An Improved Co-Kriging Multi-fidelity Surrogate Modeling Method for Non-nested Sampling Data

Xiongfeng Ruan, Ping Jiang*
The State Key Laboratory of Digital Manufacturing Equipment and Technology, School of Mechanical Science and Engineering, Huazhong University of Science & Technology, 430074 Wuhan, PR China

Qi Zhou
School of Aerospace Engineering, Huazhong University of Science & Technology, 430074 Wuhan, PR China

Yang Yang
College of Engineering, Huazhong Agricultural University, Wuhan 430074, PR China
Email: jiangping@hust.edu.cn

Abstract—The multi-fidelity surrogate model, which can effectively balance the prediction accuracy and the modeling cost shows enormous potential in the design and optimization of mechanical products. Among them, the Co-Kriging multi-fidelity surrogate model based on Bayesian theory can provide the prediction error at the non-test points, which makes it especially attractive in the field of design optimization under uncertainty. However, in the Co-Kriging modeling process, high-fidelity (HF) and low-fidelity (LF) sampling points must be nested to satisfy the Markov property. If the Co-Kriging coefficients are obtained based on the full correlation, the modeling process will be complicated and result in low modeling efficiency. Therefore, this paper proposes an improved Co-Kriging multi-fidelity surrogate modeling method for non-nested sampling data. The proposed approach makes use of the characteristics of the stochastic kriging model to take the error of the LF surrogate model into consideration. Two independent processes are used to get the hyper-parameters of the LF surrogate model and the discrepancy model, respectively. The prediction accuracy and robustness of the proposed method are compared to the existing typical multi-fidelity surrogate modeling method on a standard numerical test example and an engineering example. The comparison results indicate that the proposed approach possesses not only excellent prediction accuracy but also outstanding robustness.

Index Terms—multi-fidelity surrogate model, uncertainty, Non-nested sampling data, Co-Kriging model

I. INTRODUCTION

The surrogate model also called approximation model is constructed with an analysis model of multi-fidelity [1,2]. The high-fidelity (HF) analysis model has high credibility, but the burden to acquire the response value of the model is expensive. On the contrary, obtaining the response value of the low-fidelity (LF) analysis model is low in cost, but its reliability is not convincing. How to solve the contradiction between the cost and dependability becomes a considerable challenge [3,4].

Engineering designers have come up with a compromise that using more cheaper LF sampling data to reduce computational complexity while utilizing fewer expensive HF sampling data to ensure the prediction performance of surrogate models. The solution is the multi-fidelity (i.e., variable credibility, variable fidelity) surrogate modeling method.

The multi-fidelity surrogate model is one of the research hotspots in the field of current surrogate modeling methods. The core problem is how to integrate the data information of the HF and LF analysis model effectively. There have been mainly three kinds of multi-fidelity surrogate modeling methods. (1) Multi-fidelity surrogate modeling method based on scaling function. For example, Zhou et al. proposed an adaptive global multi-fidelity metamodeling strategy using a support vector regression based scaling function [5]. (2) Multi-fidelity surrogate modeling method based on the space mapping. Robinson et al. introduced corrected space mapping to map between the variable-parameterization design spaces [6]. (3) Co-Kriging surrogate modeling method. Elsayed et al. adopted the Co-Kriging model to optimize the cyclone separator geometry to minimize pressure drop [7].

Among these three types, the Co-Kriging model is the most widely studied. The reason for the popularity of the Co-Kriging model is that it provides not only accurate prediction but also prediction error information at the non-test points. This article focuses on this class of methods. While the existing Co-Kriging model or its extension are either not applicable for non-nested sampling data or too time-consuming to be applied. In this article, an improved Co-Kriging model has been come up with to deal with non-nested sampling data under limited computation burden.

The organization of this paper is as follows. Brief descriptions of background and terminologies are given in Section 2. The detailed process of the proposed approach is described in Section 3. Two examples are provided in Section 4 to illustrate the efficiency of the
proposed approach. Section 5 gives the conclusion and future work.

