Bayesian probabilistic propagation of hybrid uncertainties: Estimation of response expectation function, its variable importance and bounds

Chao Dang[a], Pengfei Wei[b,*], Matthias G.R. Faes[c], Michael Beer[a,d,e]

[a]*Institute for Risk and Reliability, Leibniz University Hannover, Callinstr. 34, Hannover 30167, Germany*
[b]*School of Power and Energy, Northwestern Polytechnical University, Xi'an 710072, PR China*
[c]*Chair for Reliability Engineering, TU Dortmund University, Leonhard-Euler-Str. 5, Dortmund 44227, Germany*
[d]*Institute for Risk and Uncertainty, University of Liverpool, Liverpool L69 7ZF, United Kingdom*
[e]*fInternational Joint Research Center for Resilient Infrastructure & International Joint Research Center for Engineering Reliability and Stochastic Mechanics, Tongji University, Shanghai 200092, PR China*

Abstract

Uncertainties existing in physical and engineering systems can be characterized by different kinds of mathematical models according to their respective features. However, efficient propagation of hybrid uncertainties via an expensive-to-evaluate computer simulator is still a computationally challenging task. In this contribution, estimation of response expectation function (REF), its variable importance and bounds under hybrid uncertainties in the form of precise probability models, parameterized probability-box models and interval models is investigated through a Bayesian approach. Specifically, a new method, termed "Parallel Bayesian Quadrature Optimization" (PBQO), is developed. The method starts by treating the REF estimation as a Bayesian probabilistic integration (BPI) problem with a Gaussian process (GP) prior, which in turn implies a GP posterior for the REF. Then, one acquisition function originally developed in BPI and other two in Bayesian global optimization are introduced for Bayesian experimental designs. Besides, an innovative strategy is also proposed to realize multi-point selection at each iteration. Overall, a novel advantage of PBQO is that it is capable of yielding the REF, its variable importance and bounds simultaneously via a pure single-loop procedure allowing for parallel computing. Three numerical examples are studied to demonstrate the performance of the proposed method over existing methods.

Keywords: Hybrid uncertainties, Response expectation function, Bayesian probabilistic integration, Bayesian global optimization, Bayesian experimental design, Parallel computing

1. Introduction

Uncertainty quantification (UQ) is a hot topic and even research frontier in a broad range of modern science and engineering fields. UQ is primarily aimed at the quantitative characterization and consequent

*Corresponding author
Email addresses: chao.dang@irz.uni-hannover.de (Chao Dang), pengfeiwei@nwpu.edu.cn (Pengfei Wei), matthias.faes@tu-dortmund.de (Matthias G.R. Faes), beer@irz.uni-hannover.de (Michael Beer)

reduction of uncertainties in both physical and engineering systems. Uncertainties occur when all or some aspects of the system under consideration are not exactly known. Examples of such aspects include, e.g., system parameters and operating conditions. These uncertainties generally originate from a variety of sources such as inherent variation, manufacturing error, modelling assumptions or a combination hereof. In terms of the origin of uncertainties, they are typically classified into either aleatory or epistemic types [1, 2]. Aleatory uncertainty refers to the uncertainty due to the intrinsic randomness or variability, and thus is irreducible in nature. As such, aleatory uncertainty is an inherent property of the system under consideration. Epistemic uncertainty, on the other hand, is associated with a lack of knowledge (or information) on the side of the analysts, and hence can be potentially reduced or even eliminated by acquiring more knowledge. Commonly, these two types of uncertainties occur together in both science and engineering, and many different uncertainty models might appear simultaneously in just one single problem. In addition to characterizing these uncertainties with appropriate mathematical models, uncertainty propagation through a computational model has also been of central interest from both academia and industry.

Many approaches have been indeed developed to quantitatively describe uncertain phenomena, which can be broadly categorized into three major groups: probabilistic approach, non-probabilistic approach and imprecise probability approach. The probabilistic approach is rooted in classical probability theory, and is the most traditional way to quantify uncertainties. Following this approach, non-determinism is modelled by a precise probability distribution on the basis of a set of probability axioms [3]. Despite its rigor in theory and popularity in practical applications, it is often criticized that the probabilistic approach indispensably relies on very fine information, e.g., a large amount of high-quality data, which is not always available. Alternatively, the non-probabilistic approach, including interval models [4], fuzzy sets [5] and convex models [6], is emerging for characterizing uncertainty with limited information, where the variation bounds need to be specified, instead of a precise probability distribution. However, it is argued that these methods are mostly suitable to deal with epistemic uncertainty. In recent years, the imprecise probability approach has gained increasing attention as a promising framework to quantify complex uncertainties, particularly when the available information or data is not sufficient to identify a unique probability distribution [7]. In essence, it is an extension to classical probability theory where the the uncertainty is characterised by a set of probability measures, rather than a single one. Therefore, it allows for modelling both aleatory uncertainty and epistemic uncertainty separately within a uniform framework. Typically, the aleatory uncertainty is characterized by the traditional probabilistic models, and the epistemic uncertainty is handled by the non-probabilistic models. Representative techniques include the probability box (p-box) [8], evidence theory [9] and fuzzy probability [10] among others.

As for uncertainty propagation, great efforts have been made along each line of uncertainty characterization over the past several decades. The existing approaches for propagating precise probabilistic uncertainty can be roughly divided into five categories: stochastic simulation methods [11–13], approximate analytical

methods [14, 15], surrogate-assisted methods [16–18], numerical integration methods [19–23] and probability conservation-based methods [24, 25]. Differently, the propagation of non-probabilistic uncertainty follows another district philosophy, more relaying on, e.g., interval arithmetic [26], optimization methods [27, 28], perturbation methods [29, 30] and etc. Also advanced sampling approaches for interval analysis have been introduced [31, 32]. One can refer to [5] for a good review on recent trends in propagation of non-probabilistic uncertainty. For imprecise probability propagation, however, the above two kinds of methods are not suitable, and hence new developments are necessary. The most common way to address the problem involves a double-loop procedure that uses the aforementioned two types of methods in a nested way, such as optimized parameter sampling [33] and interval Monte Carlo simulation [34], which often suffers from a heavy computational burden. To improve the computational efficiency, decoupled strategies have recently attracted increasing attention, and representative works include the augmented subset simulation [35], non-intrusive imprecise stochastic simulation [36, 37], operator norm theory [38], active learning augmented probabilistic integration [39], non-intrusive imprecise probabilistic integration (NIPI) [40], and collaborative and adaptive Bayesian optimization (CABO) [41]. For an review of the computation methods for propagating p-boxes, the reader is referred to [42]. Besides, some progress has also been made in the context of hybrid uncertainty propagation, e.g., surrogate modelling-based methods [43–48], stochastic simulation-based methods [49–51] and others [52, 53]. For propagating probabilistic-interval hybrid uncertainty, one can refer to the review [54]. Overall, propagation of hybrid uncertainties poses a more significant computational challenge in UQ community, and the existing mythologies are far from desirable for general practical applications.

In this paper, a novel method is presented to propagate hybrid uncertainties in the form of precise probabilistic models, parameterized p-box models and interval models, where the response expectation function (REF), its variable importance and bounds are of concern. The method belongs to the class of Bayesian probabilistic numerical methods [55], and can also be seen as an important extension to the NIPI [40] and CABO [41] methods originally developed for propagating parameterized p-box models. The main contributions of the present work can be summarized as follows:

- A general Bayesian framework is presented for propagating hybrid uncertainties, which is non-intrusive and fully decoupled in nature;
- Posterior means and variances of the REF and its random-sampling high-dimensional model representation (RS-HDMR) decomposition are analytically derived in closed form;
- Parallelized Bayesian experiment design is realized so as to take advantage of parallel computing at each iteration;
- A Matlab implementation of our methodology is freely available to the public [1].

[1] to be released upon acceptance of the paper

The remaining of this paper is organized as follows. We start by stating the problem to be solved in this study in Section 2. Section 3 presents the theoretical basis and numerical implementation procedure of the proposed method, with the relationship to the existing NIPI and CABO methods being discussed. How to extend the proposed method to a relatively more general case of hybrid uncertainties is briefly explained in Section 4. In Section 5, three numerical examples are studied to demonstrate the proposed method. The paper ends with some concluding remarks and perspectives in Section 6.

2. Problem statement

In this work, three kinds of uncertainty characterization models are considered to model non-deterministic inputs of a computer simulator, i.e., precise probability models, parameterized p-box models and interval models. The precise probability models that are deeply rooted in probability theory are assumed to be used for describing pure aleatory uncertainty. As a representative of imprecise probabilities, the parameterized p-box models are able to account for both aleatory uncertainty and epistemic uncertainty simultaneously. The interval models serve as a representative of non-probabilistic models and are useful to model the constant-but-unknown epistemic uncertainty. As such, the developed method is expected to work in the following four cases:

Case I: Precise probabilistic models and parameterized p-box models coexist in the model inputs;

Case II: Only parameterized p-box models exist in the model inputs;

Case III: Precise probabilistic models and interval models coexist in the model inputs;

Case IV: Precise probabilistic models, parameterized p-box models and interval models coexist in the model inputs.

Among the four cases, **Case IV** constitutes a more general situation of hybrid uncertainties. For notational clarity, however, we only take **Case III** as an example to illustrate the proposed method in the following, and when it comes to the general case (i.e., **Case IV**) one can refer to Section 4. Let $\boldsymbol{X} = [X_1, X_2, \ldots, X_{d_1}] \in \mathscr{X} \subseteq \mathbb{R}^{d_1}$ and $\boldsymbol{A} = [A_1, A_2, \ldots, A_{d_2}] \in \mathscr{A} \subseteq \mathbb{R}^{d_2}$ denote a d_1-dimensional vector of precise random variables and a d_2-dimensional vector of interval variables, respectively. The random variables are said to be 'precise' when their distribution types and distribution parameters are exactly known, and we assume that the joint probability density function (PDF) of $\boldsymbol{X}$ exists, denoted as $f_{\boldsymbol{X}}(\boldsymbol{x})$. The interval variables refer to the uncertain parameters with limited information, and can only be specified by their lower and upper bounds, i.e., $\boldsymbol{A} = [\underline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\alpha}}]$, where $\underline{\boldsymbol{\alpha}} = [\underline{\alpha}_1, \underline{\alpha}_2, \ldots, \underline{\alpha}_{d_2}]$ and $\overline{\boldsymbol{\alpha}} = [\overline{\alpha}_1, \overline{\alpha}_2, \cdots, \overline{\alpha}_{d_2}]$. As such, $\boldsymbol{A}$ represents a d_2-dimensional hyper-rectangle. In this study, these $d_1 + d_2$ variables are assumed to be independent just for the convenience of describing our method. The computer simulator is represented by a deterministic, continuous and real-valued function $g : \mathbb{R}^{d_1+d_2} \mapsto \mathbb{R}, \{x, \alpha\} \to z$, with $Z = g(\boldsymbol{X}, \boldsymbol{A})$ being a scalar quantity of interest. Due to the existence of interval variables, Z is no longer a random variable

unless $\boldsymbol{A}$ is fixed at a value $\boldsymbol{\alpha} \in \boldsymbol{A}$. Thus, the expectation of Z, is not a deterministic values anymore, but function of the interval variables. More precisely, it only assume a crisp value for a realisation of the input intervals. To formalize, the definition of the so-called REF is given as follows:

$$m(\boldsymbol{\alpha}) = \int_{\mathscr{X}} g(\boldsymbol{x}, \boldsymbol{\alpha}) f_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}, \tag{1}$$

The lower and upper bounds of $m(\boldsymbol{\alpha})$ can be defined as:

$$m_l = \min_{\boldsymbol{\alpha} \in [\underline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\alpha}}]} m(\boldsymbol{\alpha}) = \min_{\boldsymbol{\alpha} \in [\underline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\alpha}}]} \int_{\mathscr{X}} g(\boldsymbol{x}, \boldsymbol{\alpha}) f_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}, \tag{2}$$

$$m_u = \max_{\boldsymbol{\alpha} \in [\underline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\alpha}}]} m(\boldsymbol{\alpha}) = \max_{\boldsymbol{\alpha} \in [\underline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\alpha}}]} \int_{\mathscr{X}} g(\boldsymbol{x}, \boldsymbol{\alpha}) f_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}. \tag{3}$$

The REF can provide complete information about how the response expectation changes with its argument $\boldsymbol{\alpha}$, whereas the interval $[m_l, m_u]$ measures the amount of epistemic uncertainty present in the response expectation. Besides, the analyst may also concern the variable importance of the REF. Intuitively, the bounds and variable importance analysis of the REF can be proceeded straightforwardly once the REF is available. However, it is still a non-trivial task to compute the REF in an efficient manner since each evaluation of the response function $g(\boldsymbol{x}, \boldsymbol{\alpha})$ can be prohibitively expensive for a real-world problem.

3. Parallel Bayesian quadrature optimization

As the REF defined in Eq. (1) is given in the form of an integral, the Bayesian probabilistic integration (BPI) [23] can be applied to efficiently obtain an estimate for the REF. If we assign a Gaussian process (GP) prior for the integrand $g(\boldsymbol{x}, \boldsymbol{\alpha})$, the induced posterior of the REF is also a GP. Following this, the lower and upper bounds defined in Eqs. (2) and (3) may be further solved by the Bayesian global optimization (BGO) [56]. In this section, a novel Bayesian approach combining the BPI and BGO, called Parallel Bayesian Quadrature Optimization (PBQO), is presented to produce the REF, its variable importance and bounds simultaneously in an efficient manner.

3.1. Variable transformation

Before introducing our method, a pre-processing step should be performed to transform the original input variable vector $\{\boldsymbol{X}, \boldsymbol{A}\}$ to a new one so as to make the proposed method analytically tractable. In this study, the random variable vector $\boldsymbol{X}$ is transformed to be a standard normal one by a certain transformation (e.g., isoprobabilistic transformation), which is denoted as $\boldsymbol{U} = T_1(\boldsymbol{X})$. In contrast, we consider transforming the interval vector $\boldsymbol{A}$ to be a standard one (i.e., $[0, 1]^{d_2}$) by a simple linear transformation such that $\boldsymbol{V} = T_2(\boldsymbol{A})$. For convenience, the two transformations can be written in a uniform form $\boldsymbol{W} = T(\boldsymbol{X}, \boldsymbol{A})$, where $\boldsymbol{W} = \{\boldsymbol{U}, \boldsymbol{V}\}$. The REF with respect to $\boldsymbol{v}$ is defined as:

$$\mathcal{M}(\boldsymbol{v}) = \int_{\mathscr{U}} \mathcal{G}(\boldsymbol{w}) f_{\boldsymbol{U}}(\boldsymbol{u}) \mathrm{d}\boldsymbol{u}, \tag{4}$$

where $\mathcal{G}(\boldsymbol{w}) = g(T(\boldsymbol{x}, \boldsymbol{\alpha}))$, $f_{\boldsymbol{U}}(\boldsymbol{u})$ is the joint PDF of $\boldsymbol{U}$. Once $\mathcal{M}(\boldsymbol{v})$ is available, $m(\boldsymbol{\alpha})$ can be easily obtained as $m(\boldsymbol{\alpha}) = \mathcal{M}(T_2(\boldsymbol{\alpha}))$. Note that the T_1 transformation is necessary for the analytical tractability of the proposed method, while T_2 transformation is not. However, we introduce the T_2 transformation only for the purpose of producing concise analytical expressions.

3.2. Prior Gaussian process

In the proposed PBQO method, we first place a GP prior over the space $\mathscr{G}$ of functions: $\mathcal{G} : \mathcal{W} \to \mathbb{R}$, denoted as $\hat{\mathcal{G}}(\boldsymbol{w}) \sim \mathcal{GP}(\mu_0(\boldsymbol{w}), k_0(\boldsymbol{w}, \boldsymbol{w}'))$, where $\mu_0(\boldsymbol{w})$ and $k_0(\boldsymbol{w}, \boldsymbol{w}')$ are the prior mean and covariance functions, respectively. The prior mean function reflects the general trend of the GP, and can be assumed to be, e.g., zero, constant or a linear polynomial. The covariance function is a more crucial ingredient of the GP since it encodes our basic assumptions about the function to be inferred, e.g., smoothness and periodicity. In this study, the prior mean function adopts a constant, i.e., $\mu_0(\boldsymbol{w}) = \beta$, and the prior covariance function takes the squared exponential kernel:

$$\begin{aligned} k_0(\boldsymbol{w}, \boldsymbol{w}') =& s_0^2 \exp\left[-\frac{1}{2}(\boldsymbol{w}-\boldsymbol{w}')\boldsymbol{\Sigma}^{-1}(\boldsymbol{w}-\boldsymbol{w}')^{\mathrm{T}}\right] \\ =& s_0^2 \exp\left[-\frac{1}{2}(\boldsymbol{u}-\boldsymbol{u}')\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(\boldsymbol{u}-\boldsymbol{u}')^{\mathrm{T}}\right] \exp\left[-\frac{1}{2}(\boldsymbol{v}-\boldsymbol{v}')\boldsymbol{\Sigma}_{\boldsymbol{v}}^{-1}(\boldsymbol{v}-\boldsymbol{v}')^{\mathrm{T}}\right], \end{aligned} \tag{5}$$

where s_0^2 is the process variance, $\boldsymbol{\Sigma} = \mathrm{diag}\left\{l_1^2, l_2^2, \cdots, l_{d_1+d_2}^2\right\}$ with l_i being the characteristic lengthscale in i-th dimension, $\boldsymbol{\Sigma}_{\boldsymbol{u}} = \mathrm{diag}\left\{l_1^2, l_2^2, \cdots, l_{d_1}^2\right\}$ and $\boldsymbol{\Sigma}_{\boldsymbol{v}} = \mathrm{diag}\left\{l_{d_1+1}^2, l_{d_1+2}^2, \cdots, l_{d_1+d_2}^2\right\}$; Throughout the paper, the symbol $\mathrm{diag}\{\cdot\}$ means to create a square diagonal matrix with the elements of its argument when its argument is a vector or to get a column vector of the diagonal elements of its argument when its argument is a matrix. The parameters β, s_0, $l_1, l_2, \cdots, l_{d_1+d_2}$ are called hyperparameters. Note that the analytical tractability of the proposed method relies on using the squared exponential kernel.

