Bayesian updating with two-step parallel Bayesian optimization and quadrature

Masaru Kitahara [a], Chao Dang [a, *], Michael Beer [a, b, c]

[a] *Leibniz Universität Hannover, Institute for Risk and Reliability, Callinstrasse 34, Hannover, Germany*

[b] *University of Liverpool, Institute for Risk and Uncertainty, Peach Street L69 7ZF Liverpool, United Kingdom*

[c] *Tongji University, International Joint Research Center for Resilient Infrastructure & International Joint Research Center for Engineering Reliability and Stochastic Mechanics, Shanghai 200092, China*

* Correspondence author. E-mail address: chao.