x+2)(x-5)(x-12)/30$ $x \in [0, 5]$
3	$y^3 = \left(1 - e^{-\frac{1}{2x_2}}\right) \left(\frac{1000tf \times x_1^3 + 1900x_1^2 + 2092x_1 + 60}{1000tl \times x_1^3 + 500x_1^2 + 4x_1 + 20}\right)$ $y^2 = \left(1 - e^{-\frac{1}{2x_2}}\right) \left(\frac{1000tf \times x_1^3 + 1900x_1^2 + 2092x_1 + 60}{1000tl \times x_1^3 + 500x_1^2 + 4x_1 + 20}\right) + \frac{5e^{-tf} x_1^{th/2}}{x_2^{2+th} + 1}$ $y^1 = \left(1 - e^{-\frac{1}{2x_2}}\right) \left(\frac{1000tf \times x_1^3 + 1900x_1^2 + 2092x_1 + 60}{1000tl \times x_1^3 + 500x_1^2 + 4x_1 + 20}\right) + \frac{5e^{-tf} x_1^{th/2}}{x_2^{2+th} + 1} + \frac{10x_1^2 + 4x_2^2}{50x_1x_2 + 10}$ $tf = 0.2, th = 0.3, tl = 0.1; x_1, x_2 \in [0, 1]$
4	Aerodynamic modeling of wrap-around fin

property. For example, when their ratios have significantly different orders of magnitude, the small one may not be considered in the EEI criterion. In this case, the designer can analyze this influence via sensitivity studies and propose an effective formulation such as assigning different powers to the three factors based on their orders of magnitude. Different formulations used in the EEI criterion would not influence the procedure of other posterior analysis of the FC method.

Both the location of new input \mathbf{x}_{new} and the corresponding model t_{add} called to evaluate the response at \mathbf{x}_{new} are determined through maximizing the EEI function.

$$(\mathbf{x}_{new}, t_{add}) = \max_{\mathbf{x}, t} EEI(\mathbf{x}, t) \tag{33}$$

The optimization scheme stops when the current obtained optimal EEI objective value is smaller than a specified value γ , *i.e.*

$$\max_{\mathbf{x}, \mathcal{M}} EEI(\mathbf{x}, t) < \gamma = \alpha\% \times [\max(\mathbf{d}) - \min(\mathbf{d})], \tag{34}$$

where γ is the so-called “relative stopping criterion” (Huang et al. 2006), which is defined as $\alpha\%$ of the “active span” of the collected response data \mathbf{d} ; $\max(\mathbf{d})$ and $\min(\mathbf{d})$ respectively represent the maximum and minimum elements in the current collected response data set \mathbf{d} ; $\max(\mathbf{d}) - \min(\mathbf{d})$ is the “active span” of the responses; $\alpha\%$ is selected as $\alpha\% = 0.05\%$ in our study, corresponding to 95% confidence interval.

Clearly, \mathbf{x}_{new} and t_{add} found by the proposed EEI criterion can make a good balance among the response accuracy improvement, the computational cost, and the sample density

Table 2 Errors and time costs of FC, KOH and BM for example 1

		FC		KOH		BM	
		RMSE	Time(s)	RMSE	Time(s)	RMSE	Time(s)
Nested	No.1	0.1023	36.02	0.1296	19.45	0.2070	12.02
	No.2	0.1024	38.1	0.1295	19.95	0.2072	16.24
	No.3	0.1022	37.5	0.1297	21.58	0.2075	11.27
	No.4	0.1021	35.5	0.1298	20.46	0.2065	11.08
	No.5	0.1025	34.2	0.1299	21.05	0.2069	12.84
	Mean	0.1023	36.23	0.1297	20.5	0.2070	12.70
Non-nested	No.1	0.1025	38.21	0.1326	18.99	0.2100	13.16
	No.2	0.1024	35.55	0.1325	28.54	0.2098	16.88
	No.3	0.1021	37.5	0.1324	23.44	0.2107	16.5
	No.4	0.1026	39.82	0.1327	24.73	0.2102	16.45
	No.5	0.1027	41.25	0.1328	19.58	0.2103	14.85
	Mean	0.1025	38.466	0.1326	23.06	0.2102	13.57

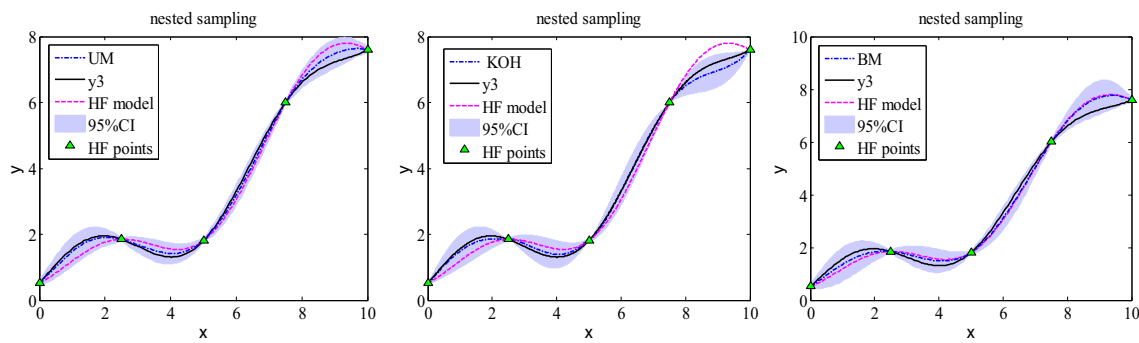


Fig. 2 Predicted responses from different models of example 1 (nested)

property, *i.e.* \mathbf{x}_{new} and t_{add} try to improve the accuracy of the cokriging metamodel as much as possible with as little computational cost as possible.

4 Comparative studies

Three numerical examples and one engineering problem (shown in Table 1) are employed to test the effectiveness of the proposed hyper-parameter estimation method (FC) during cokriging metamodeling, which is compared to KOH and BM. Then, sequential optimization employing FC and EEI is then conducted to obtain the optimal solutions. For all the examples, it is assumed that $s = 3$ levels of fidelity $y^1(\mathbf{x}), y^2(\mathbf{x}), y^3(\mathbf{x})$ exist.

4.1 Validation of FC for hyper-parameter estimation

Firstly, the FC hyper-parameter estimation method is tested. To evaluate the accuracy of the constructed cokriging metamodel, the root-mean-square mean (RSME) error is computed. A smaller RMSE corresponds to a more accurate cokriging metamodel. The RSME is calculated as

$$RSME = \sqrt{\sum(\hat{y}(\mathbf{x}_p) - \mathbf{y}^s(\mathbf{x}_p))^2 / n_p} \tag{35}$$

where \mathbf{x}_p are the test sample points generated randomly and uniformly in the design region of the input \mathbf{x} with the number

as $n_p = 1000$; $\mathbf{y}(\mathbf{x}_p)$ and $\mathbf{y}^s(\mathbf{x}_p)$ are the response predictions at test input set \mathbf{x}_p obtained by the cokriging model and the response model with the highest-fidelity level, respectively.