3.3. Bayesian posterior inference

Suppose that we have evaluated the $\mathcal{G}$-function at n points. Let a $n \times (d_1 + d_2)$ matrix $\boldsymbol{\mathcal{W}} = (\boldsymbol{\mathcal{U}}, \boldsymbol{\mathcal{V}}) = \left\{\boldsymbol{w}^{(j)}\right\}_{j=1}^{n}$ denote the n points at which the $\mathcal{G}$-function are evaluated, and a $n \times 1$ matrix $\boldsymbol{\mathcal{Z}} = \left\{z^{(j)}\right\}_{j=1}^{n}$ denote the corresponding $\mathcal{G}$-function values at $\boldsymbol{\mathcal{W}}$. Given $\boldsymbol{\mathcal{D}} = \{\boldsymbol{\mathcal{W}}, \boldsymbol{\mathcal{Z}}\}$, the hyperparameters involved in the prior mean and covariance functions can be determined by, e.g., maximum likelihood estimation [57]. Besides, conditioning on the data $\boldsymbol{\mathcal{D}}$, we can arrive at a posterior GP over functions $\mathcal{G} \in \mathscr{G}$, which is denoted as $\mathcal{GP}(\mu_n(\boldsymbol{w}), k_n(\boldsymbol{w}, \boldsymbol{w}'))$. According to [57], the posterior mean $\mu_n(\boldsymbol{w})$ and posterior covariance function $k_n(\boldsymbol{w}, \boldsymbol{w}')$ can be given by:

$$\mu_n(\boldsymbol{w}) = \mu_0(\boldsymbol{w}) + \boldsymbol{k}_0(\boldsymbol{w}, \boldsymbol{\mathcal{W}})^{\mathrm{T}} \boldsymbol{K}_0^{-1}(\boldsymbol{\mathcal{Z}} - \boldsymbol{\mu}_0(\boldsymbol{\mathcal{W}})), \tag{6}$$

$$k_n(\boldsymbol{w}, \boldsymbol{w}') = k_0(\boldsymbol{w}, \boldsymbol{w}') - \boldsymbol{k}_0(\boldsymbol{w}, \boldsymbol{\mathcal{W}})^{\mathrm{T}} \boldsymbol{K}_0^{-1} \boldsymbol{k}_0(\boldsymbol{w}', \boldsymbol{\mathcal{W}}), \tag{7}$$

where $\boldsymbol{\mu}_0(\boldsymbol{\mathcal{W}}) = [\mu_0(\boldsymbol{w}^{(1)}), \mu_0(\boldsymbol{w}^{(2)}), \cdots, \mu_0(\boldsymbol{w}^{(n)})]^{\mathrm{T}}$ is the mean vector at $\boldsymbol{\mathcal{W}}$; $\boldsymbol{k}_0(\boldsymbol{w}, \boldsymbol{\mathcal{W}}) = [k_0(\boldsymbol{w}, \boldsymbol{w}^{(1)}), k_0(\boldsymbol{w}, \boldsymbol{w}^{(2)}), \cdots, k_0(\boldsymbol{w}, \boldsymbol{w}^{(n)})]^{\mathrm{T}}$ is the covariance vector between $\boldsymbol{w}$ and $\boldsymbol{\mathcal{W}}$; $\boldsymbol{k}_0(\boldsymbol{w}', \boldsymbol{\mathcal{W}}) = [k_0(\boldsymbol{w}', \boldsymbol{w}^{(1)}), k_0(\boldsymbol{w}', \boldsymbol{w}^{(2)}), \cdots, k_0(\boldsymbol{w}', \boldsymbol{w}^{(n)})]^{\mathrm{T}}$ is the covariance vector between $\boldsymbol{w}'$ and $\boldsymbol{\mathcal{W}}$; $\boldsymbol{K}_0$ is the covariance matrix of $\boldsymbol{\mathcal{W}}$ with entry $[\boldsymbol{K}_0]_{ij} = k_0(\boldsymbol{w}^{(i)}, \boldsymbol{w}^{(j)})$.

3.3.1. Bayesian inference of REF

As an extended result of BPI [58], the posterior distribution of REF (denoted as $\hat{\mathcal{M}}(\boldsymbol{v})$), i.e., integrating $\hat{\mathcal{G}}(\boldsymbol{w})$ with respect to $\boldsymbol{u}$ under the Gaussian weight $f_{\boldsymbol{U}}(\boldsymbol{u})$, still follows a GP. By repeated application of Fubini's theorem, one can derive the analytical expressions of the posterior mean function $\mu_{\hat{\mathcal{M}}}(\boldsymbol{v})$ and posterior variance function $\sigma^2_{\hat{\mathcal{M}}}(\boldsymbol{v})$ such that:

$$\mu_{\hat{\mathcal{M}}}(\boldsymbol{v}) = \mathbb{E}_{\boldsymbol{\mathcal{D}}}\left[\hat{\mathcal{M}}(\boldsymbol{v})\right] = \Pi_{\boldsymbol{u}}\left[\mu_0(\boldsymbol{w})\right] + \Pi_{\boldsymbol{u}}\left[\boldsymbol{k}_0(\boldsymbol{w}, \boldsymbol{\mathcal{W}})^{\mathrm{T}}\right] \boldsymbol{K}_0^{-1}\left(\boldsymbol{\mathcal{Z}} - \boldsymbol{\mu}_0(\boldsymbol{\mathcal{W}})\right), \tag{8}$$

$$\sigma^2_{\hat{\mathcal{M}}}(\boldsymbol{v}) = \mathbb{V}_{\boldsymbol{\mathcal{D}}}\left[\hat{\mathcal{M}}(\boldsymbol{v})\right] = \Pi_{\boldsymbol{u}}\Pi_{\boldsymbol{u}'}\left[k_0(\boldsymbol{w}, (\boldsymbol{u}', \boldsymbol{v}))\right] - \Pi_{\boldsymbol{u}}\left[\boldsymbol{k}_0(\boldsymbol{w}, \boldsymbol{\mathcal{W}})^{\mathrm{T}}\right] \boldsymbol{K}_0^{-1}\Pi_{\boldsymbol{u}}\left[\boldsymbol{k}_0(\boldsymbol{w}, \boldsymbol{\mathcal{W}})\right], \tag{9}$$

where $\mathbb{E}_{\boldsymbol{\mathcal{D}}}[\cdot]$ and $\mathbb{V}_{\boldsymbol{\mathcal{D}}}[\cdot]$ refer to the expectation and variance operators taken with respect to the posterior distributions of their arguments given data $\boldsymbol{\mathcal{D}}$; $\Pi_{\boldsymbol{u}}[\cdot]$ denotes the integral operator taken with respect to $\boldsymbol{u}$ under Gaussian weight $f_{\boldsymbol{U}}(\boldsymbol{u})$; $\Pi_{\boldsymbol{u}'}[\cdot]$ is similarly defined; $\Pi_{\boldsymbol{u}}\Pi_{\boldsymbol{u}'}[\cdot]$ is the integral operator taken respect to both $\boldsymbol{u}$ and $\boldsymbol{u}'$ under Gaussian weights $f_{\boldsymbol{U}}(\boldsymbol{u})$ and $f_{\boldsymbol{U}}(\boldsymbol{u}')$; The term $\Pi_{\boldsymbol{u}}\left[\mu_0(\boldsymbol{w})\right]$ can be easily obtained as $\Pi_{\boldsymbol{u}}\left[\mu_0(\boldsymbol{w})\right] = \beta$; The other terms can be derived as:

$$\Pi_{\boldsymbol{u}}\left[\boldsymbol{k}_0(\boldsymbol{w}, \boldsymbol{\mathcal{W}})\right] = s_0^2\left|\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1} + I\right|^{-1/2} \exp\left[-\frac{1}{2}\mathrm{diag}\left\{\boldsymbol{\mathcal{U}}\left(\boldsymbol{\Sigma}_{\boldsymbol{u}} + \boldsymbol{I}\right)^{-1}\boldsymbol{\mathcal{U}}^{\mathrm{T}} - (\boldsymbol{v} - \boldsymbol{\mathcal{V}})\,\boldsymbol{\Sigma}_{\boldsymbol{v}}^{-1}(\boldsymbol{v} - \boldsymbol{\mathcal{V}})^{\mathrm{T}}\right\}\right], \tag{10}$$

$$\Pi_{\boldsymbol{u}}\Pi_{\boldsymbol{u}'}\left[k_0(\boldsymbol{w}, \boldsymbol{w}')\right] = s_0^2\left|2\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1} + \boldsymbol{I}\right|^{-1/2}, \tag{11}$$

where $|\cdot|$ means the determinant of its argument; $\boldsymbol{I}$ is a identity matrix of size d_1.

Note that the expressions for $\mu_{\hat{\mathcal{M}}}(\boldsymbol{v})$ and $\sigma^2_{\hat{\mathcal{M}}}(\boldsymbol{v})$ are similar in form to those of NIPI and CABO, but essentially different due to the fact that the proposed method is established on the basis of the joint space of standard normal variables and standard interval variables, while both NIPI and CABO are cast in the standard normal space. The posterior mean function $\mu_{\mathcal{M},n}(\boldsymbol{v})$ can be used as an estimate of $\mathcal{M}(\boldsymbol{v})$ and the posterior variance function $\sigma^2_{\hat{\mathcal{M}}}(\boldsymbol{v})$ measures our uncertainty of the estimate after n observations have been available. By using the linear transformation, one can easily obtain the posterior mean function $\mu_{\hat{m}}(\boldsymbol{\alpha}) = \mu_{\hat{\mathcal{M}}}(T_2(\boldsymbol{\alpha}))$ and posterior variance function $\sigma^2_{\hat{m}}(\boldsymbol{\alpha}) = \sigma^2_{\hat{\mathcal{M}}}(T_2(\boldsymbol{\alpha}))$ for $\hat{m}(\boldsymbol{\alpha})$.

3.3.2. Bayesian inference of RS-HDMR component functions of REF

In addition to the REF $\hat{m}(\boldsymbol{\alpha})$, the analyst may also be concerned about, e.g., identifying key variables among $\boldsymbol{A}$ that are more important for $m(\boldsymbol{\alpha})$. For this propose, the RS-HDMR is first employed to express

$\mathcal{M}(\boldsymbol{v})$ as the summation of a set of component functions with increasing dimensions [59]:

$$\mathcal{M}(\boldsymbol{v}) = \mathcal{M}_0 + \sum_{i=1}^{d_2} \mathcal{M}_i(v_i) + \sum_{1\le i<j\le d_2} \mathcal{M}_{ij}(v_i, v_j) + \cdots + \mathcal{M}_{ij\ldots d_2}(v_1, v_2, \cdots, v_{d_2}), \tag{12}$$

where the zeroth-order component function $\mathcal{M}_0$ is a constant representing the average value of $\mathcal{M}(\boldsymbol{v})$ over the entire domain $\mathcal{V}$, the first-order component function $\mathcal{M}_i(v_i)$ represents the independent contribution of v_i acting alone to $\mathcal{M}(\boldsymbol{v})$, the second-order component function $\mathcal{M}_{ij}(v_i, v_j)$ denotes the cooperative effects of v_i and v_j upon $\mathcal{M}(\boldsymbol{v})$, etc. The last term $\mathcal{M}_{ij\ldots d_2}(v_1, v_2, \cdots, v_{d_2})$ describes any residual cooperative effects of all input variables acting together to influence $\mathcal{M}(\boldsymbol{v})$. The component functions up to the second-order can be defined as:

$$\mathcal{M}_0 = \int_{\mathcal{V}} \mathcal{M}(\boldsymbol{v}) \mathrm{d}\boldsymbol{v} = \int_{\mathcal{V}} \int_{\mathcal{U}} \mathcal{G}(\boldsymbol{w}) f_{\boldsymbol{U}}(\boldsymbol{u}) \mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v}, \tag{13}$$

$$\mathcal{M}_i(v_i) = \int_{\mathcal{V}_{-i}} \mathcal{M}(\boldsymbol{v}) \mathrm{d}\boldsymbol{v}_{-i} - \mathcal{M}_0 = \int_{\mathcal{V}_{-i}} \int_{\mathcal{U}} \mathcal{G}(\boldsymbol{w}) f_{\boldsymbol{U}}(\boldsymbol{u}) \mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v}_{-i} - \mathcal{M}_0, \tag{14}$$

$$\mathcal{M}_{ij}(v_i, v_j) = \int_{\mathcal{V}_{-ij}} \mathcal{M}(\boldsymbol{v}) \mathrm{d}\boldsymbol{v}_{-ij} - \mathcal{M}_i(v_i) - \mathcal{M}_j(v_j) - \mathcal{M}_0 = \int_{\mathcal{V}_{-ij}} \int_{\mathcal{U}} \mathcal{G}(\boldsymbol{w}) f_{\boldsymbol{U}}(\boldsymbol{u}) \mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v}_{-ij} - \mathcal{M}_i(v_i) - \mathcal{M}_j(v_j) - \mathcal{M}_0, \tag{15}$$

where $\mathcal{V}_{-i}$ and $\boldsymbol{v}_{-i}$ denote the space $\mathcal{V}$ and the vector $\boldsymbol{v}$ excluding the i-th dimension, respectively; $\mathcal{V}_{-ij}$ and $\boldsymbol{v}_{-ij}$ are similarly defined.

As high-order component functions have small contributions for many realistic systems, the second-order truncated RS-HDMR expansion is often considered [36, 40]. For this reason, only the component functions up to the second-order are provided in the following via Bayesian inference. If necessary, high-order component functions can also be derived similarly.

Zeroth-order RS-HDMR component. As defined in Eq. (13), the zeroth-order RS-HDMR component $\mathcal{M}_0$ is actually an integral of $\mathcal{G}(\boldsymbol{w})$ with respect to $\boldsymbol{w}$. From a Bayesian quadrature perspective, the posterior distribution of $\mathcal{M}_0$ (denoted as $\hat{\mathcal{M}}_0$) is Gaussian with posterior mean $\mu_{\hat{\mathcal{M}}_0}$ and posterior variance $\sigma^2_{\hat{\mathcal{M}}_0}$ being:

$$\mu_{\hat{\mathcal{M}}_0} = \mathbb{E}_{\boldsymbol{\mathcal{D}}}\left[\hat{\mathcal{M}}_0\right] = \Pi_{\boldsymbol{w}}\left[\mu_0(\boldsymbol{w})\right] + \Pi_{\boldsymbol{w}}\left[\boldsymbol{k}_0\left(\boldsymbol{w}, \boldsymbol{\mathcal{W}}\right)^{\mathrm{T}}\right] \boldsymbol{K}_0^{-1}\left(\boldsymbol{\mathcal{Z}} - \boldsymbol{\mu}_0\left(\boldsymbol{\mathcal{W}}\right)\right), \tag{16}$$

$$\sigma^2_{\hat{\mathcal{M}}_0} = \mathbb{V}_{\boldsymbol{\mathcal{D}}}\left[\hat{\mathcal{M}}_0\right] = \Pi_{\boldsymbol{w}}\Pi_{\boldsymbol{w}'}\left[k_0(\boldsymbol{w}, \boldsymbol{w}')\right] - \Pi_{\boldsymbol{w}}\left[\boldsymbol{k}_0\left(\boldsymbol{w}, \boldsymbol{\mathcal{W}}\right)^{\mathrm{T}}\right] \boldsymbol{K}_0^{-1} \Pi_{\boldsymbol{w}'}\left[\boldsymbol{k}_0\left(\boldsymbol{w}', \boldsymbol{\mathcal{W}}\right)\right], \tag{17}$$

where $\Pi_{\boldsymbol{w}}\left[\mu_0(\boldsymbol{w})\right] = \beta$, and other terms can be derived as:

$$\begin{aligned}
\Pi_{\boldsymbol{w}}\left[\boldsymbol{k}_0\left(\boldsymbol{w}, \boldsymbol{\mathcal{W}}\right)\right] =& \Pi_{\boldsymbol{w}'}\left[\boldsymbol{k}_0\left(\boldsymbol{w}', \boldsymbol{\mathcal{W}}\right)\right] \\
=& s_0^2 \left|\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1} + \boldsymbol{I}\right|^{-1/2} \exp\left[-\frac{1}{2}\mathrm{diag}\left\{\boldsymbol{\mathcal{U}}\left(\boldsymbol{\Sigma}_{\boldsymbol{u}} + \boldsymbol{I}\right)^{-1}\boldsymbol{\mathcal{U}}^{\mathrm{T}}\right\}\right] \\
& \cdot \left(\frac{\pi}{2}\right)^{d_2/2} \mathrm{prod}_2\left\{\left[\mathrm{erf}\left(\left(1 - \boldsymbol{\mathcal{V}}\right)\left(2\boldsymbol{\Sigma}_{\boldsymbol{v}}\right)^{-1/2}\right) - \mathrm{erf}\left(-\boldsymbol{\mathcal{V}}\left(2\boldsymbol{\Sigma}_{\boldsymbol{v}}\right)^{-1/2}\right)\right] \boldsymbol{\Sigma}_{\boldsymbol{v}}^{1/2}\right\},
\end{aligned} \tag{18}$$

$$\begin{aligned}\Pi_{\boldsymbol{w}}\Pi_{\boldsymbol{w}'}\left[k_0(\boldsymbol{w},\boldsymbol{w}')\right] =& s_0^2\left|2\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}+\boldsymbol{I}\right|^{-1/2} \\ & \cdot 2^{d_2}\mathrm{prod}_1\left\{\mathrm{diag}\left\{\boldsymbol{\Sigma}_{\boldsymbol{v}}\left[-1+\exp\left[-(2\boldsymbol{\Sigma}_{\boldsymbol{v}})^{-1}\right]+(2\pi^{-1}\boldsymbol{\Sigma}_{\boldsymbol{v}})^{-1/2}\mathrm{erf}\left((2\boldsymbol{\Sigma}_{\boldsymbol{v}})^{-1/2}\right)\right]\right\}\right\},\end{aligned} \tag{19}$$

where $\mathrm{prod}_1\{\cdot\}$ means to return the product of the elements of its argument; $\mathrm{prod}_2\{\cdot\}$ is to get a column vector containing the products of each row of its argument; $\mathrm{erf}(\cdot)$ stands for the error function. Note that in Eq. (18) the argument in $\mathrm{prod}_2\{\cdot\}$ is an n-by-d_2 matrix, while in Eq. (19) the argument in $\mathrm{prod}_1\{\cdot\}$ is a d_2-by-1 vector.

First-order RS-HDMR component. The first-order RS-HDMR component function $\mathcal{M}_i(v_i)$ defined in Eq. (14) is an integral (i.e, integrating $\mathcal{G}(\boldsymbol{w})$ with respect to $\boldsymbol{w}$ excluding v_i) minus $\mathcal{M}_0$, and thus its posterior distribution $\hat{\mathcal{M}}_i(v_i)$ should follow a one-dimensional GP.

