Received: 26 September 2022 | Revised: 29 October 2022 | Accepted: 1 November 2022

DOI: 10.1002/sta4.531

ORIGINAL ARTICLE

WILEY

An integral model for high-accuracy and low-accuracy experiments

Guanying Qi | Min-Qian Liu | Jian-Feng Yang

School of Statistics and Data Science, LPMC & KLMDASR, Nankai University, Tianjin, 300071, China

Correspondence
Jian-Feng Yang, School of Statistics and Data Science, LPMC & KLMDASR, Nankai University, Tianjin 300071, China.
Email: jfyang@nankai.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Numbers: 11811033, 12131001, 12226343, 12271270; National Ten Thousand Talents Program of China; Fundamental Research Funds for the Central Universities, Grant/Award Number: 63211090; Natural Science Foundation of Tianjin, Grant/Award Numbers: 20JCYBJC01050, 111 Project, B20016

A growing trend in engineering and science is to use multiple computer codes with different levels of accuracy to study the same complex system. Strategies are needed to combine the simulation results obtained at different levels of accuracy to produce an efficient surrogate model for prediction. In this paper, we propose an integral model to borrow as much information as possible from the low-accuracy experiment. We ignore the Markov property assumed before and model the high-accuracy experiment based on an integral form of the low-accuracy experiment. The proposed model is more general thus better predictions are expected. Two explicit forms of some matrices and vectors used in our predictions are given. The effectiveness of the proposed model is illustrated with several examples.

KEYWORDS
computer experiment, Gaussian process model, Kriging, nested space-filling design

1 | INTRODUCTION

Computer experiments play a critical role in many areas of scientific research where physical experiments are too costly or infeasible. Complex mathematical models are implemented in large computer codes to study real systems, and a large computer program often can be run at different levels of complexity with vastly varying computational times. For example, two different computer experiments can be used for designing the same type of linear cellular alloy for electronic cooling. One code uses finite-element analysis, whereas the other is based on the finite difference method. The two codes differ in terms of the numerical method and resolution of the grid, resulting in an accurate but slow version and a crude but fast approximation. In this paper we focus on the situation in which two experiments are available and one is generally more accurate than the other but also more expensive to run. The two experiments considered are called the low-accuracy experiment (LE) and high-accuracy experiment (HE). It is often something experimenters need to consider that how to integrate these multiple data sources efficiently.

The issue of building prediction models based on the HE and LE data has drawn significant attention. Related work includes Kennedy and O'Hagan (2000), Qian et al. (2006) and Qian and Wu (2008), among others. The basic idea of their modelling methods is to run the HE and LE with a pair of nested space-filling designs (Haaland & Qian, 2010; Qian, 2009; Qian, Ai, & Wu, 2009; Qian, Tang, & Wu, 2009; Sun et al., 2013, 2014) or a pair of nested Latin hypercube designs (Yang et al., 2014, 2016) and then fit a prediction model based on the LE data and refine it by incorporating the more accurate HE data. In their methods, it is always assumed that a kind of Markov property between HE and LE holds; thus, their models are based on an autoregressive model which will be introduced in Section 2.2.

The method of this paper is motivated by Example 3.6 in Chapter 3.5.2 of Santner et al. (2018). The HE and LE codes of this example are plotted in Figure 1 (more details can be found in Example 1 in this paper). It can be seen that although HE and LE share similar trends, the Markov property between HE and LE and the autoregressive model are not so appropriate for the codes as there seems to be a deviation in the x-axis direction. As will be stated in Section 2.2, the Markov property assumes that we cannot get any information for the response at an HE point from any other LE point if the response of the LE point at the same position of the HE point is given. This assumption will cause the autoregressive model to be less effective if the real HE experiment also depends on the real LE experiment on another position or even a larger scope of the design region, which seems more reasonable for this example. We believe that a similar situation may occur in practice. Thus, in this paper, we

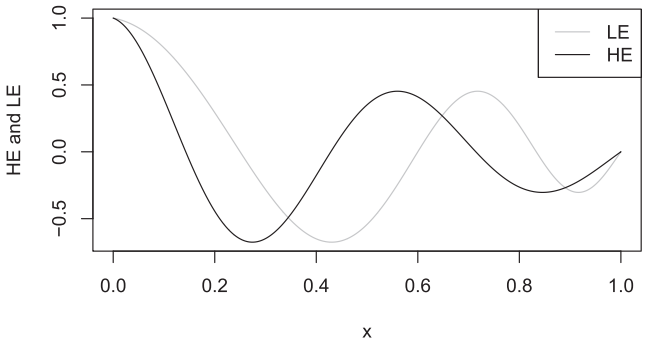


FIGURE 1 HE and LE in the motivating example

derive an integral model in order to make better predictions of HE on the whole design region. The proposed integral model not only ignores the Markov property but also is a generalization of the model in Kennedy and O'Hagan (2000). Under the assumption of Markov property being ignored, location and scale adjustments are implemented on an integral instead of a point of the surrogate model of LE. The paper is organized as follows. Section 2 briefly reviews the Gaussian process modelling and describes the proposed integral model in detail. It also gives some notation and two explicit forms of some matrices and vectors used when fitting the model. Section 3 shows some details when implementing the proposed modelling method. Section 4 uses several examples to illustrate the effectiveness of the integral model. Section 5 provides some concluding remarks. The proofs are deferred to Appendix A1.

2 | BUILDING A SURROGATE MODEL WITH BOTH HE AND LE

2.1 | Gaussian process modelling

The Gaussian process model (Kriging model) is widely used in computer experiments because of its desirable properties. Suppose the $n \times d$ matrix $\boldsymbol{X} = (\boldsymbol{x}_1,\boldsymbol{x}_2,...,\boldsymbol{x}_n)^T$ is the matrix of all inputs, where $\boldsymbol{x}_i = (x_{i1},...,x_{id})^T$ is the ith d-dimensional input vector. The corresponding response values are denoted by $\boldsymbol{z} = (z_1,...,z_n)^T$. The Gaussian process model assumes the following structure:

$$z(\boldsymbol{x}_i) = \boldsymbol{\beta}^T \boldsymbol{h}(\boldsymbol{x}_i) + \varepsilon(\boldsymbol{x}_i),\ i = 1,...,n,$$

where $\boldsymbol{h}(\boldsymbol{x}) = (h_1(\boldsymbol{x}),\ ...,\ h_q(\boldsymbol{x}))^T$ is a set of prespecified functions and $\boldsymbol{\beta} = (\beta_1,...,\beta_q)^T$ is a set of unknown coefficients. The $\varepsilon(\boldsymbol{x})$ is assumed to be a realization of a stationary Gaussian process with covariance

$$\mathrm{cov}[\varepsilon(\boldsymbol{x}_i),\varepsilon(\boldsymbol{x}_j)] = \sigma_\varepsilon^2 r_\varepsilon(\boldsymbol{x}_i,\boldsymbol{x}_j),$$

where σ_ε^2 is the variance and $r_\varepsilon(\cdot,\cdot)$ is the correlation function. One of the most commonly used forms of the correlation function is

$$r_\varepsilon(\boldsymbol{x}_i,\boldsymbol{x}_j) = \prod_{k=1}^{d} \exp\{-\theta_k |x_{ik} - x_{jk}|^p\},$$

where $\boldsymbol{\theta} = (\theta_1,...,\theta_d)^T$ is a vector of the correlation parameters, and in this paper, only $p = 1$ and $p = 2$ are considered. This form of correlation function is often adopted in the literature of computer experiments.

In this paper, a Gaussian process with mean μ, variance σ^2 and correlation parameters $\boldsymbol{\theta}$ is denoted by $GP(\mu,\sigma^2,\boldsymbol{\theta})$.

2.2 | An integral model

Suppose LE has m runs and HE has n runs. Let $\boldsymbol{X} = (\boldsymbol{x}_1,...,\boldsymbol{x}_m)^T$ and $\boldsymbol{Y} = (\boldsymbol{y}_1,...,\boldsymbol{y}_n)^T$ be the designs of LE and HE, respectively, and the corresponding response values are denoted by $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$, where $\boldsymbol{z}_1 = (z_1(\boldsymbol{x}_1),\ ...,\ z_1(\boldsymbol{x}_m))^T$ and $\boldsymbol{z}_2 = (z_2(\boldsymbol{y}_1),\ ...,\ z_2(\boldsymbol{y}_n))^T$. Let $\boldsymbol{z} = (\boldsymbol{z}_1^T,\boldsymbol{z}_2^T)^T$.

In Kennedy and O'Hagan (2000), it is assumed that

$$\mathrm{cov}[z_2(\boldsymbol{y}), z_1(\boldsymbol{x})|z_1(\boldsymbol{y})] = 0,$$

for all $\boldsymbol{y} \neq \boldsymbol{x}$. This assumption is called a kind of Markov property. Based on the assumption, Kennedy and O'Hagan (2000) proposed an autoregressive model, which is of the form

$$z_1(\boldsymbol{x}) = \varepsilon^*(\boldsymbol{x}), z_2(\boldsymbol{x}) = \rho_1 z_1(\boldsymbol{x}) + \delta^*(\boldsymbol{x}),$$

where

$$\varepsilon^*(\cdot) \sim GP(\boldsymbol{h}(\cdot)^T\boldsymbol{\beta}_\varepsilon, \sigma_\varepsilon^2, \boldsymbol{\theta}_\varepsilon), \delta^*(\cdot) \sim GP(\boldsymbol{h}(\cdot)^T\boldsymbol{\beta}_\delta, \sigma_\delta^2, \boldsymbol{\theta}_\delta),$$

ρ_1 is a kind of regression parameter and $\delta^*(\cdot)$ is independent of $z_1(\cdot)$. The corresponding correlation functions are denoted by $r_\varepsilon(\cdot,\cdot)$ and $r_\delta(\cdot,\cdot)$. In the rest of this paper, this autoregressive model together with the fitting methods in Kennedy and O'Hagan (2000) is called the KO model.