II. BACKGROUND AND TERMINOLOGY

In Co-Kriging model, the input vectors of the LF and HF analysis calculations, \( x_l \) and \( x_h \), are taken to correlate with the outputs \( y_l \) and \( y_h \) by computational functions,

\[
y_l = f_L(x_l) \\
y_h = f_H(x_h)
\]

With the corresponding responses over these computer codes, the approximation \( \hat{y}_h \) can be produced as,

\[
\hat{y}_h = \rho \hat{y}_l(x) + \hat{y}_h(x)
\]

where \( \hat{y}_h \) is the sum of two Gaussian process models, and each Gaussian process model relies on the distances between the sampling data used to form them. \( \rho \) is a scaling parameter [8]. The hat symbols demonstrate the models are approximations, and the subscript \( d \) indicates a discrepancy model between the HF and LF functions. Note that \( \hat{y}_h(x) \) is not assumed constant with regard to \( x \), but rather a function that variations will be less complex than those of \( y_h(x) \), and thus it is simpler to model as \( \hat{y}_h(x) \).

For an \( k \)-dimensional problem, the distance measure applied between two sampling points \( i \) and \( j \) is:

\[
d(x^{(i)}, x^{(j)}) = \sum_{m=1}^{d} d_m(x^{(i)}_m - x^{(j)}_m)^2
\]

where \( m \) is the dimension number, \( d_m \) and \( P_m \) are hyper-parameters tuned to the data in hand. The correlation functions between points \( x^{(i)} \) and \( x^{(j)} \) is then expressed as

\[
R(x^{(i)}, x^{(j)}) = \exp(-d(x^{(i)}, x^{(j)}))
\]

When the response at a new point \( x^{new} \) is needed, the correlation vector \( c(x^{new}) \) between the new point is formed,

\[
c(x^{new}) = \begin{bmatrix} \rho \sigma_l^2 R_l(x_l, x^{new}) \\ \rho \sigma_l^2 R_l(x_h, x^{new}) + \sigma_h^2 R_h(x_h, x^{new}) \end{bmatrix}
\]

where the \( \sigma^2 \) are the variances in the discrepancy and LF Gaussian models, and the subscripts \( l \) and \( d \) on \( R \) reveal whether LF or discrepancy hyper-parameters are used, respectively.

The prediction \( \hat{y}_h(x^{new}) \) can be formed as,

\[
\hat{y}_h(x^{new}) = \mu + c(x^{new})^T C^{-1}(y - 1\mu)
\]

where

\[
\mu = \Gamma^T C^{-1} y / \Gamma^T C^{-1}
\]

and

\[
C = \begin{bmatrix} \sigma_l^2 R_l(x_l, x_l) & \rho \sigma_l^2 R_l(x_l, x_h) \\ \rho \sigma_l^2 R_l(x_h, x_l) & \rho \sigma_l^2 R_l(x_h, x_h) + \sigma_h^2 R_h(x_h, x_h) \end{bmatrix}
\]

where \( y \) is the combination vector of HF and LF function assessments. When building Co-Kriging model, the set of hyper-parameters needs to be carefully adjusted to match the data used. For Co-Kriging model, this tuning is used for the LF data, discrepancy data representing the distinctions between the HF and LF series, and the ratio parameter \( \rho \) to connect the various datasets.

Different from assessing the set of hyper-parameters, in the hierarchical kriging (HK) model proposed by Han [9], there are a set of covariance matrices to invert, and the assessment of the correlation Kernel parameters can be performed separately. Furthermore, the scaling factor which considers the effect of the LF function on the prediction of the HF function is explicitly adjusted during the model fitting phase of Co-Kriging, while in the HK model the scaling factor is implicitly adjusted tuned.