4.1.1 Numerical example 1

The input sample points for the three models $y^1(\mathbf{x}), y^2(\mathbf{x}), y^3(\mathbf{x})$ are selected with $n_3 = 5, n_2 = 10,$ and $n_1 = 20,$ satisfying the nested property $\mathbf{D}_3 \subseteq \mathbf{D}_2 \subseteq \mathbf{D}_1$. The non-nested input sample points are also generated uniformly with the same size. For both cases in all the examples, several different sets of input sample points are tested repeatedly with the same settings to test the robustness of the FC method, based on which the mean values of RMSE and time cost are calculated. In addition to the accuracy of cokriging, the time cost in estimating the hyper-parameters on a personal computer (Intel Corei7-4710HQ, CPU 2.50GHz) is also computed. The results are shown in Table 2, from which some noteworthy observations can be made. Firstly, for both nested and non-nested sampling strategies, the accuracy of cokriging metamodel constructed by FC is the best, followed by KOH and then BM. Specially, the RMSE of BM is almost twice that of FC. Secondly, for the time cost, FC is the most time-consuming, followed by KOH and then BM. The interpretation is that for FC, the correlation of the data from all the models is computed and fully utilized during the hyper-parameter estimation. Thus the accuracy of FC is improved compared to

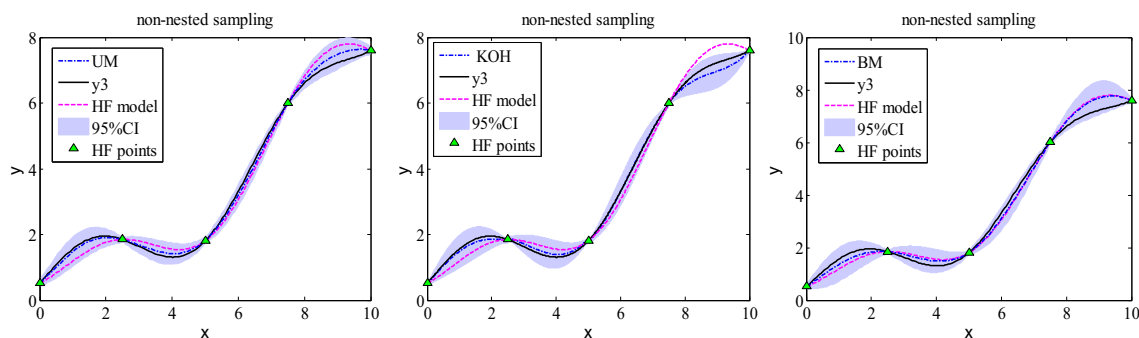


Fig. 3 Predicted responses from different models of example 1 (non-nested)

Table 3 Errors and time costs of the three approaches for example 2

		FC		KOH		BM	
		RMSE	Time(s)	RMSE	Time(s)	RMSE	Time(s)
Nested	No.1	0.0219	46.68	0.0452	21.48	0.1212	17.41
	No.2	0.0218	48.25	0.0449	19.95	0.1218	18.34
	No.3	0.0223	47.5	0.0461	22.55	0.1224	15.64
	No.4	0.0221	47.80	0.0455	20.40	0.1199	17.58
	No.5	0.0220	48.75	0.0454	21.45	0.1198	17.54
	Mean	0.0220	47.80	0.0454	21.17	0.1210	17.30
Non-nested	No.1	0.0216	40.45	0.0514	19.21	0.1217	15.54
	No.2	0.0214	42.58	0.0508	19.95	0.1218	16.8
	No.3	0.0212	45.54	0.0511	18.70	0.1211	17.5
	No.4	0.0218	41.24	0.0515	18.25	0.1223	14.25
	No.5	0.0219	41.28	0.0516	18.3	0.1225	14.5
	Mean	0.0216	42.22	0.0513	18.88	0.1220	15.72

the other two approaches. However, in solving the MLE, a covariance matrix of large size (35×35 for this example) is involved for FC, resulting in an evident increase in computational cost; meanwhile, no large-scale matrix operations are involved for KOH and BM. Thirdly, compared to the results for nested sampling, the errors with the non-nested sampling for KOH and BM are slightly larger; meanwhile for FC, the two sets of results are very close to each other.

The reason is that when the sample points for different fidelities are not nested, a GP model has to be constructed based on the collected data for each lower-fidelity model $y^t(\mathbf{x})$ ($t = 1, \dots, s-1$) for KOH and BM, which apparently would induce certain error. Since this example is relatively simpler, the GP models constructed are generally accurate. So for KOH and BM, the difference of results between nested and non-nested sampling is very small. In addition,

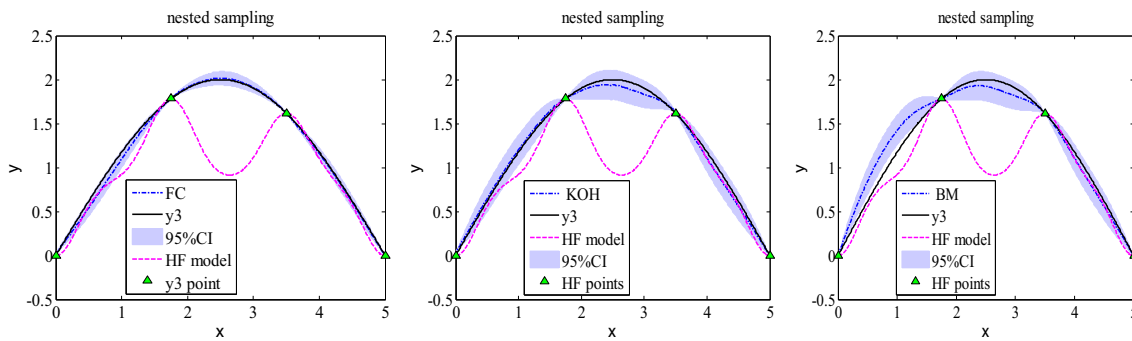


Fig. 4 Predicted responses from different models of example 2 (nested)

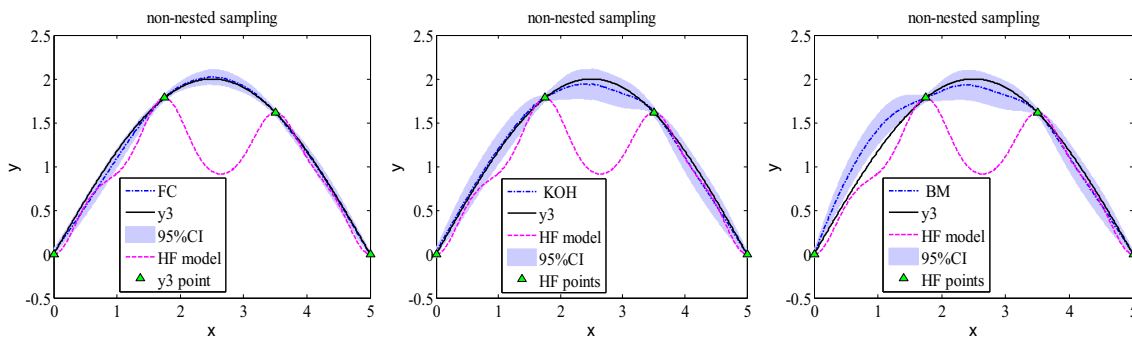


Fig. 5 Predicted responses from different models of example 2 (non-nested)

Table 4 Results of the three approaches for example 3

	FC		KOH		BM	
	RMSE	Time(s)	RMSE	Time(s)	RMSE	Time(s)
No.1	1.3824	59.45	2.4521	28.30	3.0114	24.50
No.2	1.4421	58.25	2.5613	24.95	3.3873	18.34
No.3	1.4762	57.5	2.8241	27.55	3.2448	23.62
No.4	1.4018	57.80	2.1149	28.40	3.2808	24.21
No.5	1.4767	58.75	2.2086	26.45	3.3017	17.54
No.6	1.3762	60.25	2.9548	30.80	3.1207	16.55
No.7	1.4028	54.55	2.8685	30.51	3.3844	16.85
No.8	1.4602	52.62	2.8465	29.55	3.3809	15.24
No.9	1.4967	64.58	2.4558	25.05	3.4208	16.88
No.10	1.4458	59.55	2.5503	24.85	3.4405	18.44
Mean	1.4361	58.33	2.5837	27.64	3.2973	19.22

the variation of RSME is relatively small, indicating the robustness of FC as well as other two methods.