In this paper, to borrow more information from LE, we propose an integral model as a generalization of the autoregressive model. Assume that the design region $\mathcal{X}$ is a d-dimensional hypercube. The integral model we proposed in this paper is of the form

$$z_1(\boldsymbol{x}) = \varepsilon^*(\boldsymbol{x}), z_2(\boldsymbol{y}) = \rho_1 \int_{\mathcal{X}} z_1(\boldsymbol{t}) dF_{\boldsymbol{y}}(\boldsymbol{t}) + \delta^*(\boldsymbol{y}), \tag{1}$$

where ρ_1, $\varepsilon^*(\cdot)$ and $\delta^*(\cdot)$ are defined the same as above, and $F_{\boldsymbol{y}}(\boldsymbol{t})$ is a distribution function on $\mathcal{X}$ which satisfies that

$$\int_{\mathcal{X}} dF_{\boldsymbol{y}}(\boldsymbol{t}) = 1, \text{ for any } \boldsymbol{y} \in \mathcal{X}.$$

It can be easily seen that when $F_{\boldsymbol{y}}(\boldsymbol{x}) = \mathbf{1}_{\{\boldsymbol{x} \geq \boldsymbol{y}\}}$, the integral model degenerates to the autoregressive model.

Let $\boldsymbol{y}^*$ be a new design point at which we want to predict the HE response value $z_2(\boldsymbol{y}^*)$. Given

$$\Xi = \{\rho_1, \boldsymbol{\theta}_\varepsilon, \boldsymbol{\theta}_\delta, \sigma_\varepsilon^2, \sigma_\delta^2, \boldsymbol{\beta}_\varepsilon, \boldsymbol{\beta}_\delta\},$$

as a result of our integral model, we have

$$\left(z_2(\boldsymbol{y}^*), \boldsymbol{z}^T\right)^T = \begin{pmatrix} \boldsymbol{h}_{\text{int}}^T(\boldsymbol{y}^*) \\ \boldsymbol{H}_{\text{int}} \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \rho_1 \int_{\mathcal{X}} \varepsilon(\boldsymbol{t}) dF_{\boldsymbol{y}^*}(\boldsymbol{t}) + \delta(\boldsymbol{y}^*) \\ \mathcal{E}_{\text{int}} \end{pmatrix},$$

where

$$\boldsymbol{H}_{\text{int}} = \begin{pmatrix} \boldsymbol{h}^T(\boldsymbol{x}_1), \boldsymbol{0}^T \\ \vdots \\ \boldsymbol{h}^T(\boldsymbol{x}_m), \boldsymbol{0}^T \\ \rho_1 \int_{\mathcal{X}} \boldsymbol{h}^T(\boldsymbol{t}) dF_{\boldsymbol{y}_1}(\boldsymbol{t}), \boldsymbol{h}^T(\boldsymbol{y}_1) \\ \vdots \\ \rho_1 \int_{\mathcal{X}} \boldsymbol{h}^T(\boldsymbol{t}) dF_{\boldsymbol{y}_n}(\boldsymbol{t}), \boldsymbol{h}^T(\boldsymbol{y}_n) \end{pmatrix}, \mathcal{E}_{\text{int}} = \begin{pmatrix} \varepsilon(\boldsymbol{x}_1) \\ \vdots \\ \varepsilon(\boldsymbol{x}_m) \\ \rho_1 \int_{\mathcal{X}} \varepsilon(\boldsymbol{t}) dF_{\boldsymbol{y}_1}(\boldsymbol{t}) + \delta(\boldsymbol{y}_1) \\ \vdots \\ \rho_1 \int_{\mathcal{X}} \varepsilon(\boldsymbol{t}) dF_{\boldsymbol{y}_n}(\boldsymbol{t}) + \delta(\boldsymbol{y}_n) \end{pmatrix},$$

$\boldsymbol{h}_{\text{int}}(\boldsymbol{y}^*) = \left(\rho_1 \int_{\mathcal{X}} \boldsymbol{h}^T(\boldsymbol{t}) dF_{\boldsymbol{y}^*}(\boldsymbol{t}), \boldsymbol{h}^T(\boldsymbol{y}^*)\right)^T$, and $\boldsymbol{\beta} = (\boldsymbol{\beta}_\varepsilon^T, \boldsymbol{\beta}_\delta^T)^T$. Here, we have separated the mean terms $\boldsymbol{h}^T(\cdot)\boldsymbol{\beta}_\varepsilon$ and $\boldsymbol{h}^T(\cdot)\boldsymbol{\beta}_\delta$ from the Gaussian processes $\varepsilon^*(\cdot)$ and $\delta^*(\cdot)$ to get $\varepsilon(\cdot)$ and $\delta(\cdot)$, which are two independent mean zero Gaussian processes with parameters $(0, \sigma_\varepsilon^2, \boldsymbol{\theta}_\varepsilon)$ and $(0, \sigma_\delta^2, \boldsymbol{\theta}_\delta)$, respectively.

Given Ξ, the vector $\left(z_2(\boldsymbol{y}^*), \boldsymbol{z}^T\right)^T$ is multivariate normal distributed with mean $\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{h}_{\text{int}}^T(\boldsymbol{y}^*) \\ \boldsymbol{H}_{\text{int}} \end{pmatrix} \boldsymbol{\beta}$ and the covariance matrix $\boldsymbol{C} = \begin{pmatrix} \sigma_{\boldsymbol{y}^*}^2 & \boldsymbol{r}_{\text{int}}^T(\boldsymbol{y}^*) \\ \boldsymbol{r}_{\text{int}}(\boldsymbol{y}^*) & \boldsymbol{C}_{\text{int}} \end{pmatrix}$, where $\boldsymbol{C}_{\text{int}} = \mathrm{cov}(\mathcal{E}_{\text{int}}) = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}$, and

$$
\begin{aligned}
&\boldsymbol{r}_{\text{int}}(\boldsymbol{y}^*)\\
=\ &\Big(\rho_1\sigma_\varepsilon^2\int_{\mathcal{X}} r_\epsilon(\boldsymbol{x}_1,\boldsymbol{t})dF_{\boldsymbol{y}^*}(\boldsymbol{t}),\ldots,\rho_1\sigma_\varepsilon^2\int_{\mathcal{X}} r_\epsilon(\boldsymbol{x}_m,\boldsymbol{t})dF_{\boldsymbol{y}^*}(\boldsymbol{t}),\rho_1^2\sigma_\varepsilon^2\int_{\mathcal{X}^2} r_\epsilon(\boldsymbol{t}_1,\boldsymbol{t}_2)dF_{\boldsymbol{y}_1}(\boldsymbol{t}_1)dF_{\boldsymbol{y}^*}(\boldsymbol{t}_2)\\
&+\sigma_\delta^2 r_\delta(\boldsymbol{y}_1,\boldsymbol{y}^*),\ldots,\rho_1^2\sigma_\varepsilon^2\int_{\mathcal{X}^2} r_\epsilon(\boldsymbol{t}_1,\boldsymbol{t}_2)dF_{\boldsymbol{y}_n}(\boldsymbol{t}_1)dF_{\boldsymbol{y}^*}(\boldsymbol{t}_2)+\sigma_\delta^2 r_\delta(\boldsymbol{y}_n,\boldsymbol{y}^*)\Big)^T,\\
&C_{11}=\sigma_\varepsilon^2\boldsymbol{R}_\varepsilon,\\
&C_{12}=\rho_1\sigma_\varepsilon^2\begin{pmatrix}\int_{\mathcal{X}} r_\varepsilon(\boldsymbol{x}_1,\boldsymbol{t})dF_{\boldsymbol{y}_1}(\boldsymbol{t}) & \ldots & \int_{\mathcal{X}} r_\varepsilon(\boldsymbol{x}_1,\boldsymbol{t})dF_{\boldsymbol{y}_n}(\boldsymbol{t})\\ \vdots & & \vdots\\ \int_{\mathcal{X}} r_\varepsilon(\boldsymbol{x}_m,\boldsymbol{t})dF_{\boldsymbol{y}_1}(\boldsymbol{t}) & \ldots & \int_{\mathcal{X}} r_\varepsilon(\boldsymbol{x}_m,\boldsymbol{t})dF_{\boldsymbol{y}_n}(\boldsymbol{t})\end{pmatrix}=C_{21}^T,\\
&C_{22}=\rho_1^2\sigma_\varepsilon^2\begin{pmatrix}\int_{\mathcal{X}^2} r_\varepsilon(\boldsymbol{t}_1,\boldsymbol{t}_2)dF_{\boldsymbol{y}_1}(\boldsymbol{t}_1)dF_{\boldsymbol{y}_1}(\boldsymbol{t}_2) & \ldots & \int_{\mathcal{X}^2} r_\varepsilon(\boldsymbol{t}_1,\boldsymbol{t}_2)dF_{\boldsymbol{y}_1}(\boldsymbol{t}_1)dF_{\boldsymbol{y}_n}(\boldsymbol{t}_2)\\ \vdots & & \vdots\\ \int_{\mathcal{X}^2} r_\varepsilon(\boldsymbol{t}_1,\boldsymbol{t}_2)dF_{\boldsymbol{y}_n}(\boldsymbol{t}_1)dF_{\boldsymbol{y}_1}(\boldsymbol{t}_2) & \ldots & \int_{\mathcal{X}^2} r_\varepsilon(\boldsymbol{t}_1,\boldsymbol{t}_2)dF_{\boldsymbol{y}_n}(\boldsymbol{t}_1)dF_{\boldsymbol{y}_n}(\boldsymbol{t}_2)\end{pmatrix}+\sigma_\delta^2\boldsymbol{R}_\delta,
\end{aligned}
$$

where $\boldsymbol{R}_\varepsilon$ and $\boldsymbol{R}_\delta$ are the correlation matrices obtained by applying $r_\varepsilon(\cdot,\cdot)$ on $\boldsymbol{X}$ and $r_\delta(\cdot,\cdot)$ on $\boldsymbol{Y}$, respectively. Thus, for fixed Ξ, the expectation of $z_2(\boldsymbol{y}^*)$ conditional on $\boldsymbol{z}$ is

$$\mathbb{E}[z_2(\boldsymbol{y}^*)|\boldsymbol{z}]=\boldsymbol{h}_{\text{int}}^T(\boldsymbol{y}^*)\boldsymbol{\beta}+\boldsymbol{r}_{\text{int}}^T\boldsymbol{C}_{\text{int}}^{-1}(\boldsymbol{z}-\boldsymbol{H}_{\text{int}}\boldsymbol{\beta}). \tag{2}$$