The posterior mean function $\mu_{\hat{\mathcal{M}}_i}(v_i)$ of the first-order RS-HDMR component function $\hat{\mathcal{M}}_i(v_i)$ can be expressed as:

$$\mu_{\hat{\mathcal{M}}_i}(v_i)=\mathbb{E}_{\boldsymbol{\mathcal{D}}}\left[\hat{\mathcal{M}}_i(v_i)\right]=\Pi_{-v_i}\left[\mu_0(\boldsymbol{w})\right]+\Pi_{-v_i}\left[\boldsymbol{k}_0\left(\boldsymbol{w},\boldsymbol{\mathcal{W}}\right)^{\mathrm{T}}\right]\boldsymbol{K}_0^{-1}\left(\boldsymbol{\mathcal{Z}}-\boldsymbol{\mu}_0\left(\boldsymbol{\mathcal{W}}\right)\right)-\mu_{\hat{\mathcal{M}}_0}, \tag{20}$$

where $\Pi_{-v_i}[\cdot]$ denotes the integration of its argument taken over $\boldsymbol{w}$ except v_i; it is obvious that $\Pi_{-v_i}\left[\mu_0(\boldsymbol{w})\right]=\beta$; the term $\Pi_{-v_i}\left[\boldsymbol{k}_0\left(\boldsymbol{w},\boldsymbol{\mathcal{W}}\right)\right]$ can be derived as:

$$\begin{aligned}\Pi_{-v_i}\left[\boldsymbol{k}_0\left(\boldsymbol{w},\boldsymbol{\mathcal{W}}\right)\right] =& s_0^2\left|\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}+\boldsymbol{I}\right|^{-1/2}\exp\left[-\frac{1}{2}\mathrm{diag}\left\{\boldsymbol{\mathcal{U}}\left(\boldsymbol{\Sigma}_{\boldsymbol{u}}+\boldsymbol{I}\right)^{-1}\boldsymbol{\mathcal{U}}^{\mathrm{T}}\right\}\right] \\ & \cdot\left(\frac{\pi}{2}\right)^{(d_2-1)/2}\mathrm{prod}_2\left\{\left[\mathrm{erf}\left(\left(1-\boldsymbol{\mathcal{V}}_{,-i}\right)\left(2\boldsymbol{\Sigma}_{\boldsymbol{v}_{-i}}\right)^{-1/2}\right)-\mathrm{erf}\left(-\boldsymbol{\mathcal{V}}_{,-i}\left(2\boldsymbol{\Sigma}_{\boldsymbol{v}_{-i}}\right)^{-1/2}\right)\right]\boldsymbol{\Sigma}_{\boldsymbol{v}_{-i}}^{1/2}\right\} \\ & \cdot\exp\left[-\frac{1}{2}\mathrm{diag}\left\{-\left(v_i-\boldsymbol{\mathcal{V}}_{,i}\right)\boldsymbol{\Sigma}_{\boldsymbol{v}_i}^{-1}\left(v_i-\boldsymbol{\mathcal{V}}_{,i}\right)^{\mathrm{T}}\right\}\right],\end{aligned} \tag{21}$$

in which $\boldsymbol{\mathcal{V}}_{,i}$ is the i-th column of $\boldsymbol{\mathcal{V}}$, $\boldsymbol{\mathcal{V}}_{,-i}$ represents the matrix generated by removing $\boldsymbol{\mathcal{V}}_{,i}$ from $\boldsymbol{\mathcal{V}}$, $\boldsymbol{\Sigma}_{\boldsymbol{v}_i}$ denotes the (i,i)-th element of $\boldsymbol{\Sigma}_{\boldsymbol{v}}$, and $\boldsymbol{\Sigma}_{\boldsymbol{v}_{-i}}$ stands for the matrix generated by removing the i-th column and i-th row of $\boldsymbol{\Sigma}_{\boldsymbol{v}}$.

For the posterior variance function $\sigma^2_{\hat{\mathcal{M}}_i}(v_i)$ of the first-order RS-HDMR component function $\hat{\mathcal{M}}_i(v_i)$, one can refer to Appendix A.

Second-order RS-HDMR component. Similarly, the second-order RS-HDMR component function $\mathcal{M}_{ij}(v_i,v_j)$ defined in Eq. (15) is an integral (i.e., integrating $\mathcal{G}(\boldsymbol{w})$ with respect to $\boldsymbol{w}$ excluding v_i and v_j) diminished by $\mathcal{M}_i(v_i)$, $\mathcal{M}_j(v_j)$ and $\mathcal{M}_0$, and thus its posterior distribution $\hat{\mathcal{M}}_{ij}(v_i,v_j)$ should follow a two-dimensional GP.

The posterior mean function $\mu_{\hat{\mathcal{M}}_{ij}}(v_i,v_j)$ of the first-order RS-HDMR component function $\hat{\mathcal{M}}_{ij}(v_i,v_j)$

can be given by:

$$\begin{aligned}\mu_{\hat{\mathcal{M}}_{ij}}(v_i,v_j) =&\mathbb{E}_{\boldsymbol{\mathcal{D}}}\left[\hat{\mathcal{M}}_{ij}(v_i,v_j)\right]\\ =&\Pi_{-\boldsymbol{v}_{ij}}\left[\mu_0(\boldsymbol{w})\right]+\Pi_{-\boldsymbol{v}_{ij}}\left[\boldsymbol{k}_0\left(\boldsymbol{w},\boldsymbol{\mathcal{W}}\right)^{\mathrm{T}}\right]\boldsymbol{K}_0^{-1}\left(\boldsymbol{\mathcal{Z}}-\boldsymbol{\mu}_0\left(\boldsymbol{\mathcal{W}}\right)\right)-\mu_{\hat{\mathcal{M}}_i}(v_i)-\mu_{\hat{\mathcal{M}}_j}(v_j)-\mu_{\hat{\mathcal{M}}_0},\end{aligned} \tag{22}$$

where the term $\Pi_{-\boldsymbol{v}_{ij}}\left[\mu_0(\boldsymbol{w})\right]$ is equal to β, and the term $\Pi_{-\boldsymbol{v}_{ij}}\left[\boldsymbol{k}_0\left(\boldsymbol{w},\boldsymbol{\mathcal{W}}\right)\right]$ is derived as:

$$\begin{aligned}\Pi_{-\boldsymbol{v}_{ij}}\left[\boldsymbol{k}_0\left(\boldsymbol{w},\boldsymbol{\mathcal{W}}\right)\right] =&s_0^2\left|\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}+\boldsymbol{I}\right|^{-1/2}\exp\left[-\frac{1}{2}\mathrm{diag}\left\{\boldsymbol{\mathcal{U}}\left(\boldsymbol{\Sigma}_{\boldsymbol{u}}+\boldsymbol{I}\right)^{-1}\boldsymbol{\mathcal{U}}^{\mathrm{T}}\right\}\right]\\ &\cdot\left(\frac{\pi}{2}\right)^{(d_2-2)/2}\mathrm{prod}_2\left\{\left[\mathrm{erf}\left(\left(1-\boldsymbol{\mathcal{V}}_{,-ij}\right)\left(2\boldsymbol{\Sigma}_{\boldsymbol{v}_{-ij}}\right)^{-1/2}\right)-\mathrm{erf}\left(-\boldsymbol{\mathcal{V}}_{,-ij}\left(2\boldsymbol{\Sigma}_{\boldsymbol{v}_{-ij}}\right)^{-1/2}\right)\right]\boldsymbol{\Sigma}_{\boldsymbol{v}_{-ij}}^{1/2}\right\}\\ &\cdot\exp\left[-\frac{1}{2}\mathrm{diag}\left\{-\left(v_{ij}-\boldsymbol{\mathcal{V}}_{,ij}\right)\boldsymbol{\Sigma}_{\boldsymbol{v}_{ij}}^{-1}\left(v_{ij}-\boldsymbol{\mathcal{V}}_{,ij}\right)^{\mathrm{T}}\right\}\right].\end{aligned} \tag{23}$$

For the posterior variance function $\sigma^2_{\hat{\mathcal{M}}_{ij}}(v_i,v_j)$ of the second-order RS-HDMR component function $\hat{\mathcal{M}}_{ij}(v_i,v_j)$, one can refer to Appendix B. ■

One should note that the above results are essentially different from those in NIPI. Once these RS-HDMR component functions of $\hat{\mathcal{M}}(\boldsymbol{v})$ are properly inferred, they can be transformed by a linear transformation to yield the RS-HDMR component functions for $\hat{m}(\boldsymbol{\alpha})$.

3.3.3. Bayesian inference of extrema of REF

If we stop after obtaining n observations of the $\mathcal{G}$-function, a risk-neutral choice for the minimum or maximum of the REF would be the minimum or maximum of the posterior mean function $\mu_{\hat{m}}(\boldsymbol{\alpha})$. As $\mu_{\hat{m}}(\boldsymbol{\alpha})$ has been derived in a closed-form, the extrema of the REF can be inferred from $\mu_{\hat{m}}(\boldsymbol{\alpha})$ by simply applying a global optimization algorithm such that:

$$\hat{m}_l = \min_{\boldsymbol{\alpha}\in[\underline{\boldsymbol{\alpha}},\overline{\boldsymbol{\alpha}}]}\mu_{\hat{m}}(\boldsymbol{\alpha}), \tag{24}$$

$$\hat{m}_u = \max_{\boldsymbol{\alpha}\in[\underline{\boldsymbol{\alpha}},\overline{\boldsymbol{\alpha}}]}\mu_{\hat{m}}(\boldsymbol{\alpha}). \tag{25}$$

Besides, since the posterior variance function $\sigma^2_{\hat{m}}(\boldsymbol{\alpha})$ is also available, the prediction errors regarding the minimum and maximum estimators in Eqs. (24) and (25) can be measured by the posterior variances:

$$\mathrm{Var}\left[\hat{m}_l\right] = \sigma^2_{\hat{m}}(\boldsymbol{\alpha}^-), \tag{26}$$

$$\mathrm{Var}\left[\hat{m}_u\right] = \sigma^2_{\hat{m}}(\boldsymbol{\alpha}^+), \tag{27}$$

where $\boldsymbol{\alpha}^- = \arg\min_{\boldsymbol{\alpha}\in[\underline{\boldsymbol{\alpha}},\overline{\boldsymbol{\alpha}}]}\mu_{\hat{m}}(\boldsymbol{\alpha})$ and $\boldsymbol{\alpha}^+ = \arg\max_{\boldsymbol{\alpha}\in[\underline{\boldsymbol{\alpha}},\overline{\boldsymbol{\alpha}}]}\mu_{\hat{m}}(\boldsymbol{\alpha})$ are the minimum point and maximum point, respectively.

3.4. Parallel Bayesian experimental design

Another significant advantage of the above framework is that it offers the possibility for incorporating our prior knowledge and developing a Bayesian experimental design strategy. This advantage is also realized in both NIPI and CABO. These two methods, however, are in a pure sequential manner to acquire the $\mathcal{G}$-function. That is, at each iteration only one point is allowed to be selected and a single $\mathcal{G}$-function evaluation is subsequently performed. The sequential experimental strategies would be less efficient and flexible when parallel computing architectures are available. Besides, the one for NIPI is specifically designed for inferring RS-HDMR component functions, whereas the one for CABO is only developed for inferring the extrema of the REF. Based on these considerations, a novel contribution here is to present a multi-point selection criterion that can support parallel evaluations of the $\mathcal{G}$-function and also enable us to estimate the REF, its RS-HDMR component functions and bounds at the same time. In this study, the preferred number of CPU cores or workers in a parallel pool is assumed to be an even number, denoted by c.

Stage 1: Global improvement. Supposing that we have only obtained a small set of initial observations, the first stage of our strategy aims to improve the global accuracy of the REF. The key lies in three main aspects: (1) how can we measure the global accuracy of the REF? (2) how to select c points at each iteration that are expected to improve the global accuracy of the REF? (3) when to stop the iteration at this stage?

As the zero-th order RS-HDMR component $\mathcal{M}_0$ is defined as an integral of the REF $\mathcal{M}(\boldsymbol{v})$ with respect to $\boldsymbol{v}$ (called augmented expectation), its accuracy may reflect the global accuracy of the REF to some extent. Therefore, the accuracy of $\hat{\mathcal{M}}_0$ is taken as a global accuracy measure of $\hat{\mathcal{M}}(\boldsymbol{v})$ in this study, which can be quantified by the posterior variance $\sigma^2_{\hat{\mathcal{M}}_0}$. Inspired by [23, 40, 41], a new acquisition function, called posterior variance contribution to the augmented expectation (denoted as $\mathrm{PVC}^{\mathrm{A}}$), is given by:

$$\mathrm{PVC}^{\mathrm{A}}(\boldsymbol{w}) = \Pi_{\boldsymbol{w}'}\left[k_n(\boldsymbol{w},\boldsymbol{w}')\right] \times f_{\boldsymbol{W}}(\boldsymbol{w}) = \left\{\Pi_{\boldsymbol{w}'}\left[k_0(\boldsymbol{w},\boldsymbol{w}')\right] - \boldsymbol{k}_0\left(\boldsymbol{w},\boldsymbol{\mathcal{W}}\right)^{\mathrm{T}} \boldsymbol{K}_0^{-1} \Pi_{\boldsymbol{w}'}\left[\boldsymbol{k}_0\left(\boldsymbol{w}',\boldsymbol{\mathcal{W}}\right)\right]\right\} \times f_{\boldsymbol{W}}(\boldsymbol{w}), \tag{28}$$

where the closed-form expression of $\Pi_{\boldsymbol{w}'}\left[\boldsymbol{k}_0\left(\boldsymbol{w}',\boldsymbol{\mathcal{W}}\right)\right]$ has been given in Eq. (18); Similarly, the term $\Pi_{\boldsymbol{w}'}\left[k_0(\boldsymbol{w},\boldsymbol{w}')\right]$ can be derived as:

$$\begin{aligned}\Pi_{\boldsymbol{w}'}\left[k_0(\boldsymbol{w},\boldsymbol{w}')\right] =& s_0^2 \left|\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1} + \boldsymbol{I}\right|^{-1/2} \exp\left[-\frac{1}{2}\mathrm{diag}\left\{\boldsymbol{u}\left(\boldsymbol{\Sigma}_{\boldsymbol{u}} + \boldsymbol{I}\right)^{-1}\boldsymbol{u}^{\mathrm{T}}\right\}\right] \\ & \cdot \left(\frac{\pi}{2}\right)^{d_2/2} \mathrm{prod}_2\left\{\left[\mathrm{erf}\left((1-\boldsymbol{v})\left(2\boldsymbol{\Sigma}_{\boldsymbol{v}}\right)^{-1/2}\right) - \mathrm{erf}\left(-\boldsymbol{v}\left(2\boldsymbol{\Sigma}_{\boldsymbol{v}}\right)^{-1/2}\right)\right]\boldsymbol{\Sigma}_{\boldsymbol{v}}^{1/2}\right\}.\end{aligned} \tag{29}$$

The acquisition function in Eq. (28) is said to be 'new' because it is essentially not the same as those in the cited references. It should be noted that $\sigma^2_{\hat{\mathcal{M}}_0} = \int_{\mathcal{W}} \mathrm{PVC}^{\mathrm{A}}(\boldsymbol{w})\mathrm{d}\boldsymbol{w}$ holds, which implies that the $\mathrm{PVC}^{\mathrm{A}}$ function can measure the contribution of our epistemic uncertainty at $\boldsymbol{w}$ to $\sigma^2_{\hat{\mathcal{M}}_0}$. For this reason, by selecting $\boldsymbol{w}^{(n+1)} = \arg\max_{\boldsymbol{w}\in\mathcal{W}} \mathrm{PVC}^{\mathrm{A}}(\boldsymbol{w})$ as the best next point to evaluate the $\mathcal{G}$-function, it is expected that the posterior variance of the augmented expectation will decrease the most, and hence the accuracy of

the posterior mean of the augmented expectation will be improved the most. However, adding one single point at a time may waste other useful information and cannot allow to make use of parallelization, and hence it could be inefficient especially when parallel evaluations are possible.

In this study, we propose a novel strategy to parallelize the developed PBQO method by providing c points at each iteration. This strategy is motivated by the fact that the $\mathrm{PVC}^{\mathrm{A}}$ function (defined in Eq. (28)) only explicitly depends on the sampled locations, not on function values at these points. For this reason, we can rewrite the $\mathrm{PVC}^{\mathrm{A}}$ function as $\mathrm{PVC}^{\mathrm{A}}(\boldsymbol{w}, \boldsymbol{\mathcal{W}})$. Therefore, it is possible to select c points ahead of observing their $\mathcal{G}$-function values based on the $\mathrm{PVC}^{\mathrm{A}}(\boldsymbol{w}, \boldsymbol{\mathcal{W}})$ function. Specifically, each point can be selected sequentially, with the $\mathrm{PVC}^{\mathrm{A}}$ function modified by considering the newly selected points at the current iteration. The assumption behind this strategy is that the hyper-parameters will not change, and hence the $\mathrm{PVC}^{\mathrm{A}}$ function remains the same during the process of identifying the next $c-1$ points. In fact, the hyper-parameters do change if we update immediately the GP after each point is chosen and its $\mathcal{G}$-function value is computed, which, however, corresponds to the single-point selection strategy. Our idea is expected to work since the hyper-parameters may not vary too much within the next few steps. The pseudocode of the proposed multi-point selection strategy is given in Algorithm 1. Until c points are obtained, evaluating the $\mathcal{G}$-function at these points can be run in parallel, and the GP model can be updated subsequently. This iteration process is repeated until a stopping criterion is reached, which is defined as the posterior coefficient of variation (COV) of the augmented expectation less than a pre-specified tolerance $\varepsilon^{\mathrm{BPI}}$, i.e., $\frac{\sigma_{\mathcal{M}_0}}{|\mu_{\mathcal{M}_0}|} < \varepsilon^{\mathrm{BPI}}$. To avoid possible premature convergence, the stopping criterion is required to be satisfied several (e.g., two) times in successive iterations. It should be noted that the proposed multi-point selection strategy is computationally inexpensive and can usually produce a batch of c diverse points according to our computational experience, which are thus effective and informative for parallelization.

Algorithm 1 Proposed multi-point selection strategy based on the $\mathrm{PVC}^{\mathrm{A}}(\boldsymbol{w}, \boldsymbol{\mathcal{W}})$ function

1: **Input:** c and $\mathrm{PVC}^{\mathrm{A}}(\boldsymbol{w}, \boldsymbol{\mathcal{W}})$
2: **for** $i = 1 \to c$ **do**
3: $\quad \boldsymbol{w}^{(n+i)} = \arg\max_{\boldsymbol{w} \in \mathcal{W}} \mathrm{PVC}^{\mathrm{A}}(\boldsymbol{w}, \boldsymbol{\mathcal{W}})$
4: $\quad \boldsymbol{\mathcal{W}} = \boldsymbol{\mathcal{W}} \cup \boldsymbol{w}^{(n+i)}$
5: **end for**
6: **Output:** $\boldsymbol{w}^{(n+1)}, \boldsymbol{w}^{(n+2)}, \cdots, \boldsymbol{w}^{(n+c)}$

Stage 2: Local improvement. After stage 1, it is expected that the general trend of the REF has been captured. However, the local features of the REF, e.g., minimum and maximum, may still be inaccurate. In this regard, the second stage of our strategy attempts to further improve the accuracy of the resulting REF from stage 1, with special emphasis on its extrema.