III. PROPOSED APPROACH

In Co-Kriging model, the LF surrogate model \( (\hat{y}_l) \) is constructed by the kriging model. When the HF sampling points are not the subset of LF points, the prediction value of LF surrogate model at HF sampling points \( (\hat{y}_l(x_h)) \) is not the exacted value of LF model, which contains prediction uncertainty as shown in Fig. 1. Then the kriging model is not suitable for constructing a discrepancy model here because it can only be applied to deterministic modeling. Therefore, an improved Co-Kriging (ICK) multi-fidelity surrogate modeling method, which is suitable for non-nested sampling data, has been developed by considering the LF surrogate model prediction uncertainty.

A. Predictor and Mean Square Error

Similar to Co-Kriging model, the formulation of ICK can be expressed as,

\[
\hat{y}_h(x) = \rho \hat{y}_l(x) + \hat{y}_h(x)
\]

where \( \rho \) is a scaling parameter, and \( \hat{y}_h(x) \) refers to a discrepancy model. \( \hat{y}_h(x) \) is the kriging surrogate model constructed based on LF sampling points, and then it can
be replaced by the addition of a linear combination and a random process

\[ \hat{y}_i(x) = f_i(x)\beta + z_i(x) \]  

(10)

The procedure to solve the equation can be specified in the Ref. [10].

As the HF and LF sampling points are non-nested, then the prediction value of LF surrogate model at HF sampling points contains mean square error, which means

\[ y_i(x) = \hat{y}_i(x) + e(x) \]  

(11)

with the constraint

\[ E(e(x)) = 0 \]

\[ \text{mse}(\hat{y}_i(x)) = \text{mse}(\hat{y}_i(x)) \]

where \( \text{mse}(\hat{y}_i(x)) \) is the prediction mean square error (MSE) of the surrogate model \( \hat{y}_i(x) \) at the point \( x \). In other words, \( y_i(x) \) is supposed to satisfy a normal distribution with mean \( \hat{y}_i(x) \) and variance \( \text{mse}(\hat{y}_i(x)) \). It is well to be reminded that \( e(x) \) is assumed to be independent between different inputs (i.e., \( \text{Cov}(e(x_{i}^{(h)}), e(x_{i}^{(l)})) = 0 \) for any \( i \neq j \)).

Therefore, \( y_i(x) \) can be replaced by \( \hat{y}_i(x) + e(x) \) as if the true value of \( y_i(x) \) is already known. In this case, HF and LF sampling data seems to satisfy the Markov property in a sense. The discrepancy value at HF sampling points is:

\[ y_{d}(x) = y_i(x) - \rho \hat{y}_i(x) = y_i(x) - \rho y_{d}(x) - \rho e(x) \]  

(13)

Furthermore, \( y_{d}(x) - \rho \hat{y}_{d}(x) \) can also be treated as a Gaussian process, which can be written as

\[ y_{d}(x) - \rho \hat{y}_{d}(x) = f_{d}(x)\beta + z_{d}(x) \]  

(14)

and then

\[ y_{d}(x) = f_{d}(x)\beta + z_{d}(x) - \rho e(x) \]  

(15)

The discrepancy function has a similar form to stochastic kriging model, so the similar solving procedure can be specified in the Ref. [11]. To simplify the expression, some symbols have been omitted in the following description.

Assuming that the discrepancy value at an unknown point \( x \) can be considered as the linear combination of existing discrepancy data, which has the form

\[ \hat{y}_{d}(x) = c^T Y_{d} \]  

(16)

where \( c \) is a vector of \( n_{d} \times 1 \), and \( Y_{d} = (y_{d}^{(1)} \ldots y_{d}^{(n_{d})})^T \). \( n_{d} \) is the number of HF sampling points. The prediction error can be shown as

\[ \hat{y}_{d}(x) - y_{d}(x) = c^T Y_{d} - y_{d}(x) \]

\[ = c^T (F\beta + Z - \rho \hat{E}) - (f(x)^T \beta + z) \]  

(17)

\[ = c^T (Z - \rho \hat{E}) - z + (F^T c - f(x))^T \beta \]

where \( Z = (z_{d}^{(1)} \ldots z_{d}^{(n_{d})})^T \) and \( E = (e(x_{i}^{(1)}) \ldots e(x_{i}^{(n_{d})}))^T \).