Let $\Xi\setminus\boldsymbol{\beta}$ be the parameter set $\{\rho_1,\boldsymbol{\theta}_\varepsilon,\boldsymbol{\theta}_\delta,\sigma_\varepsilon^2,\sigma_\delta^2\}$, and we know that $\boldsymbol{z}=\boldsymbol{H}_{\text{int}}\boldsymbol{\beta}+\mathcal{E}_{\text{int}}\sim N_{m+n}(\boldsymbol{H}_{\text{int}}\boldsymbol{\beta},\boldsymbol{C}_{\text{int}})$. Thus, for fixed $\Xi\setminus\boldsymbol{\beta}$, by generalized least squares, we can get the best linear unbiased estimation and also the maximum likelihood estimation (MLE) of $\boldsymbol{\beta}$ as

$$\hat{\boldsymbol{\beta}}=(\boldsymbol{H}_{\text{int}}^T\boldsymbol{C}_{\text{int}}^{-1}\boldsymbol{H}_{\text{int}})^{-1}\boldsymbol{H}_{\text{int}}^T\boldsymbol{C}_{\text{int}}^{-1}\boldsymbol{z}. \tag{3}$$

After substituting (3) into (2), we get the predictor of $z_2(\boldsymbol{y}^*)$ as

$$\hat{z}_2(\boldsymbol{y}^*)=\boldsymbol{h}_{\text{int}}^T(\boldsymbol{y}^*)\hat{\boldsymbol{\beta}}+\boldsymbol{r}_{\text{int}}^T\boldsymbol{C}_{\text{int}}^{-1}(\boldsymbol{z}-\boldsymbol{H}_{\text{int}}\hat{\boldsymbol{\beta}}). \tag{4}$$

Now, we discuss the properties of the predictor of $z_2(\boldsymbol{y}^*)$ in (4). It is straightforward that $\hat{z}_2(\boldsymbol{y}^*)$ is an unbiased and interpolated predictor for $z_2(\boldsymbol{y}^*)$ when the model is true. Rewrite $\hat{z}_2(\boldsymbol{y}^*)$ as $\boldsymbol{b}_2^T(\boldsymbol{y}^*)\boldsymbol{z}$, where

$$
\begin{aligned}
\boldsymbol{b}_2^T(\boldsymbol{y}^*)\ =&\boldsymbol{h}_{\text{int}}^T(\boldsymbol{y}^*)(\boldsymbol{H}_{\text{int}}^T\boldsymbol{C}_{\text{int}}^{-1}\boldsymbol{H}_{\text{int}})^{-1}\boldsymbol{H}_{\text{int}}^T\boldsymbol{C}_{\text{int}}^{-1}+\boldsymbol{r}_{\text{int}}^T(\boldsymbol{y}^*)\boldsymbol{C}_{\text{int}}^{-1}\\
&-\boldsymbol{r}_{\text{int}}^T(\boldsymbol{y}^*)\boldsymbol{C}_{\text{int}}^{-1}\boldsymbol{H}_{\text{int}}(\boldsymbol{H}_{\text{int}}^T\boldsymbol{C}_{\text{int}}^{-1}\boldsymbol{H}_{\text{int}})^{-1}\boldsymbol{H}_{\text{int}}^T\boldsymbol{C}_{\text{int}}^{-1}.
\end{aligned}
$$

Then, we can get the mean squared prediction error of (4) as

$$
\begin{aligned}
&\mathbb{E}\left[(\boldsymbol{b}_2^T(\boldsymbol{y}^*)\boldsymbol{z}-z_2(\boldsymbol{y}^*))^2\right]\\
=\ &\mathbb{E}\left[\left((\boldsymbol{b}_2^T(\boldsymbol{y}^*)\boldsymbol{H}_{\text{int}}-\boldsymbol{h}_{\text{int}}^T(\boldsymbol{y}^*))\boldsymbol{\beta}+\boldsymbol{b}_2^T(\boldsymbol{y}^*)\mathcal{E}_{\text{int}}-\rho_1\int_{\mathcal{X}}\varepsilon(\boldsymbol{t})dF_{\boldsymbol{y}^*}(\boldsymbol{t})-\delta(\boldsymbol{y}^*)\right)^2\right]\\
=\ &\mathbb{E}\left[\left(\boldsymbol{b}_2^T(\boldsymbol{y}^*)\mathcal{E}_{\text{int}}-\rho_1\int_{\mathcal{X}}\varepsilon(\boldsymbol{t})dF_{\boldsymbol{y}^*}(\boldsymbol{t})-\delta(\boldsymbol{y}^*)\right)^2\right]\\
=\ &\boldsymbol{b}_2^T(\boldsymbol{y}^*)\boldsymbol{C}_{\text{int}}\boldsymbol{b}_2(\boldsymbol{y}^*)-2\boldsymbol{b}_2^T(\boldsymbol{y}^*)\boldsymbol{r}_{\text{int}}(\boldsymbol{y}^*)+\rho_1^2\sigma_\varepsilon^2\int_{\mathcal{X}^2} r_\epsilon(\boldsymbol{t}_1,\boldsymbol{t}_2)dF_{\boldsymbol{y}^*}(\boldsymbol{t}_1)dF_{\boldsymbol{y}^*}(\boldsymbol{t}_2)+\sigma_\delta^2.
\end{aligned}
$$

It is not hard to find out that (4) is the best linear unbiased prediction of $z_2(\boldsymbol{y}^*)$. If we then obtain MLEs of the other unknown parameters in Ξ and plug them in (4), we can get the empirical best linear unbiased prediction of $z_2(\boldsymbol{y}^*)$.

In order to get MLEs of the parameters in $\Xi\setminus\boldsymbol{\beta}$, we should maximize the likelihood function

$$\frac{1}{\sqrt{(2\pi)^{(m+n)}|\boldsymbol{C}_{\text{int}}|}}\exp\left\{-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{H}_{\text{int}}\hat{\boldsymbol{\beta}})^T\boldsymbol{C}_{\text{int}}^{-1}(\boldsymbol{z}-\boldsymbol{H}_{\text{int}}\hat{\boldsymbol{\beta}})\right\}.$$

After taking logarithm and ignoring some constant terms, it is equivalent to minimizing the value of the objective function

$$\ln|\boldsymbol{C}_{\text{int}}| + (\boldsymbol{z} - \boldsymbol{H}_{\text{int}}\hat{\boldsymbol{\beta}})^T \boldsymbol{C}_{\text{int}}^{-1}(\boldsymbol{z} - \boldsymbol{H}_{\text{int}}\hat{\boldsymbol{\beta}}).$$

2.3 | Explicit forms

In this part, we assign the distribution function in our integral model to be a uniform distribution or truncated normal distribution and then derive explicit forms for $\boldsymbol{C}_{\text{int}}$ and $\boldsymbol{r}_{\text{int}}$. These are the two most commonly used distribution functions, and we assign them to different choices of p in the correlation functions. In this paper, we assume $\mathcal{X}$ to be a bounded hypercube region. Let $\bar{\boldsymbol{l}}$ and $\bar{\boldsymbol{u}}$ denote the lower bound vector and upper bound vector of $\mathcal{X}$, respectively, for example, the kth element of $\bar{\boldsymbol{l}}$ is the lower bound of the kth dimension of $\mathcal{X}$.

Next, we specify the form of the distribution function $F_{\boldsymbol{y}}(\boldsymbol{t})$. For the uniform distribution function, let $\mathcal{X}_{\boldsymbol{y}}$ denote a hypercube region with boundary vectors $\boldsymbol{y}+\Delta-\boldsymbol{S}$ and $\boldsymbol{y}+\Delta+\boldsymbol{S}$, where Δ and $\boldsymbol{S}$ are unknown vector parameters, and $\mathcal{X}_{\boldsymbol{y}}^* = \mathcal{X}_{\boldsymbol{y}} \cap \mathcal{X}$. Denote the lower and upper bound vectors of $\mathcal{X}_{\boldsymbol{y}}^*$ by $\tilde{\boldsymbol{l}}_{\boldsymbol{y}}$ and $\tilde{\boldsymbol{u}}_{\boldsymbol{y}}$, respectively. Then, let $U_{\boldsymbol{y}}(\boldsymbol{t})$ be a uniform distribution function on $\mathcal{X}_{\boldsymbol{y}}^*$, whose density function is

$$u_{\boldsymbol{y}}(\boldsymbol{t}) = \frac{\boldsymbol{1}_{\mathcal{X}_{\boldsymbol{y}}^*}}{\prod_k (\tilde{\boldsymbol{u}}_{\boldsymbol{y},k} - \tilde{\boldsymbol{l}}_{\boldsymbol{y},k})}.$$

For the truncated normal distribution, let $q_{\boldsymbol{y}}(\boldsymbol{t})$ denote the density function of a multivariate normal distribution with mean $\boldsymbol{y}+\Delta$ and the covariance matrix Σ, where Δ and Σ are unknown parameters. Then, the density function $g_{\boldsymbol{y}}(\boldsymbol{t})$ of the assumed truncated normal distribution can be written as

$$g_{\boldsymbol{y}}(\boldsymbol{t}) = \frac{q_{\boldsymbol{y}}(\boldsymbol{t})}{\int_{\mathcal{X}} q_{\boldsymbol{y}}(\boldsymbol{t})} \cdot \boldsymbol{1}_{\mathcal{X}}.$$

Let $G_{\boldsymbol{y}}(\boldsymbol{t})$ be the corresponding cumulative distribution function.

With $F_{\boldsymbol{y}}(\boldsymbol{t})$ being $U_{\boldsymbol{y}}(t)$, we can derive the explicit forms of $\boldsymbol{C}_{\text{int}}$ and $\boldsymbol{r}_{\text{int}}$ when $p=1$; and with $F_{\boldsymbol{y}}(\boldsymbol{t})$ being $G_{\boldsymbol{y}}(t)$, we can derive the explicit forms of $\boldsymbol{C}_{\text{int}}$ and $\boldsymbol{r}_{\text{int}}$ when $p=2$. The following theorems show the explicit forms of $\boldsymbol{C}_{\text{int}}$ and $\boldsymbol{r}_{\text{int}}$ under these two situations.

In Theorems 2.1 and 2.2, denote $\exp\{-\theta_{\epsilon k}(a_k - b_k)\}$ by $\boldsymbol{e}_b^a(k)$.