As the posterior distribution of the REF follows a GP, the expected improvement criterion originally introduced in BGO [56] could be adopted for our purposes. Let $\hat{\mathcal{M}}_l(\boldsymbol{v}^-)$ denote the current minimum, and $\boldsymbol{v}^-$ the minimum point, i.e., $\boldsymbol{v}^- = \arg\min_{\boldsymbol{v}\in\mathcal{V}} \mu_{\hat{\mathcal{M}}}(\boldsymbol{v})$. The improvement for the current minimum at the point $\boldsymbol{v}$ can be defined as $I(\boldsymbol{v}) = \max\left(\hat{\mathcal{M}}_l(\boldsymbol{v}^-) - \mu_{\hat{\mathcal{M}}}(\boldsymbol{v}), 0\right)$. The acquisition function, called expected improvement for the minimization (abbreviated as $\mathrm{EI}^{\mathrm{MIN}}$), is to simple take the expected value of $I(\boldsymbol{v})$, i.e., $\mathrm{EI}^{\mathrm{MIN}}(\boldsymbol{v}) = \mathbb{E}\left[I(\boldsymbol{v})\right]$. The closed-form expression of $\mathrm{EI}^{\mathrm{MIN}}$ can be written as [56]:

$$\mathrm{EI}^{\mathrm{MIN}}(\boldsymbol{v}) = \left(\hat{\mathcal{M}}_l(\boldsymbol{v}^-) - \mu_{\hat{\mathcal{M}}}(\boldsymbol{v})\right) \Phi\left(\frac{\hat{\mathcal{M}}_l(\boldsymbol{v}^-) - \mu_{\hat{\mathcal{M}}}(\boldsymbol{v})}{\sigma_{\hat{\mathcal{M}}}(\boldsymbol{v})}\right) + \sigma_{\hat{\mathcal{M}}}(\boldsymbol{v})\varphi\left(\frac{\hat{\mathcal{M}}_l(\boldsymbol{v}^-) - \mu_{\hat{\mathcal{M}}}(\boldsymbol{v})}{\sigma_{\hat{\mathcal{M}}}(\boldsymbol{v})}\right), \tag{30}$$

where $\varphi(\cdot)$ and $\Phi(\cdot)$ are the PDF and and cumulative distribution function (CDF) of the standard normal distribution, respectively. The $\mathrm{EI}^{\mathrm{MIN}}$ function actually measures how much improvement for the minimum is expected to achieve by sampling at $\boldsymbol{v}$. Thus, the next best point for $\boldsymbol{v}$ can be selected by maximizing the $\mathrm{EI}^{\mathrm{MIN}}$ function, i.e., $\underline{\boldsymbol{v}}^\star = \arg\max_{\boldsymbol{v}\in\mathcal{V}} \mathrm{EI}^{\mathrm{MIN}}(\boldsymbol{v})$. The first summation term in Eq. (30) is the exploitation term encouraging to sample where $\mu_{\hat{\mathcal{M}}}(\boldsymbol{v})$ is small, whereas the second summation term is the exploration term encouraging to sample where $\sigma_{\hat{\mathcal{M}}}(\boldsymbol{v})$ is large. At this stage, the associated stopping criterion can be given as [60]:

$$\frac{|\max_{\boldsymbol{v}\in\mathcal{V}} \mathrm{EI}^{\mathrm{MIN}}(\boldsymbol{v})|}{\max \boldsymbol{\mathcal{Z}} - \min \boldsymbol{\mathcal{Z}}} < \varepsilon^{\mathrm{BGO}}, \tag{31}$$

where $\varepsilon^{\mathrm{BGO}}$ is a user-defined tolerance. Similarly, the stopping criterion also needs to be met for two times in succession. Once $\underline{\boldsymbol{v}}^\star$ is identified, the best next point for $\boldsymbol{u}$ can also be specified. In order to improve the accuracy of $\mu_{\hat{\mathcal{M}}}(\underline{\boldsymbol{v}}^\star)$, an acquisition function measuring the posterior variance contribution to $\sigma^2_{\hat{\mathcal{M}}}(\underline{\boldsymbol{v}}^\star)$ (abbreviated as $\mathrm{PVC}^{\mathrm{MIN}}$), can be defined:

$$\begin{aligned}\mathrm{PVC}^{\mathrm{MIN}}(\boldsymbol{u}) &= \Pi_{\boldsymbol{u}'}\left[k_n((\boldsymbol{u}, \underline{\boldsymbol{v}}^\star), (\boldsymbol{u}', \underline{\boldsymbol{v}}^\star))\right] \times f_{\boldsymbol{U}}(\boldsymbol{u}) \\ &= \left\{\Pi_{\boldsymbol{u}'}\left[k_0((\boldsymbol{u}, \underline{\boldsymbol{v}}^\star), (\boldsymbol{u}', \underline{\boldsymbol{v}}^\star))\right] - \boldsymbol{k}_0\left((\boldsymbol{u}, \underline{\boldsymbol{v}}^\star), \boldsymbol{\mathcal{W}}\right)^{\mathrm{T}} \boldsymbol{K}_0^{-1} \Pi_{\boldsymbol{u}'}\left[\boldsymbol{k}_0\left((\boldsymbol{u}', \underline{\boldsymbol{v}}^\star), \boldsymbol{\mathcal{W}}\right)\right]\right\} \times f_{\boldsymbol{U}}(\boldsymbol{u}),\end{aligned} \tag{32}$$

where the term $\Pi_{\boldsymbol{u}'}\left[\boldsymbol{k}_0\left((\boldsymbol{u}', \underline{\boldsymbol{v}}^\star), \boldsymbol{\mathcal{W}}\right)\right]$ can be generated as Eq. (10) by replacing $\boldsymbol{v}$ by $\underline{\boldsymbol{v}}^\star$, and the term $\Pi_{\boldsymbol{u}'}\left[k_0((\boldsymbol{u}, \underline{\boldsymbol{v}}^\star), (\boldsymbol{u}', \underline{\boldsymbol{v}}^\star))\right]$ can be derived as:

$$\Pi_{\boldsymbol{u}'}\left[k_0((\boldsymbol{u}, \underline{\boldsymbol{v}}^\star), (\boldsymbol{u}', \underline{\boldsymbol{v}}^\star))\right] = s_0^2 \left|\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1} + \boldsymbol{I}\right|^{-1/2} \exp\left[-\frac{1}{2}\mathrm{diag}\left\{\boldsymbol{u}\left(\boldsymbol{\Sigma}_{\boldsymbol{u}} + \boldsymbol{I}\right)^{-1} \boldsymbol{u}^{\mathrm{T}}\right\}\right]. \tag{33}$$

In analogy to $\mathrm{PVC}^{\mathrm{A}}$ criterion (see Algorithm 1), $c/2$ points for $\boldsymbol{u}$ can be selected sequentially by maximizing the $\mathrm{PVC}^{\mathrm{MIN}}$ function, denoted as $\underline{\boldsymbol{u}}^{(n+i)}$ $(i = 1, 2, \ldots, c/2)$. The stopping criterion is defined as $\frac{\sigma_{\hat{\mathcal{M}}}(\underline{\boldsymbol{v}}^\star)}{|\mu_{\hat{\mathcal{M}}}(\underline{\boldsymbol{v}}^\star)|} < \varepsilon^{\mathrm{BPI}}$, which should be sissified two times in succession. The identified points for $\boldsymbol{w}$ can be simply formed as: $(\underline{\boldsymbol{u}}^{(n+1)}, \underline{\boldsymbol{v}}^\star)$, $(\underline{\boldsymbol{u}}^{(n+2)}, \underline{\boldsymbol{v}}^\star)$, $\cdots$, $(\underline{\boldsymbol{u}}^{(n+c/2)}, \underline{\boldsymbol{v}}^\star)$.

Similar to Eqs. (30) and (32), the expected improvement and posterior variance contribution for maximization can also be defined, which are denoted as $\mathrm{EI}^{\mathrm{MAX}}$ and $\mathrm{PVC}^{\mathrm{MAX}}$, respectively. To limit the paper length, however, we will not give them in detail. The next point for $\boldsymbol{v}$ can be determined by maximizing

the $\mathrm{EI}^{\mathrm{MAX}}$ function, i.e., $\overline{\boldsymbol{v}}^{\star} = \arg\max_{\boldsymbol{v}\in\mathcal{V}} \mathrm{EI}^{\mathrm{MAX}}(\boldsymbol{v})$. Then, based on the $\mathrm{PVC}^{\mathrm{MAX}}$ function, one can sequentially identify $c/2$ points for $\boldsymbol{u}$, denoted as $\overline{\boldsymbol{u}}^{(n+i)}$ $(i = 1, 2, \ldots, c/2)$. The remaining $c/2$ points for $\boldsymbol{w}$ can be generated as: $(\overline{\boldsymbol{u}}^{(n+1)}, \overline{\boldsymbol{v}}^{\star})$, $(\overline{\boldsymbol{u}}^{(n+2)}, \overline{\boldsymbol{v}}^{\star})$, $\cdots$, $(\overline{\boldsymbol{u}}^{(n+c/2)}, \overline{\boldsymbol{v}}^{\star})$

As a result, a total number of c points for $\boldsymbol{w}$ can be obtained, and the corresponding $\mathcal{G}$-function values can be computed at the same time by running on c cores simultaneously. After that, the GP model can be updated based on the past observations. Once pre-defined stopping criteria are reached, these quantities of interest can be extracted from the finial GP model.

3.5. Numerical implementation of PBQO

For numerical implementation of the proposed PBQO method, the basic procedures are summarized as follows, which are also illustrated by Fig. 1.

Step 1: Get initial observations

The first step consists of generating a small set of n_0 initial samples using Latin hypercube sampling (LHS), denoted as $\boldsymbol{\mathcal{W}} = (\boldsymbol{\mathcal{U}}, \boldsymbol{\mathcal{V}}) = \left\{\boldsymbol{w}^{(j)}\right\}_{i=1}^{n_0}$. The real $\mathcal{G}$-function is then evaluated at these points to obtain corresponding observations, i.e., $\boldsymbol{\mathcal{Z}} = \left\{z^{(i)} = \mathcal{G}\left(\boldsymbol{w}^{(i)}\right)\right\}_{j=1}^{n_0}$, which can be parallelized straightforwardly. The initial training dataset can be constructed: $\boldsymbol{\mathcal{D}} = \{\boldsymbol{\mathcal{W}}, \boldsymbol{\mathcal{Z}}\}$. Let $n = n_0$;

Step 2: Train a GP model

Based on data $\boldsymbol{\mathcal{D}}$, train a new GP model $\mathcal{GP}(\mu_n(\boldsymbol{w}), k_n(\boldsymbol{w}, \boldsymbol{w}'))$ for the $\mathcal{G}$-function. In this study, the *fitrgp* function in Matlab Statistics and Machine Learning Toolbox is used. The prior mean function and covariance function are specified as constant and squared exponential kernel, respectively.

Step 3: Check the stopping criterion

From the trained GP model, one can compute the posterior mean $\mu_{\hat{\mathcal{M}}_0}$ and posterior variance $\sigma^2_{\hat{\mathcal{M}}_0}$ of the augmented expectation by Eqs. (16) and (17), respectively. If the stopping criterion $\frac{\sigma_{\hat{\mathcal{M}}_0}}{|\mu_{\hat{\mathcal{M}}_0}|} < \varepsilon^{\mathrm{BPI}}$ is satisfied two times in succession, go to **Step 5**; else, go to **Step 4**;

Step 4: Identify new observations by the $\mathrm{PVC}^{\mathrm{A}}$ criterion

At this stage, one can identify c points for $\boldsymbol{W}$ by sequentially maximizing the $\mathrm{PVC}^{\mathrm{A}}$ function (Eq. (28)), denoted as $\boldsymbol{\mathcal{W}}^{\star} = \{\boldsymbol{w}^{\star}\}_{j=1}^{c}$. Then, these points are evaluated on the real $\mathcal{G}$-function in parallel to obtain corresponding observations, which are denoted as $\boldsymbol{\mathcal{Z}}^{\star} = \{z^{\star}\}_{j=1}^{c}$. At last, the training dataset $\boldsymbol{\mathcal{D}}$ can be enriched with $\boldsymbol{\mathcal{D}}^{\star} = \{\boldsymbol{\mathcal{W}}^{\star}, \boldsymbol{\mathcal{Z}}^{\star}\}$. Let $n = n + c$ and go to **Step 2**;

Step 5: Select new points by the quadruplet criteria

The next best points $\underline{\boldsymbol{v}}^{\star}$ and $\overline{\boldsymbol{v}}^{\star}$ can be selected by $\underline{\boldsymbol{v}}^{\star} = \arg\max_{\boldsymbol{v}\in\mathcal{V}} \mathrm{EI}^{\mathrm{MIN}}(\boldsymbol{v})$ and $\overline{\boldsymbol{v}}^{\star} = \arg\max_{\boldsymbol{v}\in\mathcal{V}} \mathrm{EI}^{\mathrm{MAX}}(\boldsymbol{v})$, respectively. Then, one can select $c/2$ points $(\underline{\boldsymbol{u}}^{(i)}$ $(i = 1, 2, \ldots, c/2))$ and $(\overline{\boldsymbol{u}}^{(i)}$ $(i = 1, 2, \ldots, c/2))$ by sequentially maximizing the $\mathrm{PVC}^{\mathrm{MIN}}$ function and $\mathrm{PVC}^{\mathrm{MAX}}$ function, respectively. For convenience, we denote the

$c/2$ points for minimization by $\underline{\boldsymbol{\mathcal{W}}} = \left\{\left(\underline{\boldsymbol{v}}^\star, \underline{\boldsymbol{u}}^{(1)}\right), \left(\underline{\boldsymbol{v}}^\star, \underline{\boldsymbol{u}}^{(2)}\right), \cdots, \left(\underline{\boldsymbol{v}}^\star, \underline{\boldsymbol{u}}^{(c/2)}\right)\right\}$, $c/2$ points for maximization by $\overline{\boldsymbol{\mathcal{W}}} = \left\{\left(\overline{\boldsymbol{v}}^\star, \overline{\boldsymbol{u}}^{(1)}\right), \left(\overline{\boldsymbol{v}}^\star, \overline{\boldsymbol{u}}^{(2)}\right), \cdots, \left(\overline{\boldsymbol{v}}^\star, \overline{\boldsymbol{u}}^{(c/2)}\right)\right\}$, and $\boldsymbol{\mathcal{W}}^\star = \{\underline{\boldsymbol{\mathcal{W}}}, \overline{\boldsymbol{\mathcal{W}}}\}$;

Step 6: Judge the stopping criteria

In this step, four stopping criteria should be judged, i.e., $\frac{|\max_{\boldsymbol{v}\in\mathcal{V}} \mathrm{EI}^{\mathrm{MIN}}(\boldsymbol{v})|}{\max \boldsymbol{\mathcal{Z}} - \min \boldsymbol{\mathcal{Z}}} < \varepsilon^{\mathrm{BGO}}$, $\frac{\sigma_{\hat{\mathcal{M}}}(\underline{\boldsymbol{v}}^\star)}{|\mu_{\hat{\mathcal{M}}}(\underline{\boldsymbol{v}}^\star)|} < \varepsilon^{\mathrm{BPI}}$, $\frac{|\max_{\boldsymbol{v}\in\mathcal{V}} \mathrm{EI}^{\mathrm{MAX}}(\boldsymbol{v})|}{\max \boldsymbol{\mathcal{Z}} - \min \boldsymbol{\mathcal{Z}}} < \varepsilon^{\mathrm{BGO}}$ and $\frac{\sigma_{\hat{\mathcal{M}}}(\overline{\boldsymbol{v}}^\star)}{|\mu_{\hat{\mathcal{M}}}(\overline{\boldsymbol{v}}^\star)|} < \varepsilon^{\mathrm{BPI}}$. If all these stopping criteria are met two times in succession, go to **Step 9**; else, go to **Step 7**;

Step 7: Obtain new observations by parallel computing

Evaluation of the real $\mathcal{G}$-function at these c points $\boldsymbol{\mathcal{W}}^\star$ from **Step 5** can be performed in parallel, and c observations are obtained $\boldsymbol{\mathcal{Z}}^\star = \{z^\star\}_{j=1}^{c}$. Finally, the training dataset $\boldsymbol{\mathcal{D}}$ is updated with the new data $\boldsymbol{\mathcal{D}}^\star = \{\boldsymbol{\mathcal{W}}^\star, \boldsymbol{\mathcal{Z}}^\star\}$. Let $n = n + c$;

Step 8: Train a GP model

Train a new GP model $\mathcal{GP}(\mu_n(\boldsymbol{w}), k_n(\boldsymbol{w}, \boldsymbol{w}'))$ for the $\mathcal{G}$-function with data $\boldsymbol{\mathcal{D}}$, and go to **Step 5:**.

Step 9: Return quantities of interest

The posterior means and variances of these quantities of interest, such as REF, its RS-HDMR component functions and bounds, can be extracted form the trained GP model. The posterior means can serve as estimates for these quantities, and the posterior variances measure the epistemic uncertainties (numerical errors) about our estimates.

To initialize the algorithm, there parameters n_0, $\varepsilon^{\mathrm{BPI}}$ and $\varepsilon^{\mathrm{BGO}}$ need to be specified. The initial sample size n_0 should not choose too large as we wish to enlarge the sample size sequentially. For the two thresholds $\varepsilon^{\mathrm{BPI}}$ and $\varepsilon^{\mathrm{BGO}}$, proper values are also important as they influence the accuracy and efficiency of the proposed method. According to our experience, n_0 can take values between 5 and 20 depending on the complexity of the problem at hand, and $\varepsilon^{\mathrm{BPI}}$ and $\varepsilon^{\mathrm{BGO}}$ can be set in the orders of 0.01 and 0.001 respectively. Several optimization problems are involved in the implementation procedures, one can simply use the global optimization algorithms (e.g., genetic algorithm) as the objective functions are all in closed form.