The predicted responses from the constructed cokriging models by FC, KOH and BM with nested and non-nested sample points with one of the five sample sets are shown in Figs. 2 and 3 (see the dot dash line), where the 95% confidence interval (CI) for the prediction of each cokriging model is also plotted. Meanwhile, the true response from the highest-fidelity model (denoted by y^3) and predicted response from the kriging model constructed with only the data from the highest-fidelity model y^3 (denoted by HF) are also shown in Figs. 2 and 3 for comparison.

It is found from these figures that generally the prediction responses of the cokriging models by FC are slightly closer to the true response (y^3) yielding smaller errors, compared to KOH and BM; and both KOH and FC are more accurate than BM. This exhibits great agreement to the results shown in Table 2. Meanwhile, compared to the curves (HF) produced by using only the highest-fidelity data, the cokriging models are much more accurate due to the fusion of the lower-fidelity data.

4.1.2 Numerical example 2

Similarly, nested and non-nested sample points with $n_3 = 4$, $n_2 = 11$, and $n_1 = 21$ are selected to test the effectiveness of the proposed FC for hyper-parameter estimation. The results of the three methods are shown in Table 3.

The constructed cokriging models by FC, KOH and BM, as well as the HF model constructed with only the data from $y^3(\mathbf{x})$ and the true response y^3 are shown in Figs. 4 and 5.

Similar observations can be made from Table 3 and the figures above that FC performs best with smaller errors, more accurate predicted output response curves and smaller metamodel uncertainty, compared to KOH and BM. The results of FC are the most accurate, followed by KOH and BM. For the time cost, FC is the most time-consuming, followed by KOH and then BM.

4.1.3 Numerical example 3

Since it is not easy to directly generate nested sample points as the one-dimensional examples above, only the non-nested samples are tested. For approach that can generate nested sample points, the readers can refer to Ref. (Skilling 2006). Latin Hypercube sampling (LHS) is used to select the input sample points with $n_3 = 8$, $n_2 = 15$ and $n_1 = 20$. The results are shown in Table 4.

The constructed cokriging models by FC, KOH and BM, as well as the model constructed with only the data from $y^3(\mathbf{x})$ (HF) and the true response y^3 are shown in Fig. 6. For the two-dimensional problem, the results exhibit great agreements to what has been observed above that the accuracy of FC is clearly the best, followed by KOH and then BM; and FC is the most time-consuming, followed by KOH and then BM. As Table 4 shows, RMSE of FC is much smaller than the other two approaches. Meanwhile, the response surface produced by FC is clearly much closer to the true response surface y^3 , yielding smaller metamodel uncertainty.

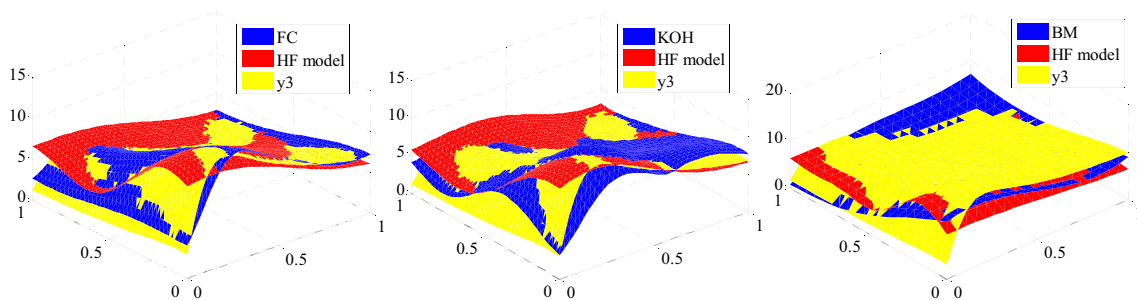
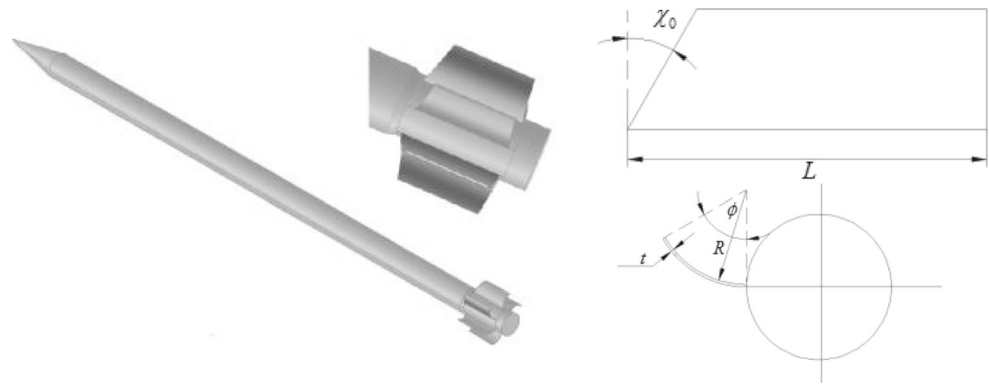


Fig. 6 Predicted responses from different models of example 3

Fig. 7 The geometric profile of the rocket



4.1.4 Aerodynamic modeling of wrap-around fin

In this part, the cokriging method is employed to construct the aerodynamic metamodel for a rocket with wrap-around fins (shown in Fig. 7), under the flight condition with Mach number as $Ma = 4$ and angle of attack as $\alpha = 6^\circ$. The input vector is $\mathbf{x} = [D, \lambda_a, t, R, \Phi, L, \chi_0]$, which is presented in Table 5.

CFD is employed for the aerodynamic analysis. Since this work focuses on investigating the effectiveness of FC in cokriging modeling and it is very time-consuming to run CFD here, we did not dig into setting up various CFD simulation models with different fidelities but directly utilized the CFD data provided in Ref. (Xue 2010) to create models with different fidelities. Generally, as the number of sample points increases, the accuracy of the kriging model is improved. Hence, the models with 3 levels of fidelity $y^1(\mathbf{x})$, $y^2(\mathbf{x})$, $y^3(\mathbf{x})$ are replaced by different kriging models that are respectively constructed based on different numbers of CFD data (10, 20 and 32). Then the LHS method is employed to generate sample points for the three models $y^1(\mathbf{x})$, $y^2(\mathbf{x})$, $y^3(\mathbf{x})$ with $n_1 = 40$, $n_2 = 20$ and $n_3 = 10$. The results are presented in Table 6, from which similar observations can be made that for the higher-dimensional problem, FC performs best with the smallest errors; while for the time cost, FC is the most time-consuming, followed by KOH and then BM.