Theorem 2.1. When $p=1$ in the correlation functions $r_\epsilon(\cdot,\cdot)$ and $r_\delta(\cdot,\cdot)$, and the distribution function in the integral model is the uniform distribution specified above, the elements in $\boldsymbol{C}_{\text{int}}$ have the following explicit forms:

$$C_{12}(i,j) = \rho_1 \sigma_\varepsilon^2 \prod_{k=1}^{d} \left(\int_{\tilde{\boldsymbol{l}}_{\boldsymbol{y}_j,k}}^{\tilde{\boldsymbol{u}}_{\boldsymbol{y}_j,k}} \exp\{-\theta_{\epsilon k}|x_{ik} - t_k|\} dt_k \right) \Big/ \prod_{k=1}^{d} \left(\tilde{\boldsymbol{u}}_{\boldsymbol{y}_j,k} - \tilde{\boldsymbol{l}}_{\boldsymbol{y}_j,k} \right),$$

where

$$\int_{\tilde{\boldsymbol{l}}_{\boldsymbol{y}_j,k}}^{\tilde{\boldsymbol{u}}_{\boldsymbol{y}_j,k}} \exp\{-\theta_{\epsilon k}|x_{ik} - t_k|\} dt_k = \begin{cases} \dfrac{1}{\theta_{\epsilon k}} \boldsymbol{e}_{x_i}^{\tilde{\boldsymbol{l}}_{\boldsymbol{y}_j}}(k) - \dfrac{1}{\theta_{\epsilon k}} \boldsymbol{e}_{x_i}^{\tilde{\boldsymbol{u}}_{\boldsymbol{y}_j}}(k), & \text{if } x_{ik} \le \tilde{\boldsymbol{l}}_{\boldsymbol{y}_j,k}; \\ \dfrac{2}{\theta_{\epsilon k}} - \dfrac{1}{\theta_{\epsilon k}} \boldsymbol{e}_{x_i}^{\tilde{\boldsymbol{l}}_{\boldsymbol{y}_j}}(k) - \dfrac{1}{\theta_{\epsilon k}} \boldsymbol{e}_{\tilde{\boldsymbol{l}}_{\boldsymbol{y}_j}}^{x_i}(k), & \text{if } \tilde{\boldsymbol{l}}_{\boldsymbol{y}_j,k} < x_{ik} \le \tilde{\boldsymbol{u}}_{\boldsymbol{y}_j,k}; \\ \dfrac{1}{\theta_{\epsilon k}} \boldsymbol{e}_{\tilde{\boldsymbol{u}}_{\boldsymbol{y}_j}}^{x_i}(k) - \dfrac{1}{\theta_{\epsilon k}} \boldsymbol{e}_{\tilde{\boldsymbol{l}}_{\boldsymbol{y}_j}}^{x_i}(k), & \text{if } \tilde{\boldsymbol{u}}_{\boldsymbol{y}_j,k} < x_{ik}, \end{cases}$$

and

$$C_{22}(i,j) = \rho_1^2 \sigma_\varepsilon^2 \prod_{k=1}^{d} \left(\int_{\mathcal{X}_{\boldsymbol{y}_i,k}^* \times \mathcal{X}_{\boldsymbol{y}_j,k}^*} \exp\{-\theta_{\epsilon k}|t_{1k} - t_{2k}|\} dt_{1k} dt_{2k} \right) \Big/ \prod_{k=1}^{d} \left[(\tilde{\boldsymbol{u}}_{\boldsymbol{y}_i,k} - \tilde{\boldsymbol{l}}_{\boldsymbol{y}_i,k}) \cdot (\tilde{\boldsymbol{u}}_{\boldsymbol{y}_j,k} - \tilde{\boldsymbol{l}}_{\boldsymbol{y}_j,k}) \right] + \sigma_\delta^2 r_\delta(\boldsymbol{y}_i, \boldsymbol{y}_j),$$

where

$$\int_{\mathcal{X}^*_{\mathbf{y}_i,k}\times\mathcal{X}^*_{\mathbf{y}_j,k}} \exp\{-\theta_{\varepsilon k}|t_{1k}-t_{2k}|\}dt_{1k}dt_{2k}$$
$$= \begin{cases} \frac{1}{\theta^2_{\varepsilon k}}\left(\boldsymbol{e}^{\tilde{\boldsymbol{u}}_{\mathbf{y}_i}}_{\tilde{\boldsymbol{l}}_{\mathbf{y}_j}}(k)+\boldsymbol{e}^{\tilde{\boldsymbol{l}}_{\mathbf{y}_i}}_{\tilde{\boldsymbol{u}}_{\mathbf{y}_j}}(k)-\boldsymbol{e}^{\tilde{\boldsymbol{l}}_{\mathbf{y}_j}}_{\tilde{\boldsymbol{l}}_{\mathbf{y}_i}}(k)-\boldsymbol{e}^{\tilde{\boldsymbol{u}}_{\mathbf{y}_j}}_{\tilde{\boldsymbol{u}}_{\mathbf{y}_i}}(k)\right), & \text{if}\tilde{\boldsymbol{l}}_{\mathbf{y}_i,k}>\tilde{\boldsymbol{u}}_{\mathbf{y}_j,k};\\ \frac{2}{\theta^2_{\varepsilon k}}\boldsymbol{e}^{\tilde{\boldsymbol{u}}_{\mathbf{y}_i}}_{\tilde{\boldsymbol{l}}_{\mathbf{y}_i}}+\frac{2}{\theta_{1k}}(\tilde{\boldsymbol{u}}_{\mathbf{y}_i}-\tilde{\boldsymbol{l}}_{\mathbf{y}_i})-\frac{2}{\theta^2_{1k}}, & \text{if}\tilde{\boldsymbol{l}}_{\mathbf{y}_i}=\tilde{\boldsymbol{l}}_{\mathbf{y}_j},\ \tilde{\boldsymbol{u}}_{\mathbf{y}_i}=\tilde{\boldsymbol{u}}_{\mathbf{y}_j};\\ \frac{1}{\theta^2_{\varepsilon k}}\left(\boldsymbol{e}^{\tilde{\boldsymbol{u}}_{\mathbf{y}_j}}_{\tilde{\boldsymbol{l}}_{\mathbf{y}_i}}(k)+\boldsymbol{e}^{\tilde{\boldsymbol{l}}_{\mathbf{y}_j}}_{\tilde{\boldsymbol{u}}_{\mathbf{y}_i}}(k)-\boldsymbol{e}^{\tilde{\boldsymbol{l}}_{\mathbf{y}_i}}_{\tilde{\boldsymbol{l}}_{\mathbf{y}_j}}(k)-\boldsymbol{e}^{\tilde{\boldsymbol{u}}_{\mathbf{y}_i}}_{\tilde{\boldsymbol{u}}_{\mathbf{y}_j}}(k)\right), & \text{if}\tilde{\boldsymbol{l}}_{\mathbf{y}_j,k}>\tilde{\boldsymbol{u}}_{\mathbf{y}_i,k}. \end{cases}$$

Other cases can be obtained by combining these three cases. When $\mathcal{X}\cap\mathcal{X}_{\boldsymbol{y}}\neq\emptyset$, $\tilde{\boldsymbol{u}}_{\boldsymbol{y},k}=\min\{\overline{\boldsymbol{u}}_k,\boldsymbol{y}_k+\Delta_k+\boldsymbol{S}_k\}$ and $\tilde{\boldsymbol{l}}_{\boldsymbol{y},k}=\max\{\overline{\boldsymbol{l}}_k,\boldsymbol{y}_k+\Delta_k-\boldsymbol{S}_k\}$.

Remark 2.1. Here, we show how the combination is done in the second part of the theorem. Suppose $\tilde{\boldsymbol{l}}_{\mathbf{y}_j,k}<\tilde{\boldsymbol{l}}_{\mathbf{y}_i,k}<\tilde{\boldsymbol{u}}_{\mathbf{y}_j,k}<\tilde{\boldsymbol{u}}_{\mathbf{y}_i,k}$, and let A denote $\exp\{-\theta_{1k}|t_{1k}-t_{2k}|\}$, then

$$\int_{\mathcal{X}^*_{\mathbf{y}_i,k}\times\mathcal{X}^*_{\mathbf{y}_j,k}} A dt_{1k}dt_{2k}=\left(\int_{\tilde{\boldsymbol{l}}_{\mathbf{y}_j,k}}^{\tilde{\boldsymbol{l}}_{\mathbf{y}_i,k}}\int_{\tilde{\boldsymbol{l}}_{\mathbf{y}_i,k}}^{\tilde{\boldsymbol{u}}_{\mathbf{y}_i,k}}+\int_{\tilde{\boldsymbol{l}}_{\mathbf{y}_i,k}}^{\tilde{\boldsymbol{u}}_{\mathbf{y}_j,k}}\int_{\tilde{\boldsymbol{l}}_{\mathbf{y}_i,k}}^{\tilde{\boldsymbol{u}}_{\mathbf{y}_j,k}}+\int_{\tilde{\boldsymbol{l}}_{\mathbf{y}_i,k}}^{\tilde{\boldsymbol{u}}_{\mathbf{y}_j,k}}\int_{\tilde{\boldsymbol{u}}_{\mathbf{y}_j,k}}^{\tilde{\boldsymbol{u}}_{\mathbf{y}_i,k}}\right)A dt_{1k}dt_{2k},$$

and the items on the right-hand side are all in the forms above.

Theorem 2.1 shows the explicit forms of the elements in $\boldsymbol{C}_{\text{int}}$ when $p=1$ and $F_{\boldsymbol{y}}(\boldsymbol{t})=U_{\boldsymbol{y}}(\boldsymbol{t})$. By a similar proof, we can get the explicit forms of the elements in $\boldsymbol{r}_{\text{int}}(\boldsymbol{y}^*)$, which are shown in Theorem 2.2.

Theorem 2.2. When $p=1$ in the correlation functions $r_\epsilon(\cdot,\cdot)$ and $r_\delta(\cdot,\cdot)$, and the distribution function in the integral model is the uniform distribution specified above, the $\boldsymbol{r}_{\text{int}}(\boldsymbol{y}^*)$ has the following explicit forms:

$$\boldsymbol{r}_{\text{int}}(\boldsymbol{y}^*)_i=\rho_1\sigma^2_\varepsilon\prod_{k=1}^{d}\left(\int_{\tilde{\boldsymbol{l}}_{\boldsymbol{y}^*,k}}^{\tilde{\boldsymbol{u}}_{\boldsymbol{y}^*,k}}\exp\{-\theta_{\varepsilon k}|x_{ik}-t_k\}dt_k\right)\Big/\prod_{k=1}^{d}(\tilde{\boldsymbol{u}}_{\boldsymbol{y}^*,k}-\tilde{\boldsymbol{l}}_{\boldsymbol{y}^*,k}),\text{and}$$
$$\boldsymbol{r}_{\text{int}}(\boldsymbol{y}^*)_{m+j}=\rho^2_1\sigma^2_\varepsilon\prod_{k=1}^{d}\left(\int_{\mathcal{X}^*_{\boldsymbol{y}^*,k}\times\mathcal{X}^*_{\mathbf{y}_j,k}}\exp\{-\theta_{\varepsilon k}|t_{1k}-t_{2k}|\}dt_{1k}dt_{2k}\right)$$
$$\Big/\prod_{k=1}^{d}\left[(\tilde{\boldsymbol{u}}_{\boldsymbol{y}^*,k}-\tilde{\boldsymbol{l}}_{\boldsymbol{y}^*,k})(\tilde{\boldsymbol{u}}_{\mathbf{y}_j,k}-\tilde{\boldsymbol{l}}_{\mathbf{y}_j,k})\right]+\sigma^2_\delta r_\delta(\boldsymbol{y}_j,\boldsymbol{y}^*),$$

where $i=1,\ldots,m$, $j=1,\ldots,n$ and other symbols and values of the integrals are similar to Theorem 2.1.