3.6. Relationship to existing NIPI and CABO methods

The proposed PBQO method does share some similarities with the NIPI method and CABO method. For example, they all rely on the use of the GP model in a Bayesian fashion, and can avoid nested loops. However, the differences among the three methods are also significant on several main aspects:

a) The proposed PBQO method transforms the interval variables (including the interval variables in p-boxes) into standard interval ones by a linear transformation. On the contrary, by assuming auxiliary

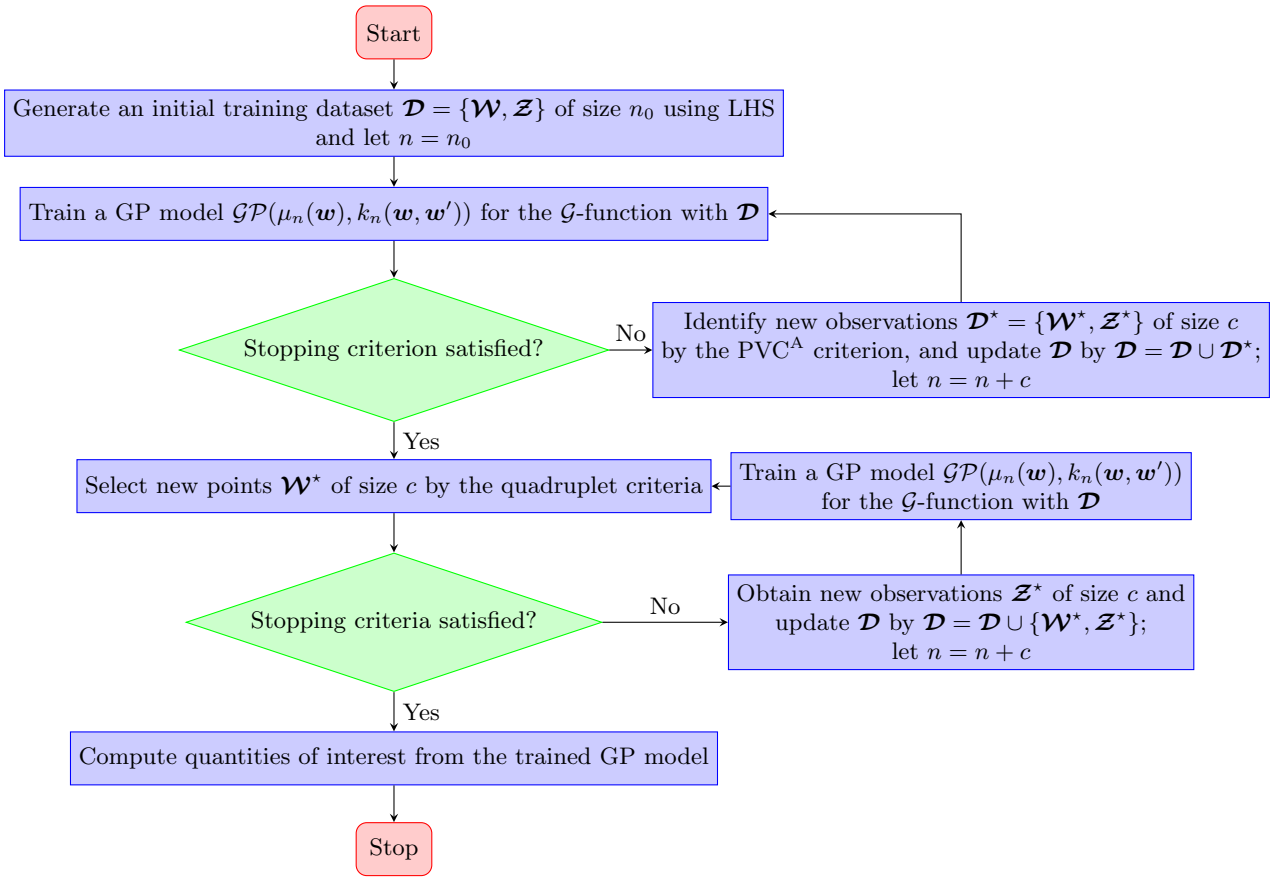


Figure 1: Flowchart of the proposed PBQO method.

uniform distributions for the interval variables, the NIPI and CABO methods convert the interval variables to standard normal ones by a nonlinear transformation. In conjunction with the squared exponential kernel, both of those two strategies can result in analytically tractable results for the REF and its HS-HDMR. However, the NIPI and CABO methods introduce an additional assumption and artificially added nonlinearity. More importantly, the transformation strategy for NIPI and CABO is the cause of poor performance near the bounds of the interval variables. To mitigate this problem, one needs to relax the support of the interval variables when applying NIPI and CABO;

b) Due to the differences in a), the posterior means and variances of the REF and its RS-HDMR component functions are re-derived in the proposed PBQO method, along with some of the acquisition functions;

c) The proposed PBQO method is able to support parallel distributed processing owing to the proposed multi-point selection strategy, while both NIPI and CABO cannot. This advantage is desired when each evaluation of the $\mathcal{G}$-function is costly and parallel computing facilities are available;

d) The proposed PBQO method is capable of yielding the REF, its variable importance and bounds simultaneously with a single run. However, the NIPI method and CABO method are only designed for evaluating the variable importance and bounds, respectively.

4. Extending the proposed method to Case IV

The proposed PBQO method is mainly illustrated in case that hybrid uncertainties present as both random variables and interval variables. When parameterized p-boxes are involved, the proposed method is also applicable, but needs slight adaptations. In this section, we will show how to extend the proposed PBQO method established in Section 3 to **Case IV**.

Let $\boldsymbol{Y} = [Y_1, Y_2, \cdots, Y_{d_3}]$ denote an imprecise random vector containing d_3 variables. These variables are assumed to be characterized by parameterized p-boxes, and their joint PDF is denoted as $f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{\theta})$, which depends on a set of d_4 interval variables $\boldsymbol{\Theta} = [\Theta_1, \Theta_2, \cdots, \Theta_{d_4}]$ with lower and upper bounds $\underline{\boldsymbol{\theta}} = [\underline{\theta}_1, \underline{\theta}_2, \cdots, \underline{\theta}_{d_4}]$ and $\overline{\boldsymbol{\theta}} = [\overline{\theta}_1, \overline{\theta}_2, \cdots, \overline{\theta}_{d_4}]$, respectively. In **Case IV**, the response function is represented by $Z = g(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{A})$. In analogy to **Case III**, an augmented response function $Z = g(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{\Theta})$ needs to be artificially constructed to account for $\boldsymbol{\Theta}$ like $\boldsymbol{A}$. Then, we map the random vector $\{\boldsymbol{X}, \boldsymbol{Y}\}$ to a standard normal one $\boldsymbol{U}$, while the interval vector $\{\boldsymbol{A}, \boldsymbol{\Theta}\}$ to a standard interval one $\boldsymbol{V}$. Accordingly, the response function is changed to be $Z = \mathcal{G}(\boldsymbol{W})$, where $\boldsymbol{W} = \{\boldsymbol{U}, \boldsymbol{V}\}$. See, e.g., [39–41], for the details of how to use an augmented response function when parameterized p-boxes are involved. Note that this does not mean that the original g-function has to be modified, but only for numerical implementation. By doing so, the remaining procedures are similar to those given in Section 3.

5. Numerical examples

In this section, three numerical examples are investigated to demonstrate the proposed method. For comparison purposes, the NIPI and CABO methods are mainly implemented in all examples. These methods are used in a similar way as the proposed PBQO method since they are originally developed for only propagating parameterized p-boxes. Besides, in both methods the support of interval variables has been increased by 10% and the stopping tolerances are specified in accordance with the proposed method.

5.1. Example 1: A test function

Consider a test function of the form:

$$Z = g(X, A_1, A_2) = X^2 + A_1 + A_2^3, \tag{34}$$

where X, A_1 and A_2 are three uncertain input variables, as listed in Tab. 1.

Table 1: Uncertainty characterization of input variables for Example 1.

Notation	Type	Mathematical model
X	Random variable	$\mathcal{N}(0, 1^2)$
A_1	Interval variable	[1 2]
A_2	Interval variable	[1 2]

Note: $\mathcal{N}$ stands for normal distribution.

We first consider the REF $m(\alpha_1, \alpha_2)$, the closed-form expression of which is obtained as $m(\alpha_1, \alpha_2) = 1 + \alpha_1 + \alpha_2^3$. The proposed PBQO method can be implemented to yield a numerical estimate of $m(\alpha_1, \alpha_2)$. In this example, we set $c = 2$, $n_0 = 5$, $\varepsilon^{\text{BPI}} = 0.02$ and $\varepsilon^{\text{BGO}} = 0.002$. Fig. 2(a) depicts the REF estimated by PBQO v.s. its analytical solution, which coincide almost perfectly. Besides, as shown in Fig. 2(b) the coefficient of variation (COV) of the PBQO estimate is quite small, indicting that the estimate is highly reliable. In order to compare with other existing methods, we also employ the NIPI and CABO methods in this example. It can be seen from Figs. 2(c) and 2(e) that both NIPI and CABO methods give poor estimates for the ERF, especially in the boundary area. In addition, Figs. 2(d) and 2(f) show that the results by these two methods also process relatively large variability.

Second, the RS-HDMR component functions of the REF are of concern. For limiting the paper length, we just show the first-order RS-HDMR component functions as an illustration. The analytical expressions of $m_1(\alpha_1)$ and $m_1(\alpha_2)$ can be derived as: $m_1(\alpha_1) = -\frac{3}{2} + \alpha_1$ and $m_2(\alpha_2) = -\frac{15}{4} + \alpha_2^3$. From Fig. 3, one can observe that for both component functions: (1) the proposed PBQO method is able to yield very close estimates to analytical solutions; (2) the 99% confidence intervals (CIs) of PBQO estimates are very narrow; (3) the NIPI and CABO methods are shown to be less accurate than the proposed method; (4) the 99% CIs of both NIPI and CABO estimates are obviously wider than these by the proposed method. These observations demonstrate the accuracy of the proposed method against both NIPI and CABO methods. Besides, through the first-order RS-HDMR component functions it is easy to know that α_2 has significantly larger influence on the REF than α_1. Therefore, if one would like to reduce the epistemic uncertainty in the REF (i.e., narrow the interval), a more rational way is to shrink A_2 by collecting more data of it.

Third, we discuss the results of the response expectation bounds. The analytical lower and upper bounds of the REF are 3 and 11, respectively. Tab. 2 compares the numerical estimates given by the PBQO, NIPI and CABO methods to the analytical solutions. It can be seen that for both lower and upper bounds: (1) PBQO and CABO methods are capable of producing close estimates to the analytical solutions, and restively small posterior COVs; (2) NIPI method gives poor estimates with large posterior COV.

At last, the efficiency and accuracy of these three methods should be emphasized. As listed in Tab. 2, the number of response function evaluations for the PBQO, NIPI and CABO is 13, 8 and 22, respectively.

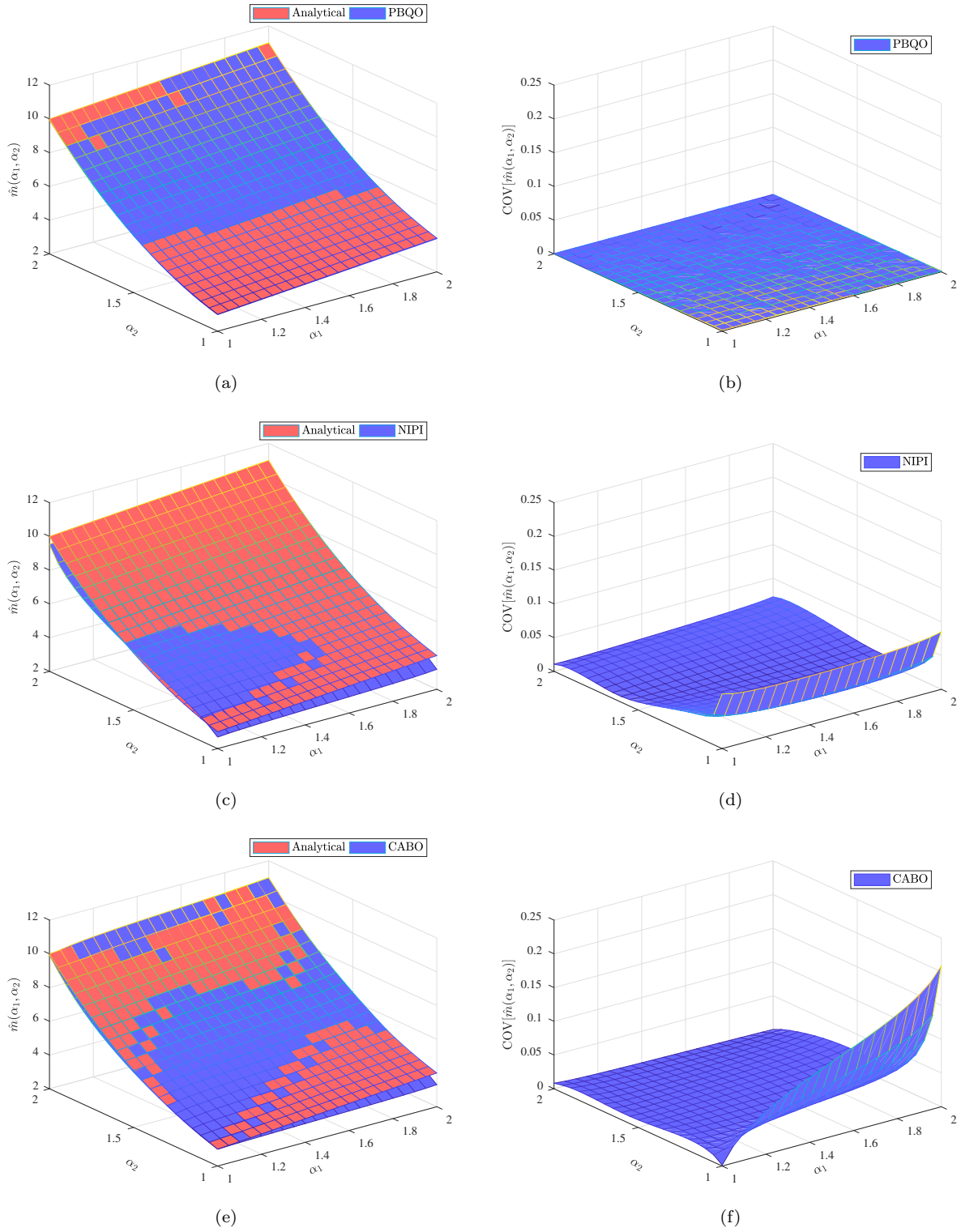
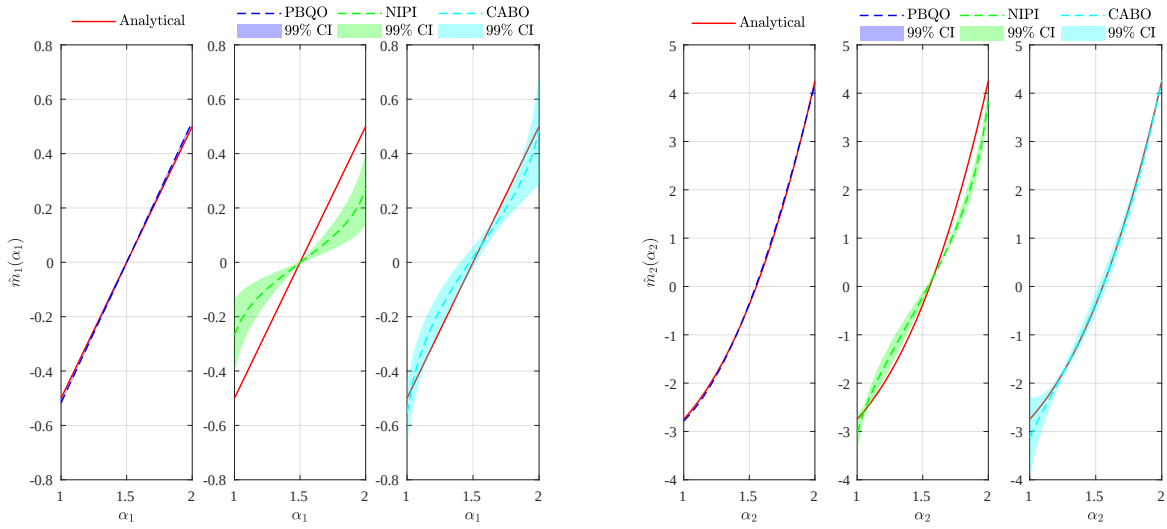


Figure 2: Response expectation function for Example 1 by different methods.



(a) $\hat{m}_1(\alpha_1)$ (b) $\hat{m}_2(\alpha_2)$

Figure 3: First-order RS-HDMR component functions for Example 1 by different methods.

However, the PBQO method can support for parallel computing, and hence its number of calls to the response function for each CPU core is only 6.5 on average. To this end, the number of effective response function evaluations required by the proposed PBQO method is close to that of the NIPI method, but less than the CABO method. Besides, the proposed PBQO method is able to produce the REF, its RS-HDMR component functions and bound simultaneously with reasonable accuracy, while the NIPI method may perform worse in all these three aspects and the CABO method could be reliable only in capturing the REF bounds.

Table 2: Response expectation bound for Example 1.

Method	$\hat{m}_l$	COV$[\hat{m}_l]/\%$	$\hat{m}_u$	COV$[\hat{m}_u]$ /%	N	$\frac{N}{c}$
Analytical	3	-	11	-	-	-
PBQO ($c=2$)	2.9820	0.22	11.0027	0.00	5+8=13	6.5
NIPI ($c=1$)	2.6795	8.11	10.0148	2.26	5+3=8	8
CABO ($c=1$)	3.0033	0.08	10.9966	0.00	5+17=22	22

Note: N is the total number of response function evaluations; c is the number of points selected at each iteration, and hence the number of CPU cores used in parallel; and N/c is referred to as the number of effective respone function evaluations.

5.2. Example 2: A non-linear oscillator

The second example considers a nonlinear undamped single degree-of-freedom (SDOF) oscillator subjected to a rectangular pulse load (as shown in Fig. 4), which was extensively studied in context of reliability analysis (see, e.g., [39, 61, 62]). The response function is defined as the maximum displacement of the oscillator:

$$Z = g(c_1, c_2, m, F_1, t_1) = \left| \frac{2F_1}{c_1 + c_2} \sin\left(\frac{t_1}{2} \sqrt{\frac{c_1 + c_2}{m}} \right) \right|, \tag{35}$$

where c_1, c_2, m, F_1, t_1 are five uncertain input variables, detailed description of which can be found in Tab. 3. For notational clarity, we denote the three intervals as $A_1 = [1\ 2]$, $A_2 = [0.1\ 0.3]$ and $A_3 = [0.5\ 1.5]$ in what follows.

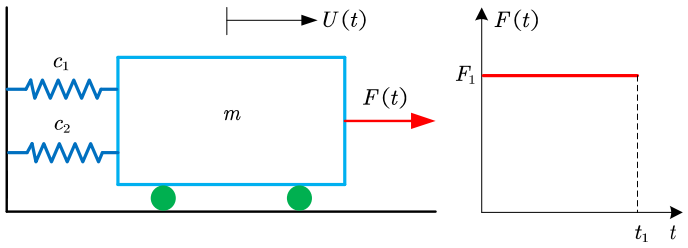


Figure 4: A nonlinear SDOF oscillator subjected to a rectangular pulse load.

Table 3: Uncertainty characterization of input variables for Example 2.

Notation	Type	Mathematical model
c_1	Random variable	$\mathcal{N}(1, 0.1^2)$
c_2	Random variable	$\mathcal{N}(0.1, 0.01^2)$
m	Random variable	$\mathcal{N}(1, 0.1^2)$
F_1	P-box variable	$\mathcal{LN}([1\ 2], [0.1\ 0.3]^2)$
t_1	Interval variable	$[0.5\ 1.5]$

Note: $\mathcal{LN}$ stands for Lognormal distribution.

In this example, the REF, its RS-HDMR component functions and bounds are also of our interest. Due to the complexity of the response function, the corresponding analytical solutions are not available, and thus we use Monte Carlo simulation (MCS) or double-loop MCS (DL-MCS) [63] to provide reference results. The initial parameters of the proposed PBQO method are specified as: $c = 4$, $n_0 = 15$, $\varepsilon^{\mathrm{BPI}} = 0.01$ and $\varepsilon^{\mathrm{BGO}} = 0.001$. It should be noted that the REF is three-dimensional, and hence we simply set $\alpha_3 = 1$ in order to visualize the results. As can be seen from Fig. 5(b), the COV of the MCS estimate is extremely small, indicting that we can take the MCS estimate as a reference result. From Figs. 5(a), 5(c) and 5(e), it is obvious that the proposed PBQO method can produce a much better REF estimate than the NIPI

and CABO methods. Besides, the posterior COV of the PBQO estimate is also much smaller than those by NIPI and CABO methods, as shown in Figs. 5(b), 5(d) and 5(f). As for the RS-HDMR component functions of the REF, we only give three first-order RS-HDMR component functions $\hat{m}_1(\alpha_1)$, $\hat{m}_2(\alpha_2)$ and $\hat{m}_3(\alpha_3)$ as an illustration. It can be seen from Fig. 6 that for all the three component functions the proposed PBQO method can produce fairly good results, in comparison to these given by MCS. However, the NIPI and CABO methods perform much worse than PBQO, especially for $\hat{m}_2(\alpha_2)$. Tab. 4 compares the lower and upper bounds of the REF by different methods. As can be seen, the PBQO and CABO methods are able to yield desirable estimates with relatively small posterior COVs, while the NIPI method does not work well. It should be noted that the proposed method only requires 7.75 effective response function evaluations to produce the above results, which are less than those by NIPI and CABO.