Table 5 Input vector and the bounds

Symbols	Explanations	LB	UB
D	diameter of the missile body (mm)	125	200
λ_a	slenderness of the missile body	10	30
t	thickness of the wrap-around fins (mm)	5	8
R	radius of curvature of the middle of cambered surface (mm)	125	200
Φ	center angle of the wrap-around fins ($^\circ$)	40	70
L	root-chord (mm)	400	700
χ_0	wing leading edge sweep angle ($^\circ$)	15	30

4.2 Sequential optimization

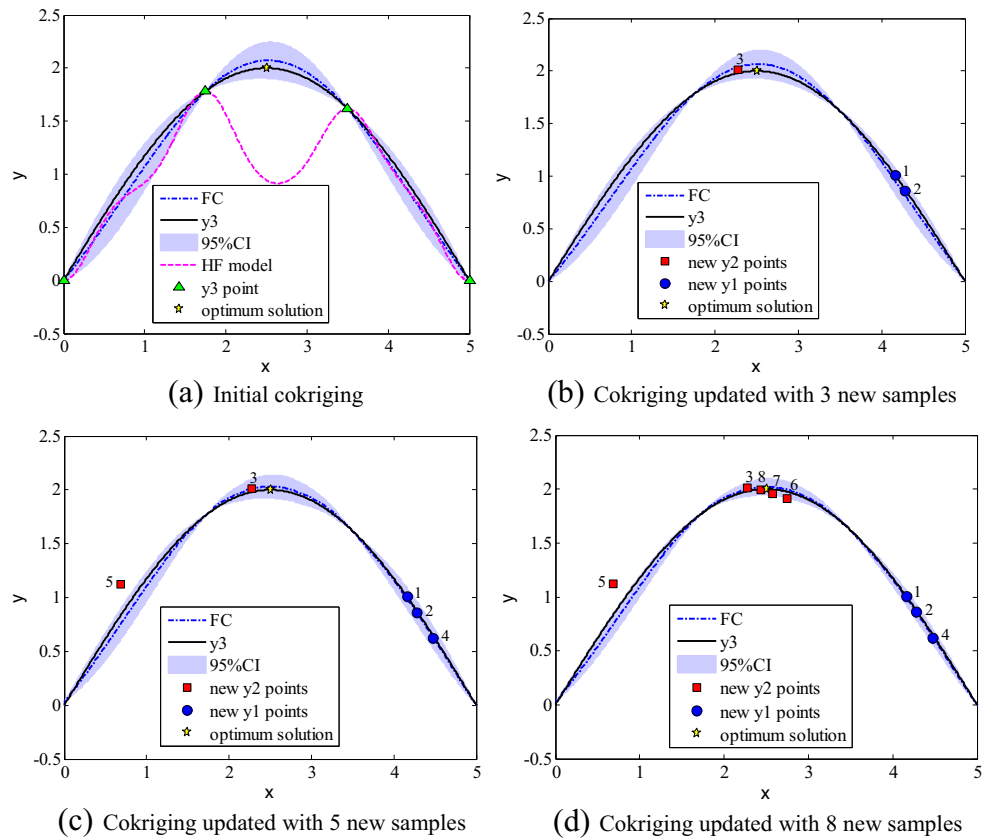
Then, the cokriging based sequential optimization method (named as cokriging-M) employing FC for hyper-parameter estimation and EEI for sequential sampling is applied to optimization. The sequential kriging method constructed based only on the high-fidelity data that employs the original EI to sequentially generate samples (denoted as kriging), and cokriging employing FC for hyper-parameter estimation that is updated with only the high-fidelity data identified by EI (denoted as cokriging-H) are also employed for optimization, of which the results are compared to cokriging-M.

For the three approaches, the same convergence criterion in inequality (34) is employed. For the mathematical examples and the wrap-around fin example, since they are just used to test the effectiveness of the proposed sequential sampling method, the models from low to high fidelities ($y^1(\mathbf{x})$, $y^2(\mathbf{x})$, $y^3(\mathbf{x})$, ...) involved are all very computationally cheap, yielding almost the same computational costs. Thus, C cannot be assigned directly as the computational time of a single model evaluation. In our work, the cost C for these examples is roughly assigned based on the accuracy of the model. As the accuracy of the model improves, C assigned to this model is increased, which evidently accords with the practical situation. Although, it is a little subjective to determine the values of C (C^3 , C^2 and C^1) in our work and different values may

Table 6 Results of three approaches for wrap-around fin aerodynamic modeling

	FC		KOH		BM	
	RMSE	Time(s)	RMSE	Time(s)	RMSE	Time(s)
No.1	2.0478	92.83	2.4521	59.46	3.0114	48.21
No.2	2.1149	96.45	2.7688	63.58	3.1816	58.82
No.3	2.4511	101.63	2.8241	55.44	3.2448	53.62
No.4	1.9629	91.24	2.1149	57.82	2.9655	44.21
No.5	1.9614	88.62	2.2086	56.78	2.9226	55.78
Mean	2.1027	94.15	2.4737	58.61	3.0652	52.13

Fig. 8 Initial and sequentially updated cokriging models for example 2



yield different total computational cost, but similar optimal solutions. Therefore, the values of C does impact the verification process and conclusions. For brevity, only the results about the maximization of examples 2 and 3 are shown here.

For the problem of maximizing example 2, the numbers of uniformly generated initial samples for the three multi-fidelity models are $n_1 = 8, n_2 = 5, n_3 = 4$. It is assumed that the cost of a single model evaluation of y^3, y^2 and y^1 is $C^3 = 15, C^2 = 2$ and $C^1 = 1$, respectively. The cokriging model constructed by FC using the initial samples are shown in Fig. 8a and the cokriging models sequentially updated with new samples generated by the proposed EEI sampling criterion are shown in Fig. 8b–d.

It is observed that with the increase of the number of samples, the predicted response of the cokriging model is getting closer to that of the high-fidelity one (y^3), and the uncertainty of the cokriging model is reduced. With EEI, 2 samples from y^1 and 1 from y^2 are sequentially identified to update the cokriging model (see Fig. 8b). The uncertainty of the cokriging model in the latter half of the design region is slightly reduced with the maximum uncertainty value reduced from 0.0075 to 0.0049, while error exists around the optimum. Subsequently, 1 sample from y^1 and 1 from y^2 are identified to update the cokriging model, shown in Fig. 8c. Clearly, the accuracy in the middle design region of the updated cokriging

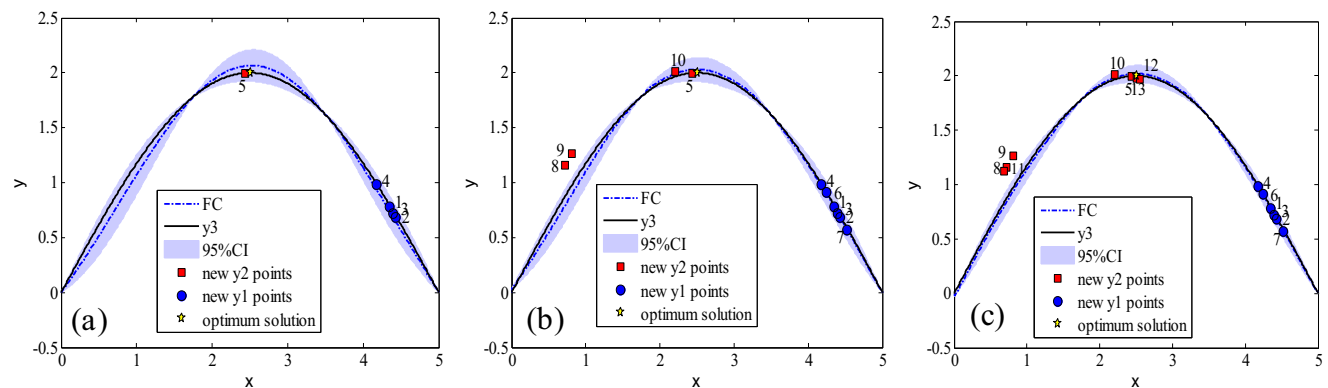


Fig. 9 Sequential samples without considering sample density (example 2)