Before presenting Theorems 2.3 and 2.4, let $\Phi_{\{\boldsymbol{\mu},B\}}(\boldsymbol{a}\sim\boldsymbol{b})$ denote the probability that a random vector $\boldsymbol{\xi}\sim N(\boldsymbol{\mu},B)$ falls in a hypercube whose lower and upper bound vectors are $\boldsymbol{a}$ and $\boldsymbol{b}$, respectively.

Theorem 2.3. When $p=2$ in the correlation functions $r_\varepsilon(\cdot,\cdot)$ and $r_\delta(\cdot,\cdot)$, and the distribution function in the integral model is the truncated normal distribution specified above, the elements in $\boldsymbol{C}_{\text{int}}$ have the following explicit forms:

$$\boldsymbol{C}_{12}(i,j)=\rho_1\sigma^2_\varepsilon\sqrt{\frac{|(2U)^{-1}|}{|\Sigma|}}\exp\left\{\boldsymbol{\mu}^T_{ij}U\boldsymbol{\mu}_{ij}-\boldsymbol{x}^T_i\text{diag}(\boldsymbol{\theta}_\varepsilon)\boldsymbol{x}_i-\frac{1}{2}\boldsymbol{y}^{\Delta T}_j\Sigma^{-1}\boldsymbol{y}^\Delta_j\right\}$$
$$\cdot\Phi_{\{\boldsymbol{\mu}_{ij},(2U)^{-1}\}}(\overline{\boldsymbol{l}}\sim\overline{\boldsymbol{u}})\Big/\int_{\overline{\boldsymbol{l}}}^{\overline{\boldsymbol{u}}}q_{\mathbf{y}_j}(\boldsymbol{t})d\boldsymbol{t},\text{and}$$
$$\boldsymbol{C}_{22}(i,j)=\rho^2_1\sigma^2_\varepsilon\frac{\sqrt{|(2V)^{-1}|}}{\nu_{ij}|\Sigma|}\exp\left\{\boldsymbol{\nu}^T_{ij}V\boldsymbol{\nu}_{ij}-\frac{1}{2}\boldsymbol{y}^{\Delta T}_{ij}\begin{pmatrix}\Sigma^{-1}&0\\0&\Sigma^{-1}\end{pmatrix}\boldsymbol{y}^\Delta_{ij}\right\}$$
$$\cdot\Phi_{\{\boldsymbol{\nu}_{ij},(2V)^{-1}\}}(\overline{\boldsymbol{l}}\sim\overline{\boldsymbol{u}})+\sigma^2_\delta r_\delta(\boldsymbol{y}_i,\boldsymbol{y}_j),$$

where $\boldsymbol{y}_i^{\Delta}=\boldsymbol{y}_i+\Delta$, $\boldsymbol{y}_j^{\Delta}=\boldsymbol{y}_j+\Delta$, $\boldsymbol{y}_{ij}^{\Delta}=(\boldsymbol{y}_i^T+\Delta^T,\boldsymbol{y}_j^T+\Delta^T)^T$, $U=\mathrm{diag}(\boldsymbol{\theta}_\varepsilon)+\frac{1}{2}\Sigma^{-1}$, $\boldsymbol{\mu}_{ij}$ is the solution to the equation $\boldsymbol{\mu}_{ij}^T U=\boldsymbol{x}_i^T\mathrm{diag}(\boldsymbol{\theta}_\varepsilon)+\frac{1}{2}\boldsymbol{y}_j^{\Delta T}\Sigma^{-1}$, $V=\begin{pmatrix} I_d \\ -I_d\end{pmatrix}\mathrm{diag}(\boldsymbol{\theta}_\varepsilon)(I_d,-I_d)+\frac{1}{2}\begin{pmatrix}\Sigma^{-1} & 0\\ 0 & \Sigma^{-1}\end{pmatrix}$, $\mathcal{V}_{ij}=\int_{\mathcal{X}^2}q_{\boldsymbol{y}_i}(\boldsymbol{t}_1)q_{\boldsymbol{y}_j}(\boldsymbol{t}_2)d\boldsymbol{t}_1d\boldsymbol{t}_2$ and $\boldsymbol{\nu}_{ij}$ is the solution to the equation $V\boldsymbol{\nu}_{ij}=\frac{1}{2}\begin{pmatrix}\Sigma^{-1} & 0\\ 0 & \Sigma^{-1}\end{pmatrix}\boldsymbol{y}_{ij}^{\Delta}$.

Theorem 2.3 shows the explicit forms of the elements in $\boldsymbol{C}_{\mathrm{int}}$ when $p=2$ and $F_{\boldsymbol{y}}(\boldsymbol{t})=G_{\boldsymbol{y}}(t)$. Moreover, by a similar proof, we can get the explicit forms of the elements in $\boldsymbol{r}_{\mathrm{int}}(\boldsymbol{y}^*)$, which are shown in Theorem 2.4.

Theorem 2.4. When $p=2$ in the correlation functions $r_\varepsilon(\cdot,\cdot)$ and $r_\delta(\cdot)$, and the distribution function in the integral model is the truncated normal distribution specified above, the $\boldsymbol{r}_{\mathrm{int}}(\boldsymbol{y}^*)$ has the following explicit forms:

$$\begin{aligned}\boldsymbol{r}_{\mathrm{int}}(\boldsymbol{y}^*)_i=\ & \rho_1\sigma_\varepsilon^2\sqrt{\frac{|(2U)^{-1}|}{|\Sigma|}}\exp\left\{\boldsymbol{\mu}_i^{*T}U\boldsymbol{\mu}_i^*-\boldsymbol{x}_i^T\mathrm{diag}(\boldsymbol{\theta}_\varepsilon)\boldsymbol{x}_i-\frac{1}{2}(\boldsymbol{y}^{*\Delta})^T\Sigma^{-1}\boldsymbol{y}^{*\Delta}\right\}\\ & \cdot\Phi_{\{\boldsymbol{\mu}_i^*,(2U)^{-1}\}}(\bar{\boldsymbol{l}}\sim\bar{\boldsymbol{u}})\Big/\int_{\bar{\boldsymbol{l}}}^{\bar{\boldsymbol{u}}}q_{\boldsymbol{y}^*}(\boldsymbol{t})d\boldsymbol{t},\text{and}\\ \boldsymbol{r}_{\mathrm{int}}(\boldsymbol{y}^*)_{m+j}=\ & \rho_1^2\sigma_\varepsilon^2\frac{\sqrt{|(2V)^{-1}|}}{\mathcal{V}_j^*\,|\Sigma|}\exp\left\{\boldsymbol{\nu}_j^{*T}V\boldsymbol{\nu}_j^*-\frac{1}{2}(\boldsymbol{y}_j^{*\Delta})^T\begin{pmatrix}\Sigma^{-1} & 0\\ 0 & \Sigma^{-1}\end{pmatrix}\boldsymbol{y}_j^{*\Delta}\right\}\\ & \cdot\Phi_{\{\boldsymbol{\nu}_j,(2V)^{-1}\}}(\bar{\boldsymbol{l}}\sim\bar{\boldsymbol{u}})+\sigma_\delta^2r_\delta(\boldsymbol{y}_j,\boldsymbol{y}^*),\end{aligned}$$

where $i=1,\ldots,m$, $j=1,\ldots,n$, $\boldsymbol{y}^{*\Delta}=\boldsymbol{y}^*+\Delta$, $\boldsymbol{y}_j^{*\Delta}=(\boldsymbol{y}^{*T}+\Delta^T,\boldsymbol{y}_j^T+\Delta^T)^T$, $\boldsymbol{\mu}_i^*$ is the solution to the equation $\boldsymbol{\mu}_i^*U=\boldsymbol{x}_i\mathrm{diag}(\boldsymbol{\theta}_\varepsilon)+\frac{1}{2}(\boldsymbol{y}^{*\Delta})^T\Sigma^{-1}$, $\mathcal{V}_j^*=\int_{\mathcal{X}^2}q_{\boldsymbol{y}^*}(\boldsymbol{t}_1)q_{\boldsymbol{y}_j}(\boldsymbol{t}_2)d\boldsymbol{t}_1d\boldsymbol{t}_2$, $\boldsymbol{\nu}_j^*$ is the solution to the equation $V\boldsymbol{\nu}_j^*=\frac{1}{2}\begin{pmatrix}\Sigma^{-1} & 0\\ 0 & \Sigma^{-1}\end{pmatrix}\boldsymbol{y}_j^{*\Delta}$ and other symbols are defined similar to Theorem 2.3.

The assumptions behind the two choices of $F_{\boldsymbol{y}}(\boldsymbol{t})$ are different. A uniform distribution assumes that the value of $z_2(\boldsymbol{y})$ is based on the integral of $z_1(\boldsymbol{x})$s on a hypercube region centred on $\boldsymbol{y}+\Delta$ with even weights; and a truncated normal distribution assumes that $z_2(\boldsymbol{y})$ is based on the integral of $z_1(\boldsymbol{x})$s on the whole design region with uneven weights.