Table 4: Response expectation bound for Example 2.

Method	$\hat{m}_l$	COV$[\hat{m}_l]$/%	$\hat{m}_u$	COV$[\hat{m}_u]$ /%	N	$\frac{N}{c}$
DL-MCS	0.4953	0.87	2.5766	0.37	10^6	-
PBQO ($c=4$)	0.4583	0.35	2.5935	0.07	15+16=31	7.75
NIPI ($c=1$)	0.4160	15.94	2.6343	3.04	15+3=18	18
CABO ($c=1$)	0.4721	0.25	2.5866	0.08	15+24=39	38

5.3. Example 3: A 56-bar spatial truss structure

The third example consists of a 56-bar spatial truss structure, as shown in Fig. 7. Nine vertical loads are applied to the structure at joints 1-9, which are denoted as $P_1 \sim P_9$. The external loads $P_1 - P_9$ are assumed to be uncertain, together with the elastic modulus E and cross-sectional area A. These uncertainties are characterized by three kinds of models, which are summarized in Tab. 5. It can be seen that four intervals are involved and we denote them as $A_1 = [20\ 30]$ kN, $A_2 = [30\ 40]$ kN, $A_3 = [200\ 220]$ Gpa and $A_4 = [150\ 250]$ mm^2. The response of concern is selected as the vertical displacement of joint 1, which can implicitly expressed as a function of $P_1 \sim P_9$, E and A, i.e., $Z = g(P_1 \sim P_9, E, A)$.

Table 5: Uncertainty characterization of input variables for Example 3.

Notation	Type	Unit	Mathematical model
$P_2 \sim P_9$	Random variable	kN	$\mathcal{LN}(20, 4^2)$
P_1	P-box variable	kN	$\mathcal{LN}([20\ 30], [30\ 40]^2)$
E	Interval variable	GPa	$[200\ 220]$
A	Interval variable	mm^2	$[150\ 250]$

The proposed PBQO method is initialized with $c = 4$, $n_0 = 20$, $\varepsilon^{\mathrm{BPI}} = 0.02$ and $\varepsilon^{\mathrm{BGO}} = 0.002$. Fig. 8 depicts the REF estimates by three methods and their corresponding posterior COVs, where we fix α_3 and

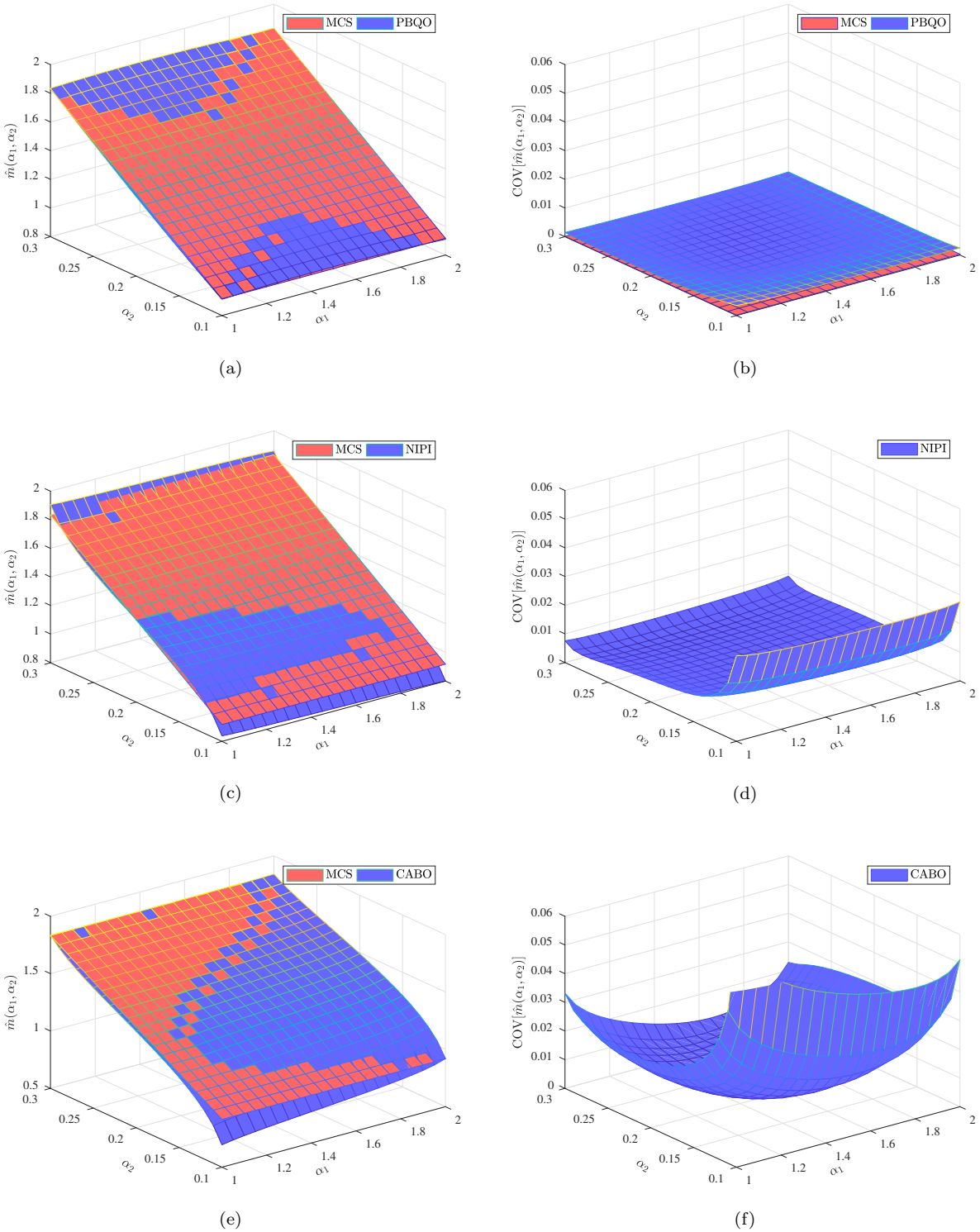
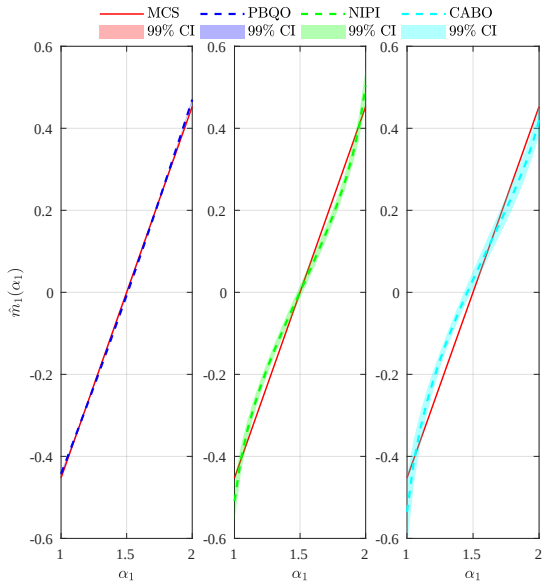
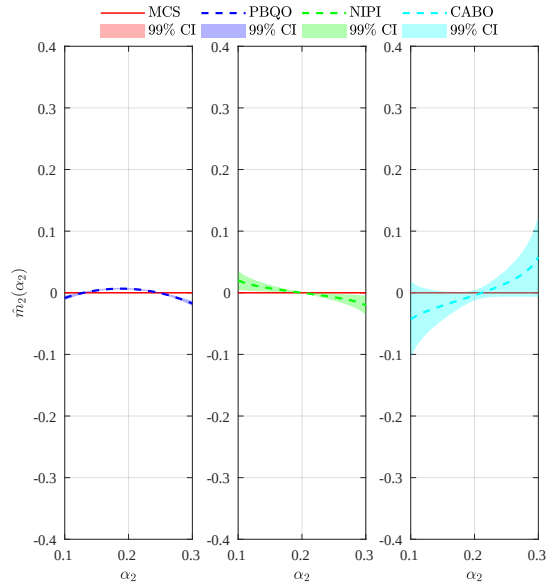


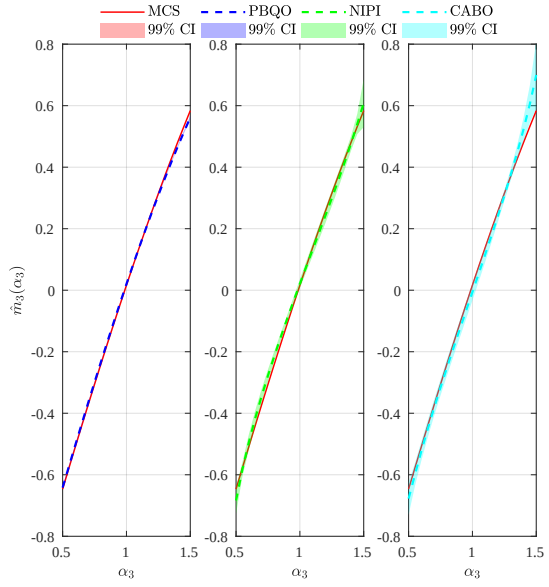
Figure 5: Response expectation function for Example 2 by different methods ($\alpha_3 = 1$).



(a) $\hat{m}_1(\alpha_1)$



(b) $\hat{m}_2(\alpha_2)$



(c) $\hat{m}_3(\alpha_3)$

Figure 6: First-order RS-HDMR component functions for Example 2 by different methods.

α_4 at their midpoints, i.e., $\alpha_3 = 210$ Gpa and $\alpha_4 = 200$ mm^2. It is shown that the posterior COV of the PBQO estimate is much smaller that those by both NIPI and CABO methods, indicating that the proposed PBQO method is more reliable for capturing the REF. The results of four first-order HDMR component functions in Fig. 9 also imply that the proposed method has better accuracy than the NIPI and CABO methods. Besides, it is easy to know from Fig. 9 that the four intervals can be ranked as $A_4 > A_1 > A_3 > A_2$

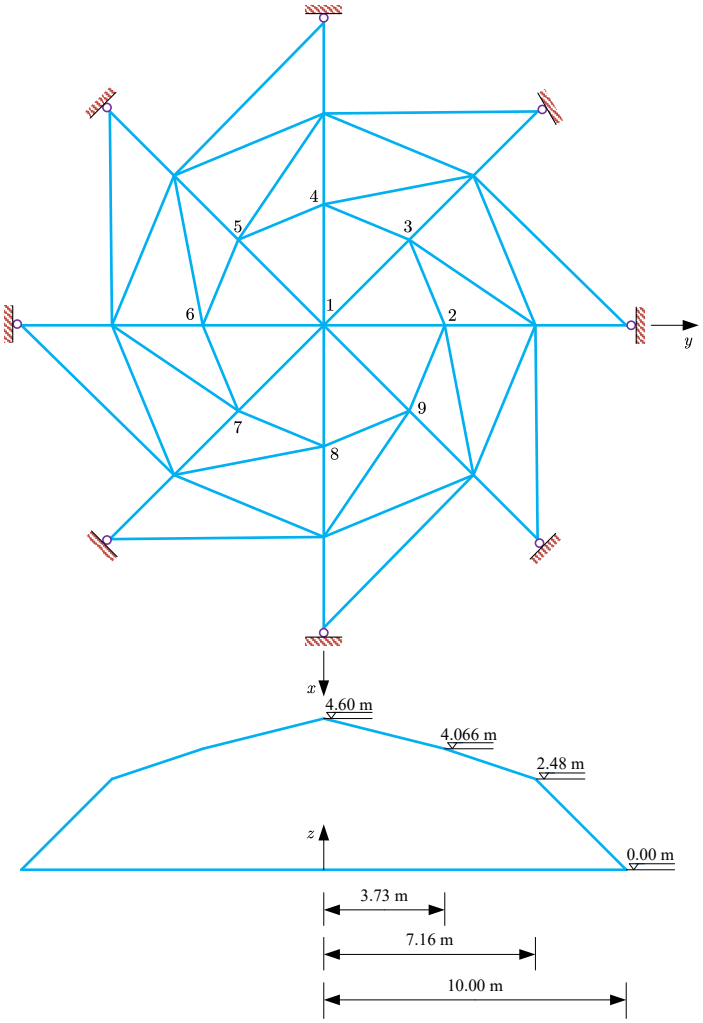


Figure 7: A 56-bar spatial truss structure.

in terms of their first-order importance to the REF. Through Tab. 6, one can find that for both lower and upper bounds of the REF the PBQO and CABO can yield better estimates than the NIPI, indicating by their posterior COVs. It should be emphasized that by taking advantage of parallel computing the effective response function calls required by the proposed PBQO method are much less than that of CABO.

Table 6: Response expectation bound for Example 3.

Method	$\hat{m}_l$	COV$[\hat{m}_l]/\%$	$\hat{m}_u$	COV$[\hat{m}_u]$ /%	N	$\frac{N}{c}$
DL-MCS	11.1793	2.11	35.2535	1.64	10^4	-
PBQO ($c=4$)	11.8785	0.23	35.4109	0.08	20+12=32	8
NIPI ($c=1$)	12.5791	4.74	34.4007	2.36	20+2=22	22
CABO ($c=1$)	11.5818	0.12	35.3302	0.07	20+23=43	43

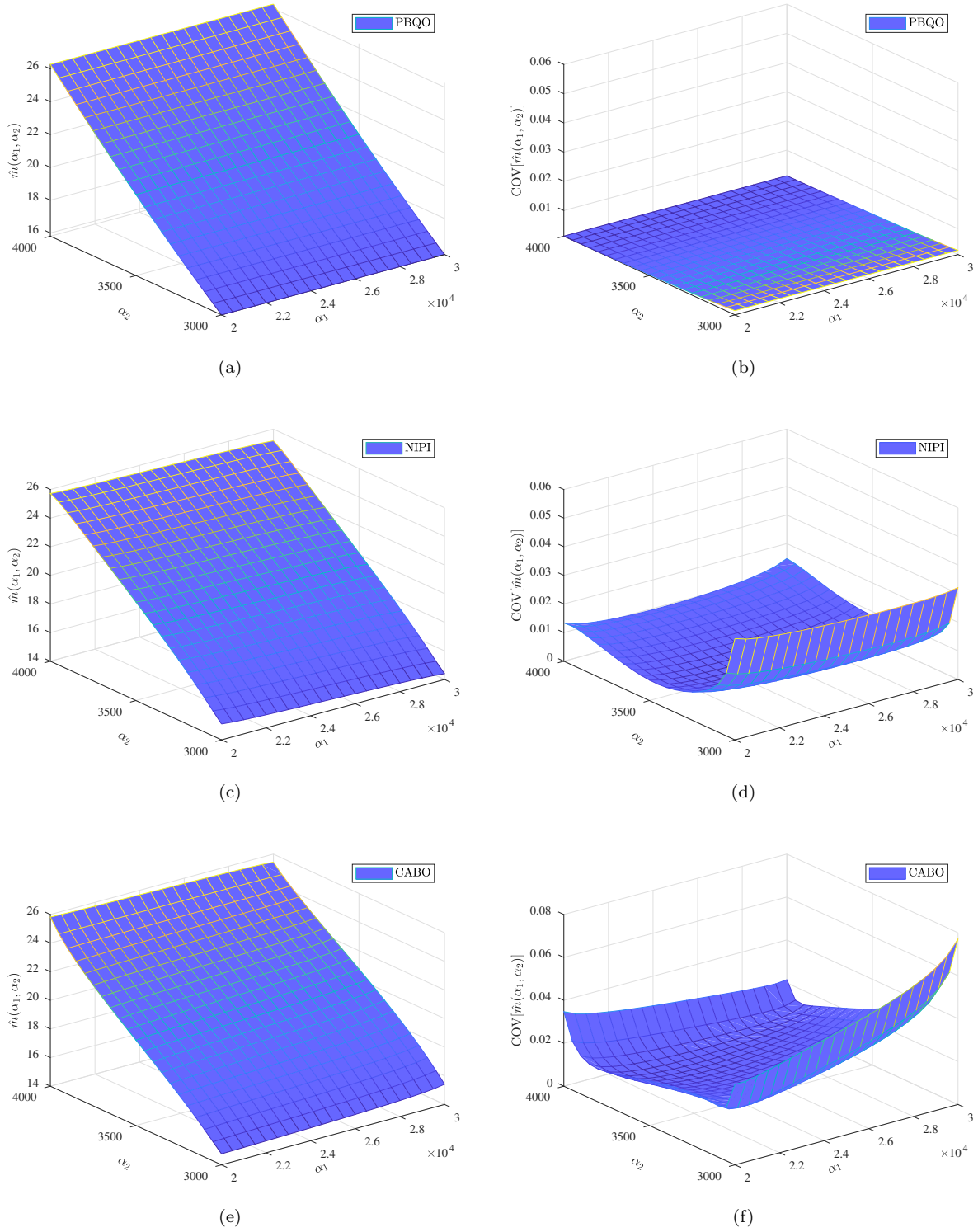
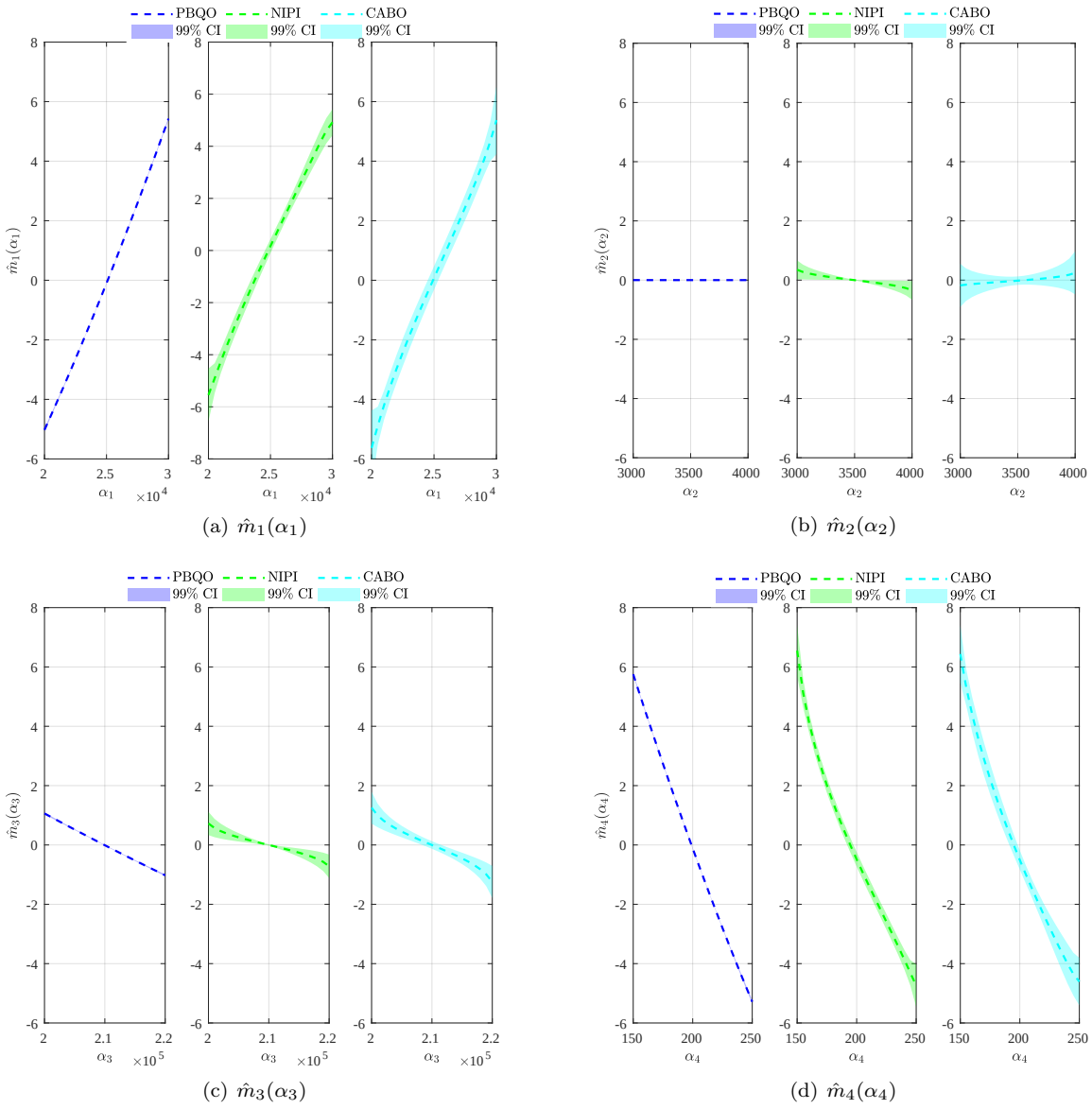


Figure 8: Response expectation function for Example 3 by different methods ($\alpha_3 = 210$ Gpa and $\alpha_4 = 200$ mm^2).