Table 7 Optimal solution and computational cost of three approaches (example 2)

Methods	Evaluation number			Total cost	x^*	y^*
	$y^1(1)$	$y^2(2)$	$y^3(15)$			
Cokriging-M	11	10	4	91	2.4999	1.9999
Cokriging-H	8	5	6	108	2.5004	1.9999
Kriging	–	–	7	105	2.52	1.9998
Real opt.	–	–	–	–	2.5	2

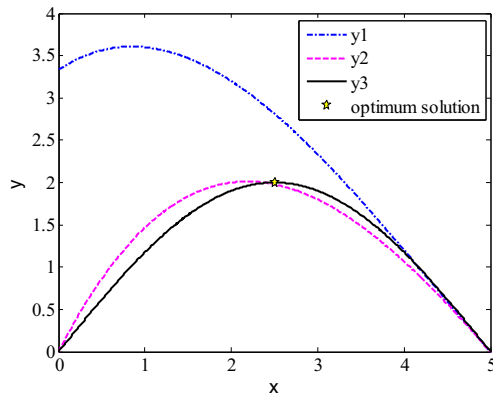


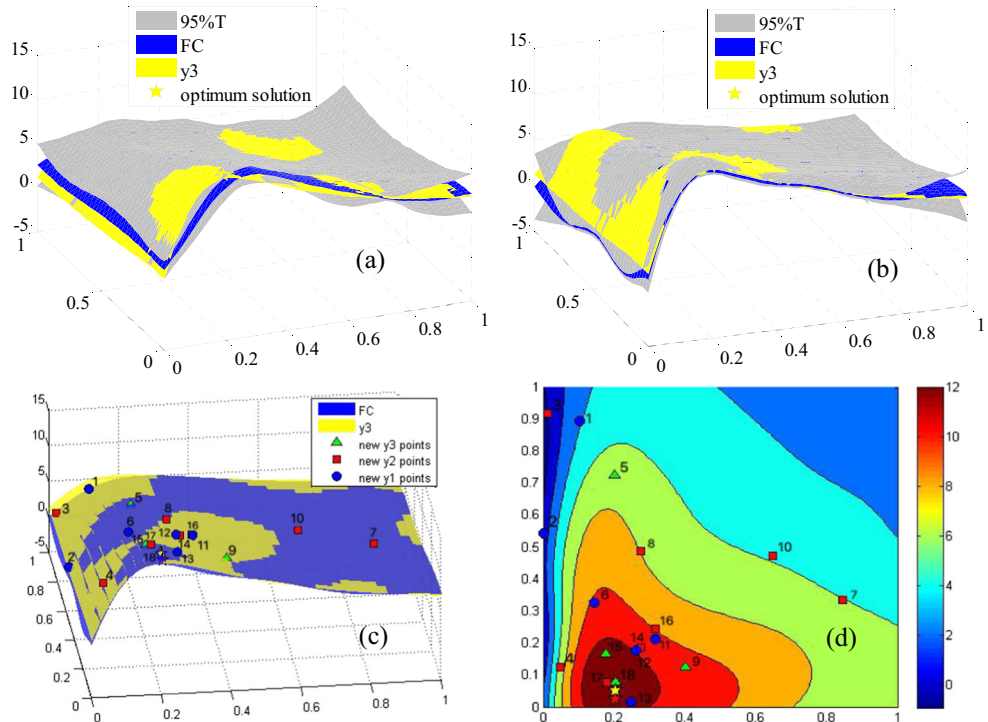
Fig. 10 Plots of 3-level fidelity models $y^3(x)$, $y^2(x)$ and $y^1(x)$

model is increased, especially for the region around the optimum. Finally, 3 new samples from y^2 that are all close to the optimum are sequentially generated to further improve the accuracy of the cokriging model around the optimum (see Fig. 8d), until the convergence criterion (see inequality (34)) is satisfied. Meanwhile, as can be seen from Fig. 8d, in the front half of the design region, evident discrepancy exists between the predicted responses of the final cokriging model and the high-fidelity model y^3 , which however does not affect the proposed optimization procedure to find the optimal solution ($x^* = 2.4999$, $y^* = 1.999$).

With the same initial samples and convergence criterion, EEI in (26) without including the sample density function $\eta(x, t)$ is also employed for sequential optimization, of which the cokriging models sequentially updated by 13 samples are shown in Fig. 9a–c. Clearly, it is noticed that the added samples (7 from y^2 and 6 from y^1) cluster around the optimum and the latter half of the design region. Although the optimal solution is accurate enough ($x^* = 2.4998$, $y^* = 1.999$), the total cost from the added samples is $20(7 \cdot 2 + 6 \cdot 1)$, while it is $14(5 \cdot 2 + 4 \cdot 1)$ for the proposed EEI sampling strategy. The interpretation is that adding dense and clustered samples has little effect on improving the accuracy of the metamodel, and adding too many samples far from the optimum is of little help to find the optimum. These results demonstrate the effectiveness of integration of the sample density function $\eta(x, t)$.

The optimal solutions and the cost are listed in Table 7, from which it is found that the optimal design variable x^* and

Fig. 11 Initial and updated cokriging models (example 3)



objective function y^* obtained by the three approaches are all very close to the real optimal solutions, while the proposed sequential optimization method is more efficient than the other two methods. From Table 7, it is also noticed that no sample from the highest-fidelity model $y^3(\mathbf{x})$ is selected during sequential sampling. The interpretation is that the objective-oriented sequential optimization prefers to select samples from $y^2(\mathbf{x})$ to improve the accuracy of metamodel near the optimum rather than $y^3(\mathbf{x})$, since the accuracy of $y^2(\mathbf{x})$ around the optimum is clearly very close to $y^3(\mathbf{x})$ (see Fig. 10), while its cost is cheaper ($C^2 = 2, C^1 = 1$). Therefore, the method chooses all the samples from $y^2(\mathbf{x})$ and $y^1(\mathbf{x})$ to make the least cost.

For the problem of maximizing example 3, it is assumed that the cost of a single model evaluation of y^3, y^2 and y^1 is $C^3 = 10, C^2 = 2$ and $C^1 = 1$, respectively. The initial samples are selected with $n_3 = 6, n_2 = 8$ and $n_1 = 10$, based on which the initial cokriging model is constructed, as shown in Fig. 11a. Clearly, large discrepancy exist between the predicted response of cokriging and the high-fidelity model y^3 . Samples sequentially identified with EEI and the updated cokrigings are shown in Fig. 11b–c. From Fig. 11b, it is observed that with the evolution of sequential optimization, the accuracy of the cokriging is clearly improved, with uncertainty clearly reduced. Meanwhile, as shown in Fig. 11c, many of the added samples are close to the optimum (indicated by star). The contour map of the final cokriging model is illustrated in Fig. 11d, from which it is also found that with the evolution of the sequential optimization, the added samples grow closer to the optimum, which is of great help to finding the optimum.

The optimal solutions and the cost are listed in Table 8, from which it is found that the optimal design variables and objective function obtained by the three approaches are all very close to the real optimal solutions, while the proposed sequential optimization method is clearly more efficient than the other two methods.