To satisfy the predictor unbiased, \( E(\hat{y}_{d}(x) - y_{d}(x)) = 0 \) is required, or

\[ F^T c(x) - f(x) = 0 \]  

(18)

Therefore, the MSE of the predictor is

\[ \phi(x) = E[(\hat{y}_{d}(x) - y_{d}(x))^2] \]

\[ = E[(c^T (Z - \rho \hat{E}) - z)^2] \]

(19)

\[ = c^T (\sigma^2 R + \rho^2 \hat{V})c + \sigma^2 (1 - 2\sigma^2 r) \]

where \( V \) is a \( n_{d} \)-dimensional diagonal matrix \( V = \text{diag}(\text{mse}(x_{i}^{(h)}), \ldots, \text{mse}(x_{i}^{(l)})) \), and \( v \) is the prediction MSE of the non-test point \( x \) of LF surrogate model.

To find the \( c \) to minimize MSE, the Lagrangian function has been used, which can be expressed as

\[ L(c, \lambda) = c^T (\sigma^2 R + \rho^2 \hat{V})c + \sigma^2 (1 - 2\sigma^2 r) - \lambda^T (F^T c - f) \]  

(20)

and with respect to \( c \), the gradient of Eq. (20) is

\[ L'_c(c, \lambda) = 2(\sigma^2 R + \rho^2 \hat{V})c - 2\sigma^2 r - F \lambda \]  

(21)

Combining the unbiased predictor condition with the first order necessary conditions for optimality, the following matrix equation is obtained

\[ \begin{bmatrix} \Lambda & F \\ F^T & 0 \end{bmatrix} \begin{bmatrix} c \\ \lambda \end{bmatrix} = \begin{bmatrix} r \\ f \end{bmatrix} \]  

(22)

where

\[ F = \{f(x^{(1)}), \ldots f(x^{(n_{d})})\}^T \]

\[ \lambda = -\frac{\lambda}{2\sigma^2} \]

\[ \Lambda = R + \rho^2 \hat{V} / \sigma^2 \]

\[ R = (R(x^{(i)}, x^{(j)}))_{i,j} \in R^{n_{d} \times n_{d}} \]

\[ r = (R(x^{(i)}, x)) \in R^{n_{d}} \]

The solution to Eq. (22) is

\[ \lambda = (F^T \Lambda^{-1} F)^{-1} (F^T \Lambda^{-1} r - f) \]

\[ c = \Lambda^{-1} (r - F\lambda) \]  

(23)

Then by inserting Eq. (23) to Eq. (16), the predictor of the discrepancy model is

\[ \hat{y}_{d}(x) = c^T Y_{d} = (r - F\lambda)^T \Lambda^{-1} Y_{d} \]

\[ = f^T \beta^* + r^T \Lambda^{-1} (Y_{d} - F\beta^*) \]

(24)

where \( \beta^* = (F^T \Lambda^{-1} F)^{-1} F^T \Lambda^{-1} Y_{d} \). Once the HF and LF sampling points are fixed, only the parameters \( f \) and \( r \) need to be recalculated for a new prediction point.

By substituting the solving parameters into Eq. (19), the expression of MSE can be shown as

\[ \phi(x) = \sigma^2 (1 + u^T (F^T \Lambda^{-1} F)^{-1} u - r^T \Lambda^{-1} r) \]  

(25)

where \( u = F^T \Lambda^{-1} r - f \).

After the LF surrogate model and discrepancy model constructed, the prediction value and MSE of the ICK model can be expressed as,
\[ \hat{y}_i = \rho \hat{y}_i(x) + \hat{y}_j(x) \]  
\[ \text{MSE} = \rho^2 \phi_i(x) + \phi_j(x) \]  

(26)

where \( \phi_i(x) \) and \( \phi_j(x) \) are prediction MSE of the LF and discrepancy surrogate model.