(a) $\hat{m}_1(\alpha_1)$ (b) $\hat{m}_2(\alpha_2)$

(c) $\hat{m}_3(\alpha_3)$ (d) $\hat{m}_4(\alpha_4)$

Figure 9: First-order RS-HDMR component functions for Example 3 by different methods.

6. Conclusions and perspectives

In this work, propagation of hybrid uncertainties in the form of precise random variables, parameterized p-boxes and interval variables is studied via Bayesian numerical analysis. The main contribution lies in the development of a novel method, termed 'Parallel Bayesian Quadrature Optimization', for estimation of response expectation function, its RS-HDMR component functions and bounds simultaneously. Compared to the state-of-the-art methods for propagating hybrid uncertainties, the proposed method has several significant advantages. First, the proposed method breaks the double-loop paradigm that typically propagates

aleatory and epistemic uncertainty separately in a nested way. That is, it can propagate both types of uncertainties simultaneously, and is a fully-decoupled procedure in nature, yielding a major improvement in computational efficiency. Second, the proposed method is able to exploit prior knowledge thanks to its Bayesian nature, and it also supports parallel computing, further leading to much higher computational efficiency. Third, the estimators (i.e., posterior means) of the response moment function and its RS-HDMR component functions are analytically derived, together with their posterior variances for indicating numerical errors.

While these advantages are encouraging, there are still some issues that need further study. For example, one should note that the analytical tractability of the proposed method is based on using the squared exponential kernel that is appropriate for modelling smooth and moderately nonlinear functions. This, however, is not always justified for a general practical problem. Besides, the proposed method relies on a total number of five acquisition functions, which could be reduced by developing more efficient Bayesian experimental design strategies. The proposed method could be extended to evaluate the second-order raw moment function, while more research efforts may still be required.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Chao Dang is mainly supported by China Scholarship Council (CSC). Pengfei Wei is grateful to the support from the National Natural Science Foundation of China (grant no. 51905430 and and 72171194). Chao Dang, Pengfei Wei and Michael Beer also would like to appreciate the support of Sino-German Mobility Program under grant number M-0175.

Appendix A. Derivation of the posterior variance for the first-order RS-HDMR component function

The posterior variance function $\sigma^2_{\hat{\mathcal{M}}_i}(v_i)$ for the first-order RS-HDMR function $\hat{\mathcal{M}}_i(v_i)$ can be given by:

$$\sigma^2_{\hat{\mathcal{M}}_i}(v_i) = \mathbb{V}_{\boldsymbol{\mathcal{D}}}\left[\hat{\mathcal{M}}_i(v_i)\right] = \mathbb{V}_{\boldsymbol{\mathcal{D}}}\left[\Pi_{-v_i}\left[\mathcal{G}(\boldsymbol{w})\right]\right] + \sigma^2_{\hat{\mathcal{M}}_0} - 2\mathbb{COV}_{\boldsymbol{\mathcal{D}}}\left[\Pi_{-v_i}\left[\hat{\mathcal{G}}(\boldsymbol{w})\right], \hat{\mathcal{M}}_0\right], \tag{A.1}$$

where $\mathbb{COV}_{\boldsymbol{\mathcal{D}}}[\cdot,\cdot]$ refers to the covariance taken with respect to the posterior distributions of its arguments given data $\boldsymbol{\mathcal{D}}$; the term $\sigma^2_{\hat{\mathcal{M}}_0}$ has been given in Eq. (17).

The term $\mathbb{V}_{\boldsymbol{\mathcal{D}}}\left[\Pi_{-v_i}\left[\mathcal{G}(\boldsymbol{w})\right]\right]$ in Eq. (A.1) can be further deduced by applying Fubini's theorem such that:

$$\mathbb{V}_{\boldsymbol{\mathcal{D}}}\left[\Pi_{-v_i}\left[\mathcal{G}(\boldsymbol{w})\right]\right] = \Pi_{-v_i}\Pi_{-v_i}\left[k_0\left(\boldsymbol{w},\left(\boldsymbol{w}'_{-v'_i},v_i\right)\right)\right] - \Pi_{-v_i}\left[\boldsymbol{k}_0\left(\boldsymbol{w},\boldsymbol{\mathcal{W}}\right)^{\mathrm{T}}\right]\boldsymbol{K}_0^{-1}\Pi_{-v_i}\left[\boldsymbol{k}_0\left(\boldsymbol{w},\boldsymbol{\mathcal{W}}\right)\right], \tag{A.2}$$

where the term $\Pi_{-v_i}\left[\boldsymbol{k}_0\left(\boldsymbol{w},\boldsymbol{\mathcal{W}}\right)\right]$ has been given in Eq. (21); the term $\Pi_{-v_i}\Pi_{-v_i}\left[k_0\left(\boldsymbol{w},\left(\boldsymbol{w}'_{-(d_1+i)},v_i\right)\right)\right]$ can be derived as:

$$\begin{aligned}
&\Pi_{-v_i}\Pi_{-v_i}\left[k_0\left(\boldsymbol{w},\left(\boldsymbol{w}'_{-(d_1+i)},v_i\right)\right)\right]\\
=&s_0^2\left|2\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}+\boldsymbol{I}\right|^{-1/2}\\
&\cdot 2^{(d_2-1)}\mathrm{prod}_1\left\{\mathrm{diag}\left\{\boldsymbol{\Sigma}_{\boldsymbol{v}_{-i}}\left[-1+\exp\left[-(2\boldsymbol{\Sigma}_{\boldsymbol{v}_{-i}})^{-1}\right]+(2\pi^{-1}\boldsymbol{\Sigma}_{\boldsymbol{v}_{-i}})^{-1/2}\mathrm{erf}\left(\left(2\boldsymbol{\Sigma}_{\boldsymbol{v}_{-i}}\right)^{-1/2}\right)\right]\right\}\right\}.
\end{aligned} \tag{A.3}$$

Likewise, the term $\mathbb{COV}_{\boldsymbol{\mathcal{D}}}\left[\Pi_{-v_i}\left[\hat{\mathcal{G}}(\boldsymbol{w})\right],\hat{\mathcal{M}}_0\right]$ in Eq. (A.1) can be formulated as:

$$\mathbb{COV}_{\boldsymbol{\mathcal{D}}}\left[\Pi_{-v_i}\left[\hat{\mathcal{G}}(\boldsymbol{w})\right],\hat{\mathcal{M}}_0\right] = \Pi_{-v_i}\Pi\left[k_0\left(\boldsymbol{w},\boldsymbol{w}'\right)\right] - \Pi_{-v_i}\left[\boldsymbol{k}_0\left(\boldsymbol{w},\boldsymbol{\mathcal{W}}\right)^{\mathrm{T}}\right]\boldsymbol{K}_0^{-1}\Pi\left[\boldsymbol{k}_0\left(\boldsymbol{w},\boldsymbol{\mathcal{W}}\right)\right], \tag{A.4}$$

where

$$\begin{aligned}
&\Pi_{-v_i}\Pi\left[k_0\left(\boldsymbol{w},\boldsymbol{w}'\right)\right]\\
=&s_0^2\left|2\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}+\boldsymbol{I}\right|^{-1/2}\\
&\cdot 2^{(d_2-1)}\mathrm{prod}_1\left\{\mathrm{diag}\left\{\boldsymbol{\Sigma}_{\boldsymbol{v}_{-i}}\left[-1+\exp\left[-(2\boldsymbol{\Sigma}_{\boldsymbol{v}_{-i}})^{-1}\right]+(2\pi^{-1}\boldsymbol{\Sigma}_{\boldsymbol{v}_{-i}})^{-1/2}\mathrm{erf}\left(\left(2\boldsymbol{\Sigma}_{\boldsymbol{v}_{-i}}\right)^{-1/2}\right)\right]\right\}\right\}\\
&\cdot\left(\frac{\pi}{2}\right)^{1/2}\mathrm{prod}_2\left\{\left[\mathrm{erf}\left((1-v_i)\left(2\boldsymbol{\Sigma}_{\boldsymbol{v}_i}\right)^{-1/2}\right)-\mathrm{erf}\left(-v_i\left(2\boldsymbol{\Sigma}_{\boldsymbol{v}_i}\right)^{-1/2}\right)\right]\boldsymbol{\Sigma}_{\boldsymbol{v}_i}^{1/2}\right\}.
\end{aligned} \tag{A.5}$$

Appendix B. Derivation of the posterior variance for the second-order RS-HDMR component function

The posterior variance function $\sigma^2_{\hat{\mathcal{M}}_{ij}}(v_i,v_j)$ for the second-order RS-HDMR component function $\hat{\mathcal{M}}_{ij}(v_i,v_j)$ can be formulated as:

$$\begin{aligned}
\sigma^2_{\hat{\mathcal{M}}_{ij}}(v_i,v_j) =&\mathbb{V}_{\boldsymbol{\mathcal{D}}}\left[\hat{\mathcal{M}}_{ij}(v_i,v_j)\right]\\
=&\mathbb{V}_{\boldsymbol{\mathcal{D}}}\left[\Pi_{-\boldsymbol{v}_{ij}}\left[\mathcal{G}(\boldsymbol{w})\right]\right]+\sigma^2_{\hat{\mathcal{M}}_i}(v_i)+\sigma^2_{\hat{\mathcal{M}}_j}(v_j)+\sigma^2_{\hat{\mathcal{M}}_0}\\
&-2\mathbb{COV}_{\boldsymbol{\mathcal{D}}}\left[\Pi_{-\boldsymbol{v}_{ij}}\left[\mathcal{G}(\boldsymbol{w})\right],\Pi_{-v_i}\left[\mathcal{G}(\boldsymbol{w})\right]\right]-2\mathbb{COV}_{\boldsymbol{\mathcal{D}}}\left[\Pi_{-\boldsymbol{v}_{ij}}\left[\mathcal{G}(\boldsymbol{w})\right],\Pi_{-v_j}\left[\mathcal{G}(\boldsymbol{w})\right]\right]\\
&+2\mathbb{COV}_{\boldsymbol{\mathcal{D}}}\left[\Pi_{-\boldsymbol{v}_{ij}}\left[\mathcal{G}(\boldsymbol{w})\right],\Pi\left[\mathcal{G}(\boldsymbol{w})\right]\right]+2\mathbb{COV}_{\boldsymbol{\mathcal{D}}}\left[\Pi_{-v_i}\left[\mathcal{G}(\boldsymbol{w})\right],\Pi_{-v_j}\left[\mathcal{G}(\boldsymbol{w})\right]\right]\\
&-2\mathbb{COV}_{\boldsymbol{\mathcal{D}}}\left[\Pi_{-v_i}\left[\mathcal{G}(\boldsymbol{w})\right],\Pi\left[\mathcal{G}(\boldsymbol{w})\right]\right]-2\mathbb{COV}_{\boldsymbol{\mathcal{D}}}\left[\Pi_{-v_j}\left[\mathcal{G}(\boldsymbol{w})\right],\Pi\left[\mathcal{G}(\boldsymbol{w})\right]\right],
\end{aligned} \tag{B.1}$$

where the terms $\sigma^2_{\hat{\mathcal{M}}_i}(v_i)$ and $\sigma^2_{\hat{\mathcal{M}}_j}(v_j)$ can refer to Eq. (A.1); the term $\sigma^2_{\hat{\mathcal{M}}_0}$ has been derived in Eq. (17); the last two covariance terms $\mathbb{COV}_{\boldsymbol{\mathcal{D}}}\left[\Pi_{-v_i}\left[\mathcal{G}(\boldsymbol{w})\right],\Pi\left[\mathcal{G}(\boldsymbol{w})\right]\right]$ and $\mathbb{COV}_{\boldsymbol{\mathcal{D}}}\left[\Pi_{-v_j}\left[\mathcal{G}(\boldsymbol{w})\right],\Pi\left[\mathcal{G}(\boldsymbol{w})\right]\right]$ has been given in Eq. (A.4).

The term $\mathbb{V}_{\boldsymbol{\mathcal{D}}}\left[\Pi_{-\boldsymbol{v}_{ij}}\left[\mathcal{G}(\boldsymbol{w})\right]\right]$ in Eq. (B.1) can be derived as:

$$\mathbb{V}_{\boldsymbol{\mathcal{D}}}\left[\Pi_{-\boldsymbol{v}_{ij}}\left[\mathcal{G}(\boldsymbol{w})\right]\right] = \Pi_{-\boldsymbol{v}_{ij}}\Pi_{-\boldsymbol{v}_{ij}}\left[k_0\left(\boldsymbol{w},\left(\boldsymbol{w}'_{-\boldsymbol{v}'_{ij}},\boldsymbol{v}_{ij}\right)\right)\right] - \Pi_{-\boldsymbol{v}_{ij}}\left[\boldsymbol{k}_0\left(\boldsymbol{w},\boldsymbol{\mathcal{W}}\right)^{\mathrm{T}}\right]\boldsymbol{K}_0^{-1}\Pi_{-\boldsymbol{v}_{ij}}\left[\boldsymbol{k}_0\left(\boldsymbol{w},\boldsymbol{\mathcal{W}}\right)\right], \tag{B.2}$$

where the term $\Pi_{-\boldsymbol{v}_{ij}}\left[\boldsymbol{k}_0\left(\boldsymbol{w},\boldsymbol{\mathcal{W}}\right)\right]$ has been given in Eq. (23); the term $\Pi_{-\boldsymbol{v}_{ij}}\Pi_{-\boldsymbol{v}_{ij}}\left[k_0\left(\boldsymbol{w},\left(\boldsymbol{w}'_{-\boldsymbol{v}'_{ij}},\boldsymbol{v}_{ij}\right)\right)\right]$ can be derived as:

$$\begin{aligned}
&\Pi_{-\boldsymbol{v}_{ij}}\Pi_{-\boldsymbol{v}_{ij}}\left[k_0\left(\boldsymbol{w},\left(\boldsymbol{w}'_{-\boldsymbol{v}'_{ij}},\boldsymbol{v}_{ij}\right)\right)\right]\\
=&s_0^2\left|2\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}+\boldsymbol{I}\right|^{-1/2}\\
&\cdot 2^{(d_2-2)}\mathrm{prod}_1\left\{\mathrm{diag}\left\{\boldsymbol{\Sigma}_{\boldsymbol{v}_{-ij}}\left[-1+\exp\left[-(2\boldsymbol{\Sigma}_{\boldsymbol{v}_{-ij}})^{-1}\right]+(2\pi^{-1}\boldsymbol{\Sigma}_{\boldsymbol{v}_{-ij}})^{-1/2}\mathrm{erf}\left(\left(2\boldsymbol{\Sigma}_{\boldsymbol{v}_{-ij}}\right)^{-1/2}\right)\right]\right\}\right\}.
\end{aligned}\tag{B.3}$$

The term $\mathbb{COV}_{\boldsymbol{\mathcal{D}}}\left[\Pi_{-\boldsymbol{v}_{ij}}\left[\mathcal{G}(\boldsymbol{w})\right],\Pi_{-v_i}\left[\mathcal{G}(\boldsymbol{w})\right]\right]$ in Eq. (B.1) is formulated as:

$$\mathbb{COV}_{\boldsymbol{\mathcal{D}}}\left[\Pi_{-\boldsymbol{v}_{ij}}\left[\mathcal{G}(\boldsymbol{w})\right],\Pi_{-v_i}\left[\mathcal{G}(\boldsymbol{w})\right]\right]=\Pi_{-\boldsymbol{v}_{ij}}\Pi_{-v_i}\left[k_0\left(\boldsymbol{w},\left(\boldsymbol{w}'_{-v'_i},v_i\right)\right)\right]-\Pi_{-\boldsymbol{v}_{ij}}\left[\boldsymbol{k}_0\left(\boldsymbol{w},\boldsymbol{\mathcal{W}}\right)^{\mathrm{T}}\right]\boldsymbol{K}_0^{-1}\Pi_{-v_i}\left[\boldsymbol{k}_0\left(\boldsymbol{w},\boldsymbol{\mathcal{W}}\right)\right],\tag{B.4}$$

where the terms $\Pi_{-\boldsymbol{v}_{ij}}\left[\boldsymbol{k}_0\left(\boldsymbol{w},\boldsymbol{\mathcal{W}}\right)\right]$ and $\Pi_{-v_i}\left[\boldsymbol{k}_0\left(\boldsymbol{w},\boldsymbol{\mathcal{W}}\right)\right]$ have been given in Eq. (23) and Eq. (21) respectively; the term $\Pi_{-\boldsymbol{v}_{ij}}\Pi_{-v_i}\left[k_0\left(\boldsymbol{w},\left(\boldsymbol{w}'_{-v'_i},v_i\right)\right)\right]$ can be derived as:

$$\begin{aligned}
&\Pi_{-\boldsymbol{v}_{ij}}\Pi_{-v_i}\left[k_0\left(\boldsymbol{w},\left(\boldsymbol{w}'_{-v'_i},v_i\right)\right)\right]\\
=&s_0^2\left|2\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}+\boldsymbol{I}\right|^{-1/2}\\
&\cdot 2^{(d_2-2)}\mathrm{prod}_1\left\{\mathrm{diag}\left\{\boldsymbol{\Sigma}_{\boldsymbol{v}_{-ij}}\left[-1+\exp\left[-(2\boldsymbol{\Sigma}_{\boldsymbol{v}_{-ij}})^{-1}\right]+(2\pi^{-1}\boldsymbol{\Sigma}_{\boldsymbol{v}_{-ij}})^{-1/2}\mathrm{erf}\left(\left(2\boldsymbol{\Sigma}_{\boldsymbol{v}_{-ij}}\right)^{-1/2}\right)\right]\right\}\right\}\\
&\cdot\left(\frac{\pi}{2}\right)^{1/2}\mathrm{prod}_2\left\{\left[\mathrm{erf}\left((1-v_j)\left(2\boldsymbol{\Sigma}_{\boldsymbol{v}_j}\right)^{-1/2}\right)-\mathrm{erf}\left(-v_j\left(2\boldsymbol{\Sigma}_{\boldsymbol{v}_j}\right)^{-1/2}\right)\right]\boldsymbol{\Sigma}_{\boldsymbol{v}_j}^{1/2}\right\}.
\end{aligned}\tag{B.5}$$

Note that the term $\mathbb{COV}_{\boldsymbol{\mathcal{D}}}\left[\Pi_{-\boldsymbol{v}_{ij}}\left[\mathcal{G}(\boldsymbol{w})\right],\Pi_{-v_j}\left[\mathcal{G}(\boldsymbol{w})\right]\right]$ in Eq. (B.1) can be similarly derived as the term $\Pi_{-\boldsymbol{v}_{ij}}\Pi_{-v_i}\left[k_0\left(\boldsymbol{w},\left(\boldsymbol{w}'_{-(d_1+i)},v_i\right)\right)\right]$ given in Eq. (B.4).