Table 8 Optimal solution and computational cost of three approaches (example 3)

Methods	Evaluation number			Total cost	\mathbf{x}^*	y^*
	$y^1(1)$	$y^2(2)$	$y^3(15)$			
Cokriging-M	16	14	12	164	(0.2051,0.0426)	13.3606
Cokriging-H	10	8	18	206	(0.2108,0.0578)	13.3458
Kriging	–	–	20	200	(0.2311,0.0768)	13.2298
Real opt.	–	–	–	–	(0.2051,0.0256)	13.3607

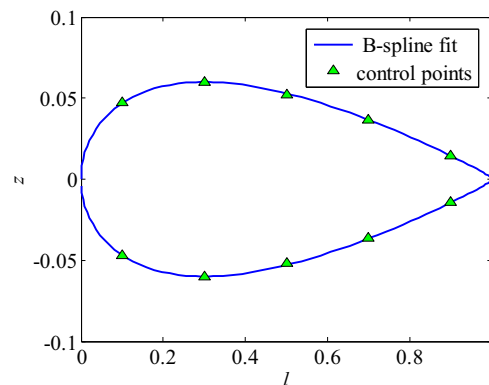


Fig. 12 B-spine parameterization for the airfoil

5 Application to drag minimization of the NACA 0012 airfoil

The proposed sequential optimization method is applied to a benchmark aerodynamic design problem simply based on the AIAA aerodynamic design optimization discussion group (Ren et al. 2016), which is a viscous case. It aims at maximizing the lift-to-drag ratio L/D of the modified NACA 0012 airfoil section at a free-stream Mach number of $M_\infty = 0.8$ and an angle of attack $\alpha = 2.5^\circ$, subject to the thickness constraint. The optimization problem is stated as

$$\begin{aligned} & \max_{1 \leq \mathbf{x} \leq \mathbf{u}} L(\mathbf{x})/D(\mathbf{x}) \\ & s.t. z(\mathbf{x})_{\max} \geq z_{\max\text{baseline}} \end{aligned} \tag{36}$$

where \mathbf{x} is the vector of design variables, depicted in detail below; $z(\mathbf{x})_{\max}$ is the airfoil maximum thickness and $z_{\max\text{baseline}}$ ($z_{\max\text{baseline}} = 0.12$) is the maximum thickness of the original baseline airfoil.

In this work, the B-spline curve (Farin 1993) with 10 control points is employed for the shape parameterization, where the horizontal locations of the 10 control points are fixed as $l = [0.1 \ 0.3 \ 0.5 \ 0.7 \ 0.9 \ 0.9 \ 0.7 \ 0.5 \ 0.3 \ 0.1]$ and the vertical locations are free to move, as shown in Fig. 12. The original airfoil is

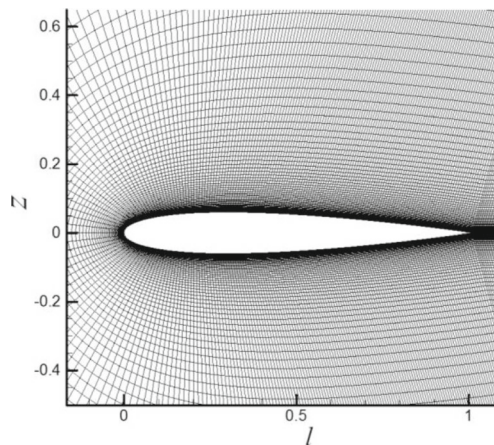


Fig. 13 Views of the airfoil with the mesh D

Table 9 Grid density convergence study for the NACA0012 airfoil

Mesh index	Grid size	L/D	Simulation time (s)
A	64×64	8.3271	37.5
B	128×128	8.4728	68.5
C	256×256	8.5944	152
D	512×512	8.6652	324.5
E	1024×1024	8.6684	2208

considered as the baseline with the vertical locations of the 10 control points as $\mathbf{z}_{\text{baseline}} = [0.04698 \ 0.06000 \ 0.05216 \ 0.03667 \ 0.01450 \ -0.01450 \ -0.03667 \ -0.05216 \ -0.06000 \ -0.04698]$. The vertical locations of the 10 control points are considered as design variables $\mathbf{x} = [x_1, \dots, x_{10}] = [z_1, \dots, z_{10}]$.

CFD is employed for aerodynamic analysis and a mesh with size 512×512 is generated using the integrated computer engineering and manufacturing code for CFD in the Ansys software as shown in Fig. 13. The mesh density is controlled by both the numbers of cells on and normal to the airfoil surface. It should be noted that the density of the mesh significantly impacts the simulation time cost and the accuracy of CFD analysis result. As shown in Table 9, when the mesh density is increased from mesh D to mesh E, the result of L/D does not change clearly. So for this application example, CFD with mesh D is considered as the highest-fidelity model $y^3(\mathbf{x})$, CFD with mesh C as the intermediate-fidelity model $y^2(\mathbf{x})$, and CFD with mesh B as the lowest-fidelity model $y^1(\mathbf{x})$.

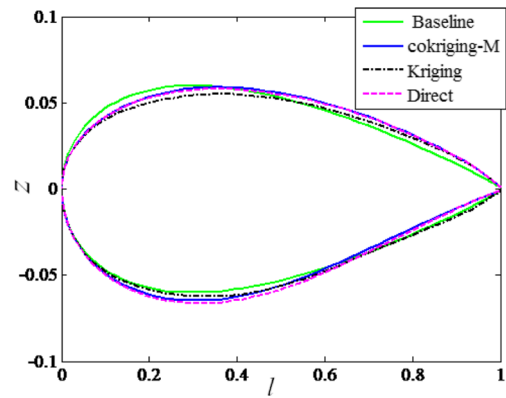


Fig. 14 The airfoil curve after optimization

Cokriging-M, cokriging-H and kriging are all employed to solve this optimization problem in (36). Furthermore, the optimization is also conducted directly on the CFD simulation model with mesh D (*i.e.* the highest-fidelity model) without employing any metamodel technique (denoted as direct), to verify the accuracy and effectiveness of the proposed cokriging-M. For cokriging-M, cokriging-H, kriging and direct, the genetic algorithm is used for searching the optima. For this drag minimization engineering problem, it is a practical problem and thus the models with different fidelities yield clearly different computational time (see Table 9). For this reason, C is directly set as the computational time of a single model evaluation for each model. Based on Table 9, it is obtained that the cost for y^3 , y^2 and y^1 is $C^3 = 324.5$ (s), $C^2 = 152$ (s) and $C^1 = 68.5$ (s).

Table 10 Comparative results of the three approaches

Methods	Evaluation number			Cost (min)	\mathbf{x}_{opt}	L/D_{opt}	$z(\mathbf{x})_{\text{max}}$
	y^1	y^2	y^3				
Cokriging-M	74	53	46	468	[0.042204 0.059528 0.054698 0.039769 0.018209 -0.015566 -0.039573 -0.05455 -0.057029 -0.044078]	14.4906	0.1273
Cokriging-H	50	40	72	548	[0.042204 0.055538 0.054698 0.039769 0.018209 -0.016154 -0.033246 -0.053171 -0.059029 -0.044082]	14.3021	0.1237
Kriging	-	-	119	643	[0.042204 0.055538 0.054698 0.037713 0.018209 -0.015566 -0.036004 -0.05655 -0.57584 -0.048778]	13.0776	0.1228
Direct	-	-	907	4905	[0.042204 0.059128 0.054198 0.039769 0.018209 -0.015566 -0.038683 -0.055124 -0.056329 -0.044078]	14.6180	0.1272