### B. Correlation Model

The correlation model is assumed to only depend on the spatial distance between two points \( x \) and \( x' \). The form of the correlation model can be given by

\[ R(x,x') = \prod_{i=1}^{n} R_i(\theta_i, x - x'_i) \]  

(27)

where \( \theta = (\theta_1, ..., \theta_m) \in \mathbb{R}^m \) are parameters that control the decrease rate of the correlation, and larger \( \theta \) leads to faster decrease. Among them, the most popular correlation model is the Gaussian exponential function, which is also selected in this paper. The expression of the Gaussian exponential function is

\[ R(x,x') = \prod_{i=1}^{n} \exp(-\theta_i(x - x'_i)^2) \]  

(28)

### C. Hyper-parameter Turning Strategy

The covariance matrix of the HF and LF sampling data can be simplified because of the Markov property by considering the LF approximation uncertainty. The probability of the data conditional on \( \phi \) can be converted to

\[ p(\hat{y}_i | \phi) = p(y_i | \gamma_i, \rho, \theta_i, \sigma_i^2) \]  

(29)

where \( \phi \) is the parameter of multi-fidelity surrogate model needed to be estimated. \( p(\hat{y}_i | \phi) \) can be solved by maximizing the two probability functions independently [12]. Each of them has a normal distribution, and then the following function should be maximized by estimating the parameters \( (\theta_i, \sigma_i^2) \) and \( (\rho, \theta_j, \sigma_j^2) \).

\[ L_\theta(\sigma_i^2, \theta_i) = \frac{1}{\sqrt{(2\pi\sigma_i^2)^{|R_i|}}} \times \exp\left(-\frac{1}{2} \frac{(Y_i - F_i \beta_i)^T R_i^{-1}(Y_i - F_i \beta_i)}{\sigma_i^2}\right) \]  

(30)

and

\[ L_\rho(\sigma_j^2, \theta_j, \rho) = \frac{1}{\sqrt{(2\pi\sigma_j^2)^{|A|}}} \times \exp\left(-\frac{1}{2} \frac{(Y_j - F_a \beta_a)^T A^{-1}(Y_j - F_a \beta_a)}{\sigma_j^2}\right) \]  

(31)

where subscript \( l \) and \( d \) refer to the parameters of LF and discrepancy surrogate model respectively.

Since the Eqs. (30) and (31) are implicit, so the parameters cannot be expressed by an expression directly. The Eqs. (30) and (31) can be solved by optimization algorithm, such as a genetic algorithm.

### IV. EXAMPLES AND RESULTS

In this paper, a numerical example and an engineering example are taken to verify the effectiveness of the ICK model. A widely used extended Co-Kriging, HK model, is compared to the ICK model. Maximum absolute error (MAE) and root mean square error (RMSE) are chosen as performance criteria. Smaller MAE implies the model provides better local prediction accuracy, while RMSE corresponds to global prediction precision.

#### A. Numerical Example

Forrester function is taken as the numerical example, whose LF function is dependent on the HF function [13]. The expressions of HF and LF functions are

\[ \text{HF} : y_i = (6z - 2)^2 \sin(2(6z - 2)), z \in (0,1) \]  

(32)

\[ \text{LF} : y_i = x_i - (A + 0.5)(6z - 2)^2 - 6z, x_i \in (0,1), A = 0.01 \]

where \( A \) reflects the correlation between HF and LF functions, and smaller \( A \) means stronger correlation. The LF sampling points are 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 1.0, while the HF sampling points are 0.05, 0.35, 0.75, and 0.95. It is obvious that HF and LF sampling points are non-nested. \( A \) is chosen as 0, 0.2, 0.4, 0.6, 0.8 and 1.0. The range and initial value of hyper-parameters are \( [10^{-2} \theta_{a}, 10^2 \theta_{a}] \) and \( \theta_{a} = 0.1 \) respectively [14].