3 | IMPLEMENTATION

While fitting an integral model, several things should be considered. Firstly, the input and corresponding output data are standardized in advance, just like fitting a Gaussian process model. Secondly, we seek MLEs of the parameters to determine their values, and the determination is an optimization problem with box constraints. The same algorithm as Matlab toolbox DACE with successive coordinate search and pattern moves is used to solve this problem (Kowalik & Osborne, 1968, Section 2.4; Lophaven et al., 2002, Section 6). However, a minor adjustment is made to the algorithm. The purpose of the adjustment is to ensure that the algorithm can continue to work when the value of objective function does not exist after a certain step of the algorithm. For example, for a correlation function with $p=1$ and $F_{\boldsymbol{y}}(\boldsymbol{t})$ being $U_{\boldsymbol{y}}(\boldsymbol{t})$, after the search process reduces the value of $\boldsymbol{S}_k$ and the value of Δ_k is relatively large, it would happen that $\mathcal{X}\cap\mathcal{X}_{\boldsymbol{y}}=\emptyset$ on some $\boldsymbol{y}$s with $\boldsymbol{y}_k$ close to the upper bound in the kth dimension of $\mathcal{X}$. For the same reason, in order to ensure the existence of all predictions, it is preferred that all inputs which reach the bound of $\mathcal{X}$ in any dimension should be included in $\boldsymbol{Y}$. Thirdly, when we let $p=2$ in the correlation function and use the truncated normal distribution as $F_{\boldsymbol{y}}(\boldsymbol{t})$, the explicit forms contain the cumulative distribution function of the multivariate normal distribution. In our examples we use the "pmvnorm" function in the R package "mvtnorm" to get this value, but the computation time grows greatly as the dimension of input increases. So it is usually infeasible to find the MLEs directly. In this case, we can fit a KO model first and take the resulting parameter estimates to be the initial values in our integral model. The initial values for other parameters can be randomly chosen multiple times, and the algorithm stated above is applied to them separately. Similarly to ordinary Kriging, we assume $h(x)=1$ in the rest of the paper.

4 | NUMERICAL STUDIES

We use three examples to illustrate the proposed model. The first two are simulation studies and the third one is a real data example. For the convenience of comparison, all designs used in the examples have the property that the HE points are nested in the LE points, which is not necessary

for the proposed integral model. In this section, we use "int1" and "int2" to denote the applications of two different explicit forms with $p = 1$ and $p = 2$, respectively, and compare them with the KO model.

4.1 | Example 1

We use the method proposed in this paper to solve the problem in the motivating example from Chapter 3.5.2 of Santner et al. (2018).

It is supposed that the HE code is

$$z_2(x) = e^{-1.4x}\cos(7\pi x/2),\ x \in [0,1],$$

and the LE code is

$$z_1(x) = z_2(x/(2-x)),\ x \in [0,1].$$

In this example, we just choose $\mathcal{X}$ to be the interval $[0,1]$, and let $\boldsymbol{X} = \{0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1\}$, $\boldsymbol{Y} = \{0.1,0.3,0.5,0.7,0.9\}$. It can be seen that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are both uniformly scattered and satisfy $\boldsymbol{Y} \subset \boldsymbol{X}$. The HE code, LE code, $\boldsymbol{X}$ and $\boldsymbol{Y}$ we choose above are plotted in Figure 2a. The black line and points correspond to z_2, and the grey ones correspond to z_1.

Here, for the KO method, we use the "optim" function with the "L-BFGS-B" method in R to minimize the objective function. For our integral model, the objective function is minimized based on the algorithm we mentioned in Section 3. By minimizing the objective functions, we get MLEs of the unknown parameters. Note that all estimates are obtained after standardizing the input and output data.

For the KO model, the estimates are

$$\hat{\rho} = 0.3191, \hat{\boldsymbol{\theta}}_\varepsilon = 1.5537, \hat{\boldsymbol{\theta}}_\delta = 53.3038, \hat{\sigma}_\epsilon^2 = 3.2693, \hat{\sigma}_\delta^2 = 0.5425,$$

and the value of the objective function is -6.6578.

For the int1 model, the estimates are

$$\hat{\rho} = 1.6028, \hat{\boldsymbol{\theta}}_\varepsilon = 0.9259, \hat{\boldsymbol{\theta}}_\delta = 0.0074, \hat{\sigma}_\epsilon^2 = 1.1863, \hat{\sigma}_\delta^2 = 0, \hat{\boldsymbol{S}} = 0.3487, \hat{\Delta} = 0.4067,$$

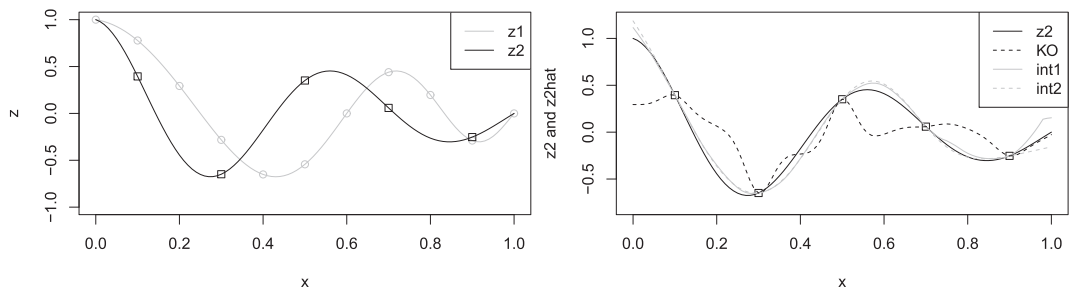
and the value of the objective function is -4.0993.

For the int2 model, the estimates are

$$\hat{\rho} = 1.6132, \hat{\boldsymbol{\theta}}_\varepsilon = 1.5673, \hat{\boldsymbol{\theta}}_\delta = 0.2717, \hat{\sigma}_\epsilon^2 = 3.1674, \hat{\sigma}_\delta^2 = 0.0628, \hat{\Sigma} = 0.0464, \hat{\Delta} = 0.4276,$$

and the value of the objective function is -29.8651.

We use these three models to predict z_2 on the 101 uniformly scattered points $\{0,0.01,0.02,\ldots,1\}$ in the design region, and the predictors are plotted in Figure 2b. It can be seen that all three predictors interpolate the HE points but the predictors from the proposed integral models can fit z_2 better than the KO model.



(a) z_1 and z_2 (b) z_2 and the predictors of z_2

FIGURE 2 The plots of z_1, z_2 and the predictors of z_2 in Example 1

The mean squared errors (MSEs) of these three predictors on these 101 points are also calculated. We get $MSE_{KO} = 0.0832$, $MSE_{int1} = 0.0043$, $MSE_{int2} = 0.0044$, which confirms our conclusion from the figure.

4.2 | Example 2

This example is modified from Example 1 in Xiong et al. (2013). In their paper, the HE code is

$$y_h = \left[1 - \exp\left(-\frac{1}{2x_2}\right)\right] \frac{2300x_1^3 + 1900x_1^2 + 2092x_1 + 60}{100x_1^3 + 500x_1^2 + 4x_1 + 20},$$

and the LE code is

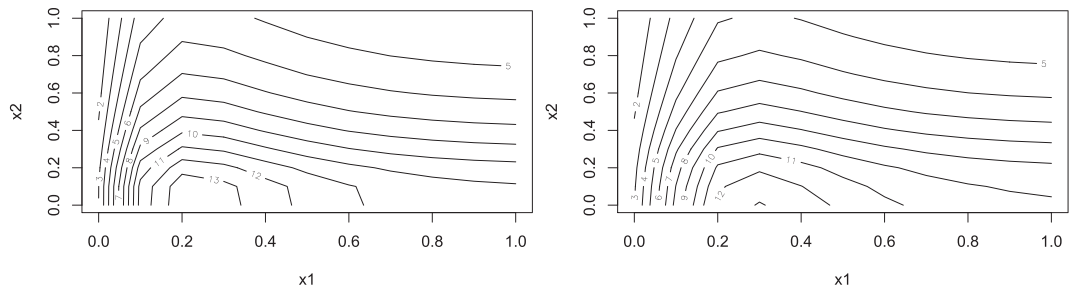
$$\begin{aligned} y_l &= [y_h(x_1 + 1/20, x_2 + 1/20) + y_h(x_1 + 1/20, \max(0, x_2 - 1/20)) \\ &\quad + y_h(x_1 - 1/20, x_2 + 1/20) + y_h(x_1 - 1/20, \max(0, x_2 - 1/20))]/4. \end{aligned}$$

Here, we assume the accuracy of the LE code is lower and replace the 1/20 in the expression for y_l with 1/10. Thus, in our example,

$$\begin{aligned} z_2(\boldsymbol{x}) &= \left[1 - \exp\left(-\frac{1}{2x_2}\right)\right] \frac{2300x_1^3 + 1900x_1^2 + 2092x_1 + 60}{100x_1^3 + 500x_1^2 + 4x_1 + 20}, \text{and} \\ z_1(\boldsymbol{x}) &= [z_2(x_1 + 1/10, x_2 + 1/10) + z_2(x_1 + 1/10, \max(0, x_2 - 1/10)) \\ &\quad + z_2(x_1 - 1/10, x_2 + 1/10) + z_2(x_1 - 1/10, \max(0, x_2 - 1/10))]/4. \end{aligned}$$

The contours of the two codes are shown in Figure 3.

We take $\boldsymbol{X}$ to be a sliced Latin hypercube design with 60 rows, two columns and four slices generated by the function "maximinLHD" in the R package "SLHD" (Ba et al., 2015), and let $\boldsymbol{Y}$ be the third slice of $\boldsymbol{X}$. Figure 4 shows the scatter plot of $\boldsymbol{X}$ and $\boldsymbol{Y}$ on the design region $\mathcal{X}$. In this



(a) The contour of z_2. (b) The contour of z_1.