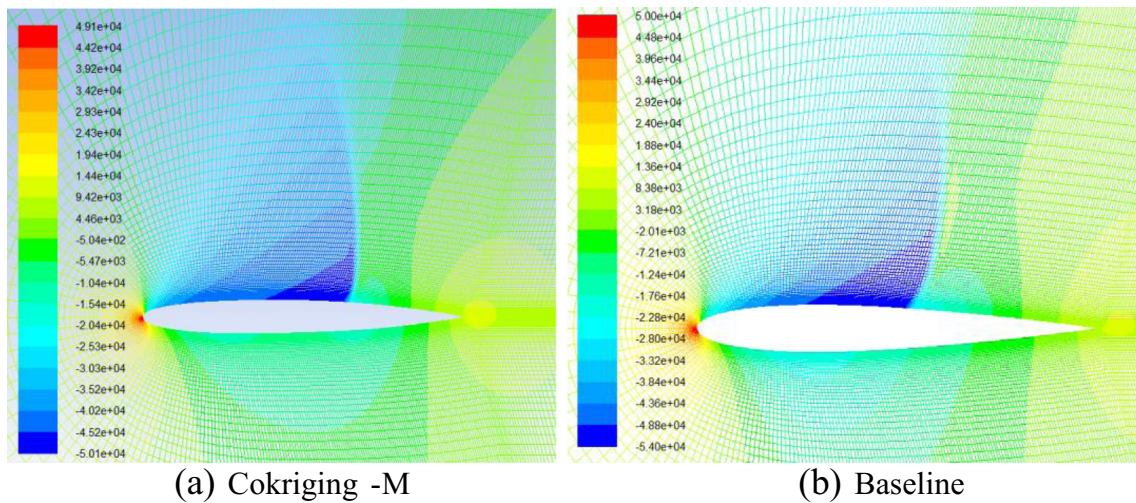


Fig. 15 Static pressure contours from baseline and cokriging-M

For cokriging-M and cokriging-H, $n_3 = 20$, $n_2 = 40$, and $n_1 = 50$ sample points are respectively generated by LHS to run the three selected multi-fidelity CFD models for initial cokriging modeling; meanwhile, for kriging, $n_3 = 20$

high fidelity samples are generated by LHS to construct the initial model.

The results are shown in Table 10, including the optimal design variable \mathbf{x}_{opt} , its objective function L/D_{opt} and con-

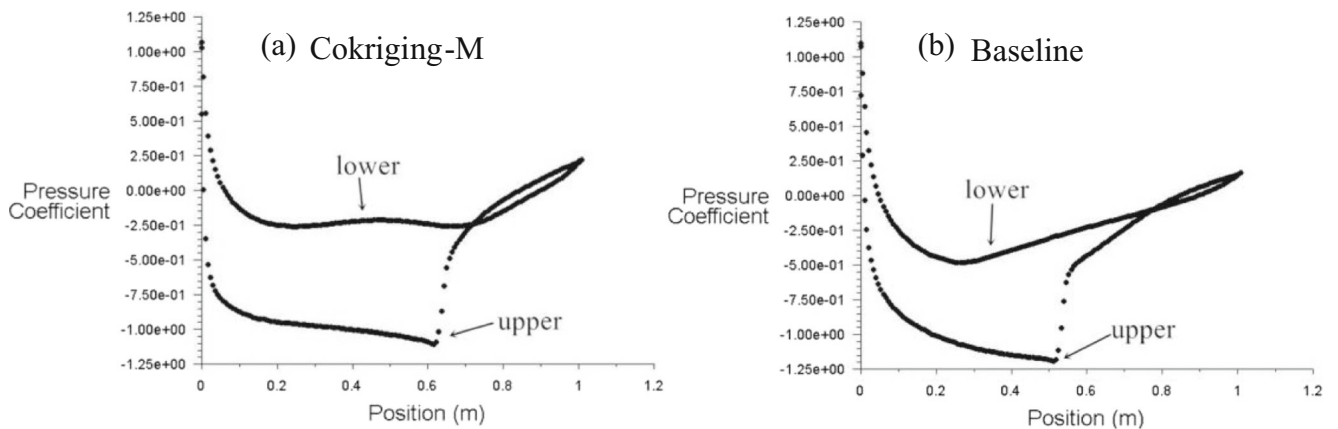


Fig. 16 Pressure coefficient distribution from baseline and cokriging-M

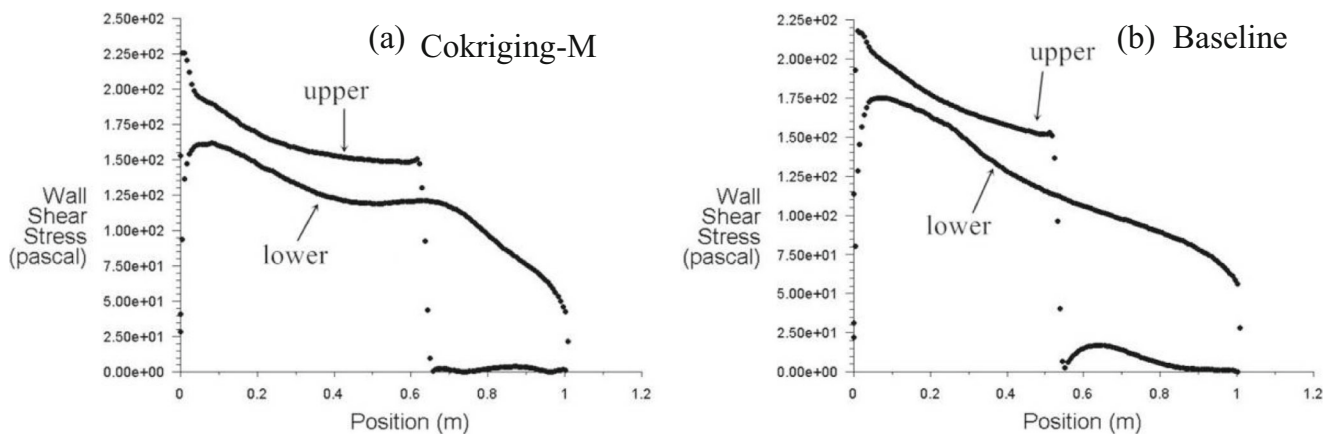


Fig. 17 Wall shear stress of cokriging-M and baseline

straint $z(\mathbf{x}_{\text{opt}})_{\text{max}}$, as well as the computational cost. From Table 10, it is found that \mathbf{x}_{opt} and L/D_{opt} obtained by the three approaches are all very close to those generated by direct; while generally the results of proposed cokriging-M are slightly closer to that of direct yielding smaller errors, and are more efficient than the other two methods. In the optimization process, direct requires 907 highest-fidelity model evaluations, while cokriging-M just needs 46 along with 53 intermediate-fidelity model evaluations and 74 low-fidelity ones, with much less total computational cost (468) compared to that of direct (4905). L/D of the airfoil obtained by cokriging-M is clearly improved after optimization compared to the original one. These results further demonstrate the effectiveness of the cokriging based optimization method proposed in this paper.

The airfoil curves after optimization by the proposed cokriging-M method and other approaches are shown in Fig. 14, from which it is found that compared to the baseline airfoil, the curvature and thickness of leading edge of the upper surface obtained by cokriging-M are both reduced, resulting the increase of the lift. Meanwhile, it is observed that the location of the maximum thickness of the airfoil moves backwards (baseline: $l = 0.3$; optimized: $l = 0.37$), which can make the turbulent transition delayed, and then reduce the drag. Therefore, the lift-to-drag ratio obtained by the proposed cokriging-M is increased ultimately.