The MAE, RMSE, and R-square of both approaches are shown in Fig. 2. R-square indicates the correlation between the HF model and multi-fidelity surrogate model. Compared to the HK model, the MAE and RMSE of the ICK model are significantly smaller, which means it provides both better local and global prediction accuracy. On the other hand, the MAE and RMSE of HK model increase monotonously with the increasing of \( A \), which indicates the correlation of HF and LF models have a significant impact on the prediction accuracy of HK model. However, the ICK model can overcome the influence of the model correlation. Similarly, the R-square between the HF function and the multi-fidelity surrogate model provides the same conclusion, which reveals that the ICK model is also more robust than the existing HK method.

![Figure 2. Performance citation of HK and ICK model](image-url)
Here \( A = 0, 0.6, \) and \( 1.0 \) are taken as an example. As is shown in Fig. 3, when \( A = 0 \), ICK model and HK model have similar prediction performance. As \( A \) increases (i.e., the correlation between HF and LF drops), the prediction accuracy of HK model drops much faster than ICK model. The HK model even fails to fit the HF function when \( A = 1.0 \). But the ICK method provides even better results compared to \( A = 0 \) and 0.6 .

Figure 3. Multi-fidelity surrogate model of Forrestdal function

B. Engineering Example

In this section, a three-dimensional aircraft drag coefficient prediction is taken as an engineering example. Because of the lack of empirical formula, it is difficult to describe the relationship between the angle of attack (\( \alpha \)), Mach number (\( Ma \)) and drag coefficient (\( C_L \)) [15]. Therefore, multi-fidelity surrogate model is taken as an effective solution. The ranges of variables are \(-4^\circ \leq \alpha \leq 6^\circ\) and \(0.6 \leq Ma \leq 0.9\) . The computational grids of HF and LF models are shown in Fig. 4. The LF model consists of 566,000 elements, which takes about 1.5 hours to simulate on a 2.6GHz CPU computer. As to the HF model, the element number is 2,630,000, and the computing time increases to 6.5 hours. The Spalart–Allmaras one-equation turbulence model is applied to both models on Fluent 15.0 simulation platform.

Figure 4. Computation grids for LF model (left) and HF model (right)

Figure 5. The distribution of the LF and HF sampling points

Figure 6. Coefficient prediction of HK model (L) and ICK model (R)

40 LF sampling points and 15 HF sampling points are generated by Latin hypercube sampling. The distribution of sampling points is shown in Fig. 5, where circles and triangles represent LF and HF sampling points, respectively. The response surface of HK model and the ICK model has been a plot in Fig. 6. Another 25 test points have been randomly generated in the feasible domain to validate the prediction accuracy, and the prediction performance has been shown in Table 1. The MAE of the ICK method is 33.4% smaller than that of HK model and the RMSE value is reduced by 29.6%. It is concluded that the ICK model significantly improves the prediction accuracy compared to the HK model, either globally or locally.

TABLE I. ACCURACY COMPARISON OF THE THREE-DIMENSIONAL AIRCRAFT EXAMPLE

<table>
<thead>
<tr>
<th>Methods</th>
<th>MAE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>HK model</td>
<td>0.0719</td>
<td>0.0250</td>
</tr>
<tr>
<td>ICK model</td>
<td>0.0479</td>
<td>0.0176</td>
</tr>
</tbody>
</table>

V. CONCLUSION

In this paper, an improved Co-Kriging multi-fidelity surrogate modeling method has been developed. The LF surrogate model is constructed by the kriging model, while the stochastic kriging model is applied to construct the discrepancy model. The hyper-parameters of two models are solved independently, which reduces
computation costs significantly compared to the conventional full-correlational Co-Kriging model. To validate the effectiveness of the ICK model, a numerical case and an engineering example are solved. The comparison results show that the ICK model provides both better local and global prediction accuracy compared to HK model. In addition, the ICK model is also more robust.

As future work, how to integrate the uncertainties of the sampling data into the ICK model will be investigated.

ACKNOWLEDGMENT

This research has been supported by National Natural Science Foundation of China (NSFC) under Grant No. 51775203, No. 51505163, and No. 51805179.

REFERENCES