FIGURE 3 The contours of z_2 and z_1 in Example 2

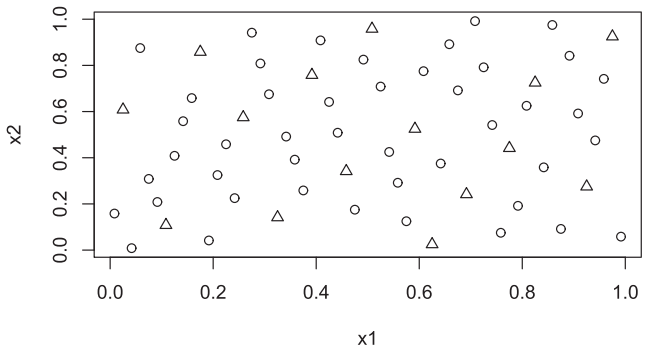
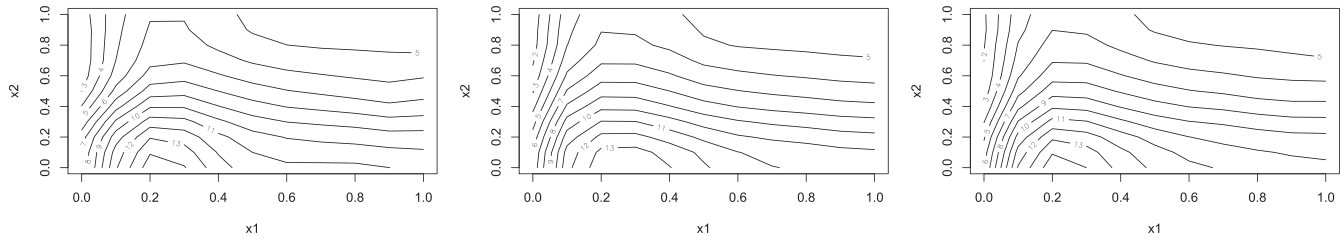


FIGURE 4 The scatter plot of $\boldsymbol{X}$ and $\boldsymbol{Y}$ with the triangles being the points of $\boldsymbol{Y}$ in Example 2

TABLE 1 Results of Example 2

	KO	int1	int2
$\hat{\rho}$	0.9353	1.2311	1.2035
$\hat{\boldsymbol{\theta}}_e$	0.8115	0.0996	0.8115
$\hat{\boldsymbol{\theta}}_\delta$	0.5047	6.1494	0.3945
$\hat{\sigma}_\epsilon^2$	1.7080	1.7687	1.7080
$\hat{\sigma}_\delta^2$	0.2299	0.0001	0.1365
$\hat{\boldsymbol{S}}(\text{int1})/\hat{\Sigma}(\text{int2})$		$(0.145; 0.536)^T$	$\text{diag}(0.0036; 0.0091)^T$
$\hat{\Delta}$		$(0.2010; 0.1860)^T$	$(0.1733; 0.0315)^T$
Objective function	−191.8456	−138.8033	−210.4528
RMSE	0.7958	0.5695	0.4891
SRMSE	0.3040	0.2054	0.1766



(a) The contour of KO model. (b) The contour of int1 model. (c) The contour of int2 model.

FIGURE 5 The contours of KO model, int1 model and int2 model in Example 2

TABLE 2 RMSEs and SRMSEs after ignoring points with $x_1 = 0$

	LE code	KO	int1	int2
RMSE	0.7721	0.2976	0.1719	0.2050
SRMSE	0.0903	0.0388	0.0227	0.0253

example, we use isotropic correlation functions, and when $p = 2$ in the correlation function, we restrict the covariance matrix Σ of the truncated normal distribution to be diagonal.

Table 1 gives the estimated parameters, values of the objective functions, root mean squared errors (RMSEs) and standardized RMSEs (SRMSEs) of all fitted models. The RMSEs and SRMSEs are calculated on the 121 gird points $\{0,0.1,...,1\} \times \{0,0.1,...,1\}$. Denote the grid points by $s_1,...,s_{121}$, the expression of SRMSE is

$$\sqrt{\frac{\sum_{i=1}^{121}\{[\hat{z}_2(s_i) - z_2(s_i)]/z_2(s_i)\}^2}{121}}.$$

The corresponding contours of the predictions are shown in Figure 5.

From Table 1, we can see that both integral models outperform the KO model in prediction, and it is better to take the combination of $p = 2$ and the truncated normal distribution for the correlation function and $F_{\boldsymbol{Y}}(\boldsymbol{t})$, respectively.

We also compare the fitted models with the LE code here. Regarding the LE code as a model, the RMSE value is 0.7363 and the SRMSE value is 0.0862. It is worth noting that the KO model is worse than the low-accuracy code under both criteria, and two integral models are worse than the low-accuracy code in terms of the SRMSE values. The most possible reason is that the responses z_1 and z_2 change rapidly with x_1 when x_1 is small and we did not choose enough design points in that area for the models to fit the change. Sometimes, this case cannot be avoided when we have no prior information. So we ignore the 11 grid points with $x_1 = 0$ and make the comparison on the other 110 points. The results are shown in Table 2. It can be seen that all three models outperform the LE code and the two integral models are better than the KO model. The combination of $p = 1$ and the uniform distribution performs the best.

4.3 | Example 3

We consider the linear cellular alloy example used by Qian et al. (2006) and Qian and Wu (2008). The inputs are four-dimensional, including the mass flow rate of entry air $\dot{m}$, the temperature of entry air T_{in}, the temperature of the heat source T_{wall} and the solid material thermal conductivity k. A detailed but slow simulation based on FLUENT finite-element analysis (HE) and an approximate but fast simulation using the finite difference method (LE) are used. The responses of the two experiments are denoted by y_h and y_l, respectively. Details on the engineering background can be found in Qian et al. (2006) and Qian and Wu (2008).

We use the same data as Example 1 of Qian and Wu (2008) with a small adaptation. Since the dimension of the data is higher than in the first two examples and the space-filling property of the data is not as good, we take the 20th point out of the training set and put the first point in it

TABLE 3 Data from the linear cellular alloy experiment

Run	$\dot{m}$ (kg/s)	T_{in} (K)	k (W/mK)	T_{wall} (K)	y_l	y_h	Status
1	0.000500	293.15	362.73	393.15	27.24	25.82	**Train**
2	0.000550	315.00	310.00	365.00	7.02	7.48	Train
3	0.000552	293.53	318.63	388.29	25.61	23.54	Train
4	0.000560	277.01	354.98	374.00	25.53	19.77	Test
5	0.000566	285.77	266.71	367.27	21.23	20.15	Train
6	0.000578	302.17	358.13	343.72	11.44	10.17	Train
7	0.000580	272.26	211.71	333.65	15.03	15.29	Train
8	0.000589	278.16	225.78	351.83	18.55	18.39	Train
9	0.000594	279.54	258.51	360.13	20.74	20.52	Test
10	0.000612	280.83	291.53	394.72	30.22	30.12	Train
11	0.000620	275.00	225.00	340.00	16.40	18.78	Test
12	0.000626	284.89	350.46	352.29	18.13	18.17	Train
13	0.000627	287.60	243.96	382.54	25.02	24.68	Test
14	0.000639	270.45	241.21	341.81	17.92	19.05	Train
15	0.000643	276.17	216.99	371.60	24.20	24.96	Train
16	0.000652	298.04	303.96	361.58	17.47	16.95	Train
17	0.000657	294.24	330.63	375.53	22.48	22.30	Test
18	0.000680	313.28	259.12	350.00	10.23	4.55	Test
19	0.000700	288.15	300.00	400.00	30.90	34.45	Train
20	0.000751	287.99	326.02	354.08	18.17	19.57	**Test**
21	0.000763	292.82	254.84	373.38	21.96	23.33	Test
22	0.000780	292.73	267.84	369.00	20.92	21.97	Train
23	0.000800	303.15	250.00	350.00	13.08	14.83	Train
24	0.000814	286.39	339.92	332.40	12.68	14.36	Train
25	0.000842	294.39	203.45	346.05	13.75	15.12	Train
26	0.000850	270.00	325.00	385.00	31.14	32.85	Test
27	0.000850	301.31	317.85	341.00	11.30	11.92	Train
28	0.000851	273.71	315.27	381.14	29.08	34.80	Test
29	0.000857	282.12	262.30	350.10	18.25	21.31	Train
30	0.000874	282.50	253.25	396.36	30.90	36.11	Test
31	0.000882	299.22	288.45	385.07	24.45	27.36	Test
32	0.000903	284.25	290.90	364.99	22.22	25.37	Train
33	0.000910	248.87	206.74	398.00	36.56	47.05	Train
34	0.000940	271.32	362.73	400.00	35.53	42.93	Train
35	0.000950	280.00	270.00	330.00	13.54	17.41	Train
36	0.001000	293.15	202.40	373.15	21.60	22.89	Train

TABLE 4 Results of Example 3

	KO	int1	int2
$\hat{\rho}$	1.2025	1.2176	1.3381
$\hat{\boldsymbol{\theta}}_{\varepsilon}$	0.0574	0.0335	0.0633
$\hat{\boldsymbol{\theta}}_{\delta}$	0.2357	0.0347	0.1839
$\hat{\sigma}_{\epsilon}^2$	2.4519	1.4780	2.1013
$\hat{\sigma}_{\delta}^2$	0.0929	0.1226	0.0515
$\hat{\boldsymbol{S}}(\text{int1})/\hat{\Sigma}(\text{int2})$		(1.33e−5, 9.58e−1, 1.05e−5, 5.66e−2)^T	diag(8.11e−3, 9.60e−3, 6.99e−2, 9.50e−4)^T
$\hat{\Delta}$		(0, 0.8844, 0, −0.0208)^T	(−0.5820, 0.1661, −1, −0.0989)^T
Objective function	−137.8099	−109.1277	−149.4909
RMSE	2.1529	2.2502	2.1515
SRMSE	0.0966	0.1056	0.0935

so that all inputs that reach the bound of $\mathcal{X}$ in any dimension are included in $\boldsymbol{Y}$ as recommended in Section 3. The data are shown in Table 3, where bold texts in the last column indicate the change we made to the data.

Similar to Example 2, isotropic correlation functions are used, and when $p = 2$ in the correlation function, we restrict the covariance matrix Σ of the truncated normal distribution to be diagonal. After fitting all three models (KO, int1 and int2), the estimated parameters, values of the objective functions, RMSEs and SRMSEs are shown in Table 4. Similar to Qian and Wu (2008), the results for run 18 are suppressed.

For this example, the two integral models do not show much in common. The results for int1 are worse than the KO model, and the results for int2 are slightly better than the KO model. In Qian and Wu (2008), the KO model performs the best among all models. As the fitting methods of the KO model use a correlation function with $p = 2$, a conclusion may be made that a correlation function with $p = 1$ cannot fit the data well, and the KO model is somehow good enough for the data. However, the KO model can still be improved by our int2 model, which is a generalization of it.

5 | CONCLUDING REMARKS

In this paper, we have developed an integral Gaussian process model for modelling and integrating LE and HE data. The model in (1) together with some explicit forms of matrices and vectors works well for integrating the HE and LE data in the numerical examples in Section 4. The implementation of location and scale adjustments on an integral of the surrogate model of LE makes the prediction much closer to HE in some cases, and the explicit forms allow us to do the fitting procedure. Compared with the existing methods that model the scale parameter in various ways, the proposed modelling approach provides another perspective which may perform better under certain circumstances. Further, one can consider using both methods simultaneously, but a significant increase in the number of parameters may cause problems.