In addition, Fig. 15 shows the static pressure contours. It is found that the static pressure of the upper surface optimized by cokriging-M is smaller than that of the baseline airfoil (the area in blue is larger for cokriging-M); while the static pressure is smaller for the lower surface (the area in green is larger for cokriging-M), which means that the lift of the optimized airfoil by cokriging-M is increased. In Fig. 16, the pressure coefficient distributions from cokriging-M and baseline are also shown, from which it is noticed that the pressure coefficient for the upper surface is reduced after optimization; meanwhile the pressure coefficient for the lower surface is increased, resulting in increased pressure differentials, *i.e.* increased lift. The results in Figs. 15 and 16 show great agreements on the optimal lift coefficient obtained by the proposed method, which is increased from 0.34922 (baseline) to 0.492546 (optimized). Fig. 17 illustrates the wall shear stress of the upper and lower surfaces, where the wall shear stress optimized by cokriging-M is generally smaller than that of the baseline. Since the viscous drag force is the integration of the wall shear stress, the drag is clearly reduced after optimization with cokriging-M, exhibiting great agreement to the obtained optimal drag coefficient 0.033991 that is clearly reduced compared to the baseline 0.040307.

6 Conclusions

In this paper, a cokriging based sequential optimization method is proposed. With the proposed approach, the covariance among data from all the multi-fidelity models is fully utilized in estimating the hyper-parameter through establishing a unified maximum likelihood function to improve the accuracy of cokriging modeling. Moreover, an extended expected improvement sequential sampling criterion considering the sample cluster issue is developed to more reasonably identify the location and fidelity level of the next response evaluation with reduced computational cost. Through comparative studies, it is noticed that the proposed method can make more accurate cokriging response prediction than the existing popular multi-level cokriging modeling approaches due to the consideration of the full correlation of all the data, which however requires longer computational time of modeling. Meanwhile, it is more computationally efficient compared to some existing sequential optimization methods. The effectiveness and advantage of the proposed method are further demonstrated using an AIAA benchmark aerodynamic optimization problem.

References

- Chen S, Jiang Z, Yang S, Apley DW, Chen W (2015) Nonhierarchical multi-model fusion using spatial random processes. *Int J Numer Methods Eng* 106(7):503–526
- El-Beltagy MA, Wright WA (2001) Gaussian processes for model fusion. *Artificial Neural Networks-ICANN 2001*, Vienna, pp 376–383
- Farin G (1993) *Curves and surfaces for computer aided geometric design*. Academic Press, Boston
- Forrester AI, Söbester A, Keane AJ (2007) Multi-fidelity optimization via surrogate modelling. *Proc R Soc Math Phys Eng Sci* 463(2088): 3251–3269
- Gratiet LL, Cannamela C (2015) Cokriging-based sequential design strategies using fast cross-validation techniques for multi-fidelity computer codes. *Technometrics* 57(3):418–427
- Han Z, Zimmerman R, Görtz S (2012) Alternative cokriging method for variable-fidelity surrogate modeling. *AIAA J* 50(5):1205–1210
- Huang D, Allen T, Notz W, Miller R (2006) Sequential kriging optimization using multiple-fidelity evaluations. *Struct Multidiscip Optim* 32(5):369–382
- Huang L, Gao Z, Zhang D (2013) Research on multi-fidelity aerodynamic optimization methods. *Chin J Aeronaut* 26(2):279–286
- Jin R, Chen W, Simpson TW (2001) Comparative studies of metamodelling techniques under multiple modelling criteria. *Struct Multidiscip Optim* 23(1):1–13
- Jin R, Chen W, Sudjianto A (2002) On sequential sampling for global metamodeling in engineering design. *ASME 2002 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Montreal, September 29–October 2, DETC2002/DAC-34092
- Jones DR, Schonlau M, Welch WJ (1998) Efficient global optimization of expensive black-box functions. *J Glob Optim* 13(4):455–492
- Keane AJ (2015) Cokriging for robust design optimization. *AIAA J* 50(11):2351–2364

- Kennedy MC, O'Hagan A (2000) Predicting the output from a complex computer code when fast approximations are available. *Biometrika* 87(1):1–13
- Laurenceau J, Sagaut P (2008) Building efficient response surfaces of aerodynamic functions with kriging and cokriging. *AIAA J* 46(2): 498–507
- Le Gratiet L (2013) Bayesian analysis of hierarchical multifidelity codes. *SIAM/ASA J Uncertain Quantif* 1(1):244–269
- Ng LW-T, Eldred M (2012) Multifidelity uncertainty quantification using nonintrusive polynomial chaos and stochastic collocation. The 14th AIAA Non-Deterministic Approaches Conference, Honolulu, April 23–26, AIAA-2012-1852
- Park C, Hafika RT, Kim NH (2016) Investigation of the effectiveness of multi-fidelity surrogates on extrapolation. *ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* Aug. 21–24, 2016, Charlotte, V02BT03A057
- Park C, Hafika RT, Kim NH (2017) Remarks on multi-fidelity surrogates. *Struct Multidiscip Optim* 55:1029–1050
- Qian PZ, Wu CJ (2008) Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments. *Technometrics* 50(2): 192–204
- Ren J, Leifsson LS, Koziel, Tesfahunegn YA (2016) Multi-fidelity aerodynamic shape optimization using manifold mapping. In 54th AIAA Aerospace Sciences Meeting, Science and Technology Forum, San Diego, Jan 4–8, 2016
- Sasena MJ (2002) Flexibility and efficiency enhancements for constrained global design optimization with kriging approximations. Ph.D dissertation, University of Michigan, Ann Arbor
- Shi Y, Xiong F, Xiu R, Liu Y (2013) A comparative study of relevant vector machine and support vector machine in uncertainty analysis. 2013 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (QR2MSE), pp 469–472
- Simpson TW, Poplinski JD, Koch PN, Allen JK (2001) Metamodels for computer-based engineering design: survey and recommendations. *Eng Comput* 17(2):129–150
- Skilling J (2006) Nested Sampling for general Bayesian computation. *Bayesian Anal* 1(1):3833–3860
- Toal DJJ, Keane AJ (2011) Efficient multipoint aerodynamic design optimization via cokriging. *J Aircr* 48(5):1685–1695
- Tuo R, Wu CJ, Yu D (2013) Surrogate modeling of computer experiments with different mesh densities. *Technometrics* 56(3):372–380
- Xiong Y, Chen W, Tsui KL (2008) A new variable-fidelity optimization framework based on model fusion and objective-oriented sequential sampling. *J Mech Des* 130(11):111401
- Xiong F, Xiong Y, Chen W et al (2009) Optimizing latin hypercube design for sequential sampling of computer experiments. *Eng Optim* 41(8):793–810
- Xue SH (2010) Research on multidisciplinary design optimization of aerodynamic & structural on wrap-around-wing rockets. Beijing Institute of Technology, Ph.D dissertation, (in Chinese)
- Yang RJ, Wang N, Tho CH, Bobineau JP, Wang BP (2005) Metamodeling development for vehicle frontal impact simulation. *J Mech Des* 127(5):1014–1020
- Yang Q, Luo W, Jiang Z et al (2016) Improve the prediction of soil bulk density by cokriging with predicted soil water content as auxiliary variable. *J Soils Sediments* 16(1):77–84
- Zheng J, Shao XY, Gao L, Jiang P, Li ZL (2013) A hybrid variable-fidelity global approximation modelling method combining tuned radial basis function base and kriging correction. *J Eng Des* 24(8): 604–622
- Zhu P, Zhang S, Chen W (2015) Multi-point objective-oriented sequential sampling strategy for constrained robust design. *Eng Optim* 47(3): 287–307