The covariance term $\mathbb{COV}_{\boldsymbol{\mathcal{D}}}\left[\Pi_{-\boldsymbol{v}_{ij}}\left[\mathcal{G}(\boldsymbol{w})\right],\Pi\left[\mathcal{G}(\boldsymbol{w})\right]\right]$ in Eq. (B.1) can be formulated as:

$$\mathbb{COV}_{\boldsymbol{\mathcal{D}}}\left[\Pi_{-\boldsymbol{v}_{ij}}\left[\mathcal{G}(\boldsymbol{w})\right],\Pi\left[\mathcal{G}(\boldsymbol{w})\right]\right]=\Pi_{-\boldsymbol{v}_{ij}}\Pi\left[k_0\left(\boldsymbol{w},\boldsymbol{w}'\right)\right]-\Pi_{-\boldsymbol{v}_{ij}}\left[\boldsymbol{k}_0\left(\boldsymbol{w},\boldsymbol{\mathcal{W}}\right)^{\mathrm{T}}\right]\boldsymbol{K}_0^{-1}\Pi\left[\boldsymbol{k}_0\left(\boldsymbol{w},\boldsymbol{\mathcal{W}}\right)\right],\tag{B.6}$$

where the terms $\Pi\left[\boldsymbol{k}_0\left(\boldsymbol{w},\boldsymbol{\mathcal{W}}\right)\right]$ and $\Pi_{-\boldsymbol{v}_{ij}}\left[\boldsymbol{k}_0\left(\boldsymbol{w},\boldsymbol{\mathcal{W}}\right)\right]$ have been given in Eq. (18) and Eq. 23 respectively; the term $\Pi_{-\boldsymbol{v}_{ij}}\Pi\left[k_0\left(\boldsymbol{w},\boldsymbol{w}'\right)\right]$ can be derived as:

$$\begin{aligned}
&\Pi_{-\boldsymbol{v}_{ij}}\Pi\left[k_0\left(\boldsymbol{w},\boldsymbol{w}'\right)\right]\\
=&s_0^2\left|2\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}+\boldsymbol{I}\right|^{-1/2}\\
&\cdot 2^{(d_2-2)}\mathrm{prod}_1\left\{\mathrm{diag}\left\{\boldsymbol{\Sigma}_{\boldsymbol{v}_{-ij}}\left[-1+\exp\left[-(2\boldsymbol{\Sigma}_{\boldsymbol{v}_{-ij}})^{-1}\right]+(2\pi^{-1}\boldsymbol{\Sigma}_{\boldsymbol{v}_{-ij}})^{-1/2}\mathrm{erf}\left(\left(2\boldsymbol{\Sigma}_{\boldsymbol{v}_{-ij}}\right)^{-1/2}\right)\right]\right\}\right\}\\
&\cdot\left(\frac{\pi}{2}\right)^{2/2}\mathrm{prod}_2\left\{\left[\mathrm{erf}\left((1-\boldsymbol{v}_{ij})\left(2\boldsymbol{\Sigma}_{\boldsymbol{v}_{ij}}\right)^{-1/2}\right)-\mathrm{erf}\left(-\boldsymbol{v}_{ij}\left(2\boldsymbol{\Sigma}_{\boldsymbol{v}_{ij}}\right)^{-1/2}\right)\right]\boldsymbol{\Sigma}_{\boldsymbol{v}_{ij}}^{1/2}\right\}.
\end{aligned}\tag{B.7}$$

The covariance term $\mathbb{COV}_{\boldsymbol{\mathcal{D}}}\left[\Pi_{-v_i}\left[\mathcal{G}(\boldsymbol{w})\right],\Pi_{-v_j}\left[\mathcal{G}(\boldsymbol{w})\right]\right]$ in Eq. (B.1) can be formulated as:

$$\mathbb{COV}_{\boldsymbol{\mathcal{D}}}\left[\Pi_{-v_i}\left[\mathcal{G}(\boldsymbol{w})\right],\Pi_{-v_j}\left[\mathcal{G}(\boldsymbol{w})\right]\right]=\Pi_{-v_i}\Pi_{-v_j}\left[k_0\left(\boldsymbol{w},\left(\boldsymbol{w}'_{-v'_j},v_j\right)\right)\right]-\Pi_{-v_i}\left[\boldsymbol{k}_0\left(\boldsymbol{w},\boldsymbol{\mathcal{W}}\right)^{\mathrm{T}}\right]\boldsymbol{K}_0^{-1}\Pi_{-v_j}\left[\boldsymbol{k}_0\left(\boldsymbol{w},\boldsymbol{\mathcal{W}}\right)\right],\tag{B.8}$$

where the terms $\Pi_{-v_i}\left[\boldsymbol{k}_0\left(\boldsymbol{w},\boldsymbol{\mathcal{W}}\right)\right]$ and $\Pi_{-v_j}\left[\boldsymbol{k}_0\left(\boldsymbol{w},\boldsymbol{\mathcal{W}}\right)\right]$ have been given in Eq. (21); the term $\Pi_{-v_i}\Pi_{-v_j}\left[k_0\left(\boldsymbol{w},\left(\boldsymbol{w}'_{-v'_j},v_j\right)\right.\right.$ is actually equal to $\Pi_{-\boldsymbol{v}_{ij}}\Pi\left[k_0\left(\boldsymbol{w},\boldsymbol{w}'\right)\right]$ as given in Eq. (B.7).

References

[1] A. Der Kiureghian, O. Ditlevsen, Aleatory or epistemic? does it matter?, Structural Safety 31 (2) (2009) 105–112.

[2] M. Beer, S. Ferson, V. Kreinovich, Imprecise probabilities in engineering analyses, Mechanical Systems and Signal Processing 37 (1-2) (2013) 4–29.

[3] A. N. Kolmogorov, A. T. Bharucha-Reid, Foundations of the theory of probability: Second English Edition, Courier Dover Publications, 2018.

[4] D. Moens, D. Vandepitte, A survey of non-probabilistic uncertainty treatment in finite element analysis, Computer Methods in Applied Mechanics and Engineering 194 (12-16) (2005) 1527–1555.

[5] M. Faes, D. Moens, Recent trends in the modeling and quantification of non-probabilistic uncertainty, Archives of Computational Methods in Engineering 27 (3) (2020) 633–671.

[6] C. Jiang, R. Bi, G. Lu, X. Han, Structural reliability analysis using non-probabilistic convex model, Computer Methods in Applied Mechanics and Engineering 254 (2013) 83–98.

[7] T. Augustin, F. P. Coolen, G. De Cooman, M. C. Troffaes, Introduction to imprecise probabilities, John Wiley & Sons, 2014.

[8] S. Ferson, V. Kreinovich, L. Grinzburg, D. Myers, K. Sentz, Constructing probability boxes and dempster-shafer structures, Tech. rep., Sandia National Lab.(SNL-NM), Albuquerque, NM (United States) (2015).

[9] Z. Zhang, C. Jiang, Evidence-theory-based structural reliability analysis with epistemic uncertainty: a review, Structural and Multidisciplinary Optimization (2021) 1–19.

[10] B. Möller, M. Beer, Fuzzy randomness: uncertainty in civil engineering and computational mechanics, Springer Science & Business Media, 2004.

[11] S.-K. Au, J. L. Beck, Estimation of small failure probabilities in high dimensions by subset simulation, Probabilistic Engineering Mechanics 16 (4) (2001) 263–277.

[12] M. D. Shields, J. Zhang, The generalization of latin hypercube sampling, Reliability Engineering & System Safety 148 (2016) 96–108.

[13] S. Geyer, I. Papaioannou, D. Straub, Cross entropy-based importance sampling using gaussian densities revisited, Structural Safety 76 (2019) 15–27.

[14] B. Keshtegar, Z. Meng, A hybrid relaxed first-order reliability method for efficient structural reliability analysis, Structural Safety 66 (2017) 84–93.

[15] X. Huang, Y. Li, Y. Zhang, X. Zhang, A new direct second-order reliability analysis method, Applied Mathematical Modelling 55 (2018) 68–80.

[16] B. Echard, N. Gayton, M. Lemaire, Ak-mcs: an active learning reliability method combining kriging and monte carlo simulation, Structural Safety 33 (2) (2011) 145–154.

[17] G. Blatman, B. Sudret, Adaptive sparse polynomial chaos expansion based on least angle regression, Journal of Computational Physics 230 (6) (2011) 2345–2367.

[18] R. Teixeira, M. Nogal, A. O'Connor, Adaptive approaches in metamodel-based reliability analysis: A review, Structural Safety 89 (2021) 102019.

[19] Y.-G. Zhao, T. Ono, New point estimates for probability moments, Journal of Engineering Mechanics 126 (4) (2000) 433–436.

[20] S. Xiao, Z. Lu, Reliability analysis by combining higher-order unscented transformation and fourth-moment method, ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering 4 (1) (2018) 04017034.

[21] J. Xu, C. Dang, A new bivariate dimension reduction method for efficient structural reliability analysis, Mechanical Systems and Signal Processing 115 (2019) 281–300.

[22] R. Liu, W. Fan, Y. Wang, A. H.-S. Ang, Z. Li, Adaptive estimation for statistical moments of response based on the exact dimension reduction method in terms of vector, Mechanical Systems and Signal Processing 126 (2019) 609–625.

[23] P. Wei, X. Zhang, M. Beer, Adaptive experiment design for probabilistic integration, Computer Methods in Applied Mechanics and Engineering 365 (2020) 113035.

[24] J. Li, J. Chen, Probability density evolution method for dynamic response analysis of structures with uncertain parameters, Computational Mechanics 34 (5) (2004) 400–409.

[25] G. Chen, D. Yang, A unified analysis framework of static and dynamic structural reliabilities based on direct probability integral method, Mechanical Systems and Signal Processing 158 (2021) 107783.

[26] D. Moens, M. Hanss, Non-probabilistic finite element analysis for parametric uncertainty treatment in applied mechanics: Recent advances, Finite Elements in Analysis and Design 47 (1) (2011) 4–16.

[27] B. Möller, W. Graf, M. Beer, Fuzzy structural analysis using α-level optimization, Computational Mechanics 26 (6) (2000) 547–565.

[28] C. Dang, P. Wei, M. G. Faes, M. A. Valdebenito, M. Beer, Interval uncertainty propagation by a parallel bayesian global optimization method, Applied Mathematical Modelling 108 (2022) 220–235.

[29] Z. Qiu, I. Elishakoff, Antioptimization of structures with large uncertain-but-non-random parameters via interval analysis, Computer Methods in Applied Mechanics and Engineering 152 (3-4) (1998) 361–372.

[30] S. Chen, H. Lian, X. Yang, Interval static displacement analysis for structures with interval parameters, International Journal for Numerical Methods in Engineering 53 (2) (2002) 393–407.

[31] R. R. Callens, M. G. Faess, D. Moens, Multilevel quasi-monte carlo for interval analysis, International Journal for Uncertainty Quantification 12 (4) (2022).

[32] B. Ni, C. Jiang, P. Wu, Z. Wang, W. Tian, A sequential simulation strategy for response bounds analysis of structures with interval uncertainties, Computers & Structures 266 (2022) 106785.

[33] M. C. Bruns, Propagation of imprecise probabilities through black box models, Ph.D. thesis, Georgia Institute of Technology (2006).

[34] H. Zhang, R. L. Mullen, R. L. Muhanna, Interval monte carlo methods for structural reliability, Structural Safety 32 (3) (2010) 183–190.

[35] S. Au, Reliability-based design sensitivity by efficient simulation, Computers & Structures 83 (14) (2005) 1048–1061.

[36] P. Wei, J. Song, S. Bi, M. Broggi, M. Beer, Z. Lu, Z. Yue, Non-intrusive stochastic analysis with parameterized imprecise probability models: I. performance estimation, Mechanical Systems and Signal Processing 124 (2019) 349–368.

[37] P. Wei, J. Song, S. Bi, M. Broggi, M. Beer, Z. Lu, Z. Yue, Non-intrusive stochastic analysis with parameterized imprecise probability models: II. reliability and rare events analysis, Mechanical Systems and Signal Processing 126 (2019) 227–247.

[38] M. G. Faes, M. A. Valdebenito, D. Moens, M. Beer, Bounding the first excursion probability of linear structures subjected to imprecise stochastic loading, Computers & Structures 239 (2020) 106320.

[39] C. Dang, P. Wei, J. Song, M. Beer, Estimation of failure probability function under imprecise probabilities by active learning–augmented probabilistic integration, ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering 7 (4) (2021) 04021054.

[40] P. Wei, F. Liu, M. Valdebenito, M. Beer, Bayesian probabilistic propagation of imprecise probabilities with large epistemic uncertainty, Mechanical Systems and Signal Processing 149 (2021) 107219.

[41] P. Wei, F. Hong, K.-K. Phoon, M. Beer, Bounds optimization of model response moments: a twin-engine bayesian active

learning method, Computational Mechanics 67 (5) (2021) 1273–1292.
[42] M. G. Faes, M. Daub, S. Marelli, E. Patelli, M. Beer, Engineering analysis with probability boxes: A review on computational methods, Structural Safety 93 (2021) 102092.
[43] X. Yang, Y. Liu, Y. Gao, Y. Zhang, Z. Gao, An active learning kriging model for hybrid reliability analysis with both random and interval variables, Structural and Multidisciplinary Optimization 51 (5) (2015) 1003–1016.
[44] X. Yang, Y. Liu, Y. Zhang, Z. Yue, Hybrid reliability analysis with both random and probability-box variables, Acta Mechanica 226 (5) (2015) 1341–1357.
[45] X. Yang, Y. Liu, Y. Zhang, Z. Yue, Probability and convex set hybrid reliability analysis based on active learning kriging model, Applied Mathematical Modelling 39 (14) (2015) 3954–3971.
[46] J. Zhang, M. Xiao, L. Gao, J. Fu, A novel projection outline based active learning method and its combination with kriging metamodel for hybrid reliability analysis with random and interval variables, Computer Methods in Applied Mechanics and Engineering 341 (2018) 32–52.
[47] X. Chen, Z. Qiu, A novel uncertainty analysis method for composite structures with mixed uncertainties including random and interval variables, Composite Structures 184 (2018) 400–410.
[48] M. Xiao, J. Zhang, L. Gao, S. Lee, A. T. Eshghi, An efficient kriging-based subset simulation method for hybrid reliability analysis under random and interval variables with small failure probability, Structural and Multidisciplinary Optimization 59 (6) (2019) 2077–2092.
[49] J. Song, P. Wei, M. Valdebenito, S. Bi, M. Broggi, M. Beer, Z. Lei, Generalization of non-intrusive imprecise stochastic simulation for mixed uncertain variables, Mechanical Systems and Signal Processing 134 (2019) 106316.
[50] X. Yuan, M. G. Faes, S. Liu, M. A. Valdebenito, M. Beer, Efficient imprecise reliability analysis using the augmented space integral, Reliability Engineering & System Safety 210 (2021) 107477.
[51] M. G. Faes, M. A. Valdebenito, X. Yuan, P. Wei, M. Beer, Augmented reliability analysis for estimating imprecise first excursion probabilities in stochastic linear dynamics, Advances in Engineering Software 155 (2021) 102993.
[52] C. Jiang, G. Lu, X. Han, L. Liu, A new reliability analysis method for uncertain structures with random and interval variables, International Journal of Mechanics and Materials in Design 8 (2) (2012) 169–182.
[53] M. G. Faes, M. A. Valdebenito, D. Moens, M. Beer, Operator norm theory as an efficient tool to propagate hybrid uncertainties and calculate imprecise probabilities, Mechanical Systems and Signal Processing 152 (2021) 107482.
[54] C. Jiang, J. Zheng, X. Han, Probability-interval hybrid uncertainty analysis for structures with both aleatory and epistemic uncertainties: a review, Structural and Multidisciplinary Optimization 57 (6) (2018) 2485–2502.
[55] J. Cockayne, C. J. Oates, T. J. Sullivan, M. Girolami, Bayesian probabilistic numerical methods, SIAM Review 61 (4) (2019) 756–789.
[56] D. R. Jones, M. Schonlau, W. J. Welch, Efficient global optimization of expensive black-box functions, Journal of Global Optimization 13 (4) (1998) 455–492.
[57] C. E. Rasmussen, C. K. I. Williams, Gaussian Processes for Machine Learning, MIT Press, 2006.
[58] C. E. Rasmussen, Z. Ghahramani, Bayesian monte carlo, Advances in Neural Information Processing Systems (2003) 505–512.
[59] G. Li, S.-W. Wang, H. Rabitz, High dimensional model representations (HDMR): Concepts and applications, in: Proceedings of the Institute of Mathematics and Its Applicatons Workshop on Atmospheric Modeling, Citeseer, 2000, pp. 15–19.
[60] D. Huang, T. T. Allen, W. I. Notz, N. Zeng, Global optimization of stochastic black-box systems via sequential kriging meta-models, Journal of Global Optimization 34 (3) (2006) 441–466.
[61] N. Gayton, J. M. Bourinet, M. Lemaire, CQ2RS: a new statistical approach to the response surface method for reliability analysis, Structural Safety 25 (1) (2003) 99–121.

[62] R. Schöbi, B. Sudret, Structural reliability analysis for p-boxes using multi-level meta-models, Probabilistic Engineering Mechanics 48 (2017) 27–38.

[63] D. R. Karanki, H. S. Kushwaha, A. K. Verma, S. Ajit, Uncertainty analysis based on probability bounds (p-box) approach in probabilistic safety assessment, Risk Analysis: An International Journal 29 (5) (2009) 662–675.