Extensions of the present work can be made in several directions. Firstly, we have only derived two explicit forms in Section 2. There may exist combinations of correlation functions and distribution functions that we do not consider with better predictions and less computing time. Furthermore, Monte Carlo methods may be applied to deal with the cases that explicit forms are unable to derive. We have attempted to do so but could not achieve a balance between the computing time and accuracy. Secondly, we have chosen the design region $\mathcal{X}$ to be the integral region in the proposed integral model. However, this is not necessary. It is of interest that what would happen to the predictors if we make some change to the integral region, and how to choose a "best" one. Thirdly, in some cases, the proposed integral model can only give a little improvement to the KO model, and sometimes, the predictions may get worse as the likelihood function value gets bigger. Methods may be found out to determine when and how to use the integral model and how to avoid overfitting as well.

ACKNOWLEDGEMENTS

The authors are grateful to two anonymous referees for their insightful comments and suggestions. This work was supported by the National Natural Science Foundation of China (Grants 11811033, 12131001 12226343 and 12271270), National Ten Thousand Talents Program of China, Fundamental Research Funds for the Central Universities (63211090), Natural Science Foundation of Tianjin (20JCYBJC01050) and the 111 Project (B20016). The three authors contributed equally to this work.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analysed in this study.

ORCID

Jian-Feng Yang https://orcid.org/0000-0002-2271-4798

REFERENCES

Ba, S., Myers, W. R., & Brenneman, W. A. (2015). Optimal sliced Latin hypercube designs. *Technometrics*, *57*(4), 479–487.

Haaland, B., & Qian, P. Z. G. (2010). An approach to constructing nested space-filling designs for multi-fidelity computer experiments. *Statistica Sinica*, *20*(3), 1063–1075.

Kennedy, M. C., & O'Hagan, A. (2000). Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, *87*(1), 1–13.

Kowalik, J. S., & Osborne, M. R. (1968). *Methods for unconstrained optimization problems*, Vol. 13: Elsevier Publishing Company.

Lophaven, S. N., Nielsen, H. B., & Søndergaard, J. (2002). *Aspects of the Matlab Toolbox DACE. IMM, Informatics and Mathematical Modelling*: The Technical University of Denmark.

Qian, P. Z. G. (2009). Nested Latin hypercube designs. *Biometrika*, *96*(4), 957–970.

Qian, P. Z. G., Ai, M., & Wu, C. F. J. (2009). Construction of nested space-filling designs. *Annals of Statistics*, *37*(6A), 3616–3643.

Qian, Z., Seepersad, C. C., Joseph, V. R., Allen, J. K., & Wu, C. F. J. (2006). Building surrogate models based on detailed and approximate simulations. *Journal of Mechanical Design*, *128*(4), 668–677.

Qian, P. Z. G., Tang, B., & Wu, C. F. J. (2009). Nested space-filling designs for computer experiments with two levels of accuracy. *Statistica Sinica*, *19*(1), 287–300.

Qian, P. Z. G., & Wu, C. F. J. (2008). Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments. *Technometrics*, *50*(2), 192–204.

Santner, T. J., Williams, B. J., & Notz, W. I. (2018). *The design and analysis of computer experiments* (2nd ed.).: Springer.

Sun, F., Liu, M. Q., & Qian, P. Z. G. (2014). On the construction of nested space-filling designs. *Annals of Statistics*, *42*(4), 1394–1425.

Sun, F., Yin, Y., & Liu, M. Q. (2013). Construction of nested space-filling designs using difference matrices. *Journal of Statistical Planning and Inference*, *143*(1), 160–166.

Xiong, S., Qian, P. Z. G., & Wu, C. F. J. (2013). Sequential design and analysis of high-accuracy and low-accuracy computer codes. *Technometrics*, *55*(1), 37–46.

Yang, J., Liu, M. Q., & Lin, D. K. J. (2014). Construction of nested orthogonal Latin hypercube designs. *Statistica Sinica*, *24*(1), 211–219.

Yang, X., Yang, J. F., Lin, D. K. J., & Liu, M. Q. (2016). A new class of nested (nearly) orthogonal Latin hypercube designs. *Statistica Sinica*, *26*(3), 1249–1267.

How to cite this article: Qi, G., Liu, M.-Q., & Yang, J.-F. (2022). An integral model for high-accuracy and low-accuracy experiments. *Stat*, *11*(1), e531. https://doi.org/10.1002/sta4.531

APPENDIX A

Proof of Theorem 2.1. As long as $\tilde{\boldsymbol{u}}_{\boldsymbol{y},k} = \min\{\bar{\boldsymbol{u}}_k, \boldsymbol{y}_k + \Delta_k + \boldsymbol{S}_k\}$ and $\tilde{\boldsymbol{l}}_{\boldsymbol{y},k} = \max\{\bar{\boldsymbol{l}}_k, \boldsymbol{y}_k + \Delta_k - \boldsymbol{S}_k\}$ are known, the proof is straightforward. ∎

Proof of Theorem 2.3. In this setting, we have

$$\boldsymbol{C}_{12}(i,j) = \rho_1 \sigma_\varepsilon^2 \int_{\mathcal{X}} r_\varepsilon(\boldsymbol{x}_i, \boldsymbol{t}) q_{\boldsymbol{y}_j}(\boldsymbol{t}) d\boldsymbol{t} \Big/ \int_{\mathcal{X}} q_{\boldsymbol{y}}(\boldsymbol{t}) d\boldsymbol{t},$$

and the integral part of it is

$$
\begin{aligned}
&\int_{\mathcal{X}} r_\varepsilon(\boldsymbol{x}_i,\boldsymbol{t}) q_{\boldsymbol{y}_j}(\boldsymbol{t}) d\boldsymbol{t} \\
&= \frac{1}{\sqrt{(2\pi)^d|\Sigma|}} \int_{\bar{\boldsymbol{l}}}^{\bar{\boldsymbol{u}}} \exp\left\{-(\boldsymbol{t}-\boldsymbol{x}_i)^T \mathrm{diag}(\boldsymbol{\theta}_\varepsilon)(\boldsymbol{t}-\boldsymbol{x}_i) - \frac{1}{2}(\boldsymbol{t}-\boldsymbol{y}_j^\Delta)^T \Sigma^{-1}(\boldsymbol{t}-\boldsymbol{y}_j^\Delta)\right\} d\boldsymbol{t} \\
&= \frac{1}{\sqrt{(2\pi)^d|\Sigma|}} \exp\left\{\boldsymbol{\mu}^T U \boldsymbol{\mu} - \boldsymbol{x}_i^T \mathrm{diag}(\boldsymbol{\theta}_\varepsilon)\boldsymbol{x}_i - \frac{1}{2}\boldsymbol{y}_j^{\Delta T}\Sigma^{-1}\boldsymbol{y}_j^\Delta\right\} \int_{\bar{\boldsymbol{l}}}^{\bar{\boldsymbol{u}}} \exp\left\{-\frac{1}{2}(\boldsymbol{t}-\boldsymbol{\mu})^T (2U)(\boldsymbol{t}-\boldsymbol{\mu})\right\} d\boldsymbol{t} \\
&= \sqrt{\frac{|(2U)^{-1}|}{|\Sigma|}} \exp\left\{\boldsymbol{\mu}^T U \boldsymbol{\mu} - \boldsymbol{x}_i^T \mathrm{diag}(\boldsymbol{\theta}_\varepsilon)\boldsymbol{x}_i - \frac{1}{2}\boldsymbol{y}_j^{\Delta T}\Sigma^{-1}\boldsymbol{y}_j^\Delta\right\} \Phi_{\{\boldsymbol{\mu},(2U)^{-1}\}}(\bar{\boldsymbol{l}} \sim \bar{\boldsymbol{u}}).
\end{aligned}
$$

Similarly,

$$
\boldsymbol{C}_{22}(i,j) = \frac{\rho_1^2\sigma_\varepsilon^2}{\mathcal{V}_{ij}} \int_{\mathcal{X}^2} r_\varepsilon(\boldsymbol{t}_1,\boldsymbol{t}_2) q_{\boldsymbol{y}_i}(\boldsymbol{t}_1) q_{\boldsymbol{y}_j}(\boldsymbol{t}_2) d\boldsymbol{t}_1 d\boldsymbol{t}_2 + \sigma_\delta^2 r_\delta(\boldsymbol{y}_i,\boldsymbol{y}_j),
$$

and the integral part of it is

$$
\begin{aligned}
&\int_{\mathcal{X}^2} r_\varepsilon(\boldsymbol{t}_1,\boldsymbol{t}_2) q_{\boldsymbol{y}_i}(\boldsymbol{t}_1) q_{\boldsymbol{y}_j}(\boldsymbol{t}_2) d\boldsymbol{t}_1 d\boldsymbol{t}_2 \\
&= \frac{1}{(2\pi)^d|\Sigma|} \int_{\mathcal{X}^2} \exp\Big\{-(\boldsymbol{t}_1-\boldsymbol{t}_2)^T \mathrm{diag}(\boldsymbol{\theta}_\varepsilon)(\boldsymbol{t}_1-\boldsymbol{t}_2) \\
&\quad -\frac{1}{2}(\boldsymbol{t}_1-\boldsymbol{y}_i^\Delta)^T\Sigma^{-1}(\boldsymbol{t}_1-\boldsymbol{y}_i^\Delta) - \frac{1}{2}(\boldsymbol{t}_2-\boldsymbol{y}_j^\Delta)^T\Sigma^{-1}(\boldsymbol{t}_2-\boldsymbol{y}_j^\Delta)\Big\} d\boldsymbol{t}_1 d\boldsymbol{t}_2 \\
&= \frac{1}{(2\pi)^d|\Sigma|} \int_{\mathcal{X}^2} \exp\left\{-\boldsymbol{t}^T \begin{pmatrix} I_d \\ -I_d \end{pmatrix} \mathrm{diag}(\boldsymbol{\theta}_\varepsilon)(I_d,\ -I_d)\boldsymbol{t}\right. \\
&\quad \left. -\frac{1}{2}(\boldsymbol{t}-\boldsymbol{y}_{ij}^\Delta)^T \begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & \Sigma^{-1} \end{pmatrix} (\boldsymbol{t}-\boldsymbol{y}_{ij}^\Delta)\right\} d\boldsymbol{t}_1 d\boldsymbol{t}_2 \\
&= \frac{\sqrt{|(2U)^{-1}|}}{|\Sigma|} \exp\left\{\boldsymbol{\mu}^T U \boldsymbol{\mu} - \frac{1}{2}\boldsymbol{y}_{ij}^{\Delta T} \begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & \Sigma^{-1} \end{pmatrix} \boldsymbol{y}_{ij}^\Delta\right\} \Phi_{\{\boldsymbol{\mu},(2U)^{-1}\}}(\bar{\boldsymbol{l}} \sim \bar{\boldsymbol{u}}).
\end{aligned}
$$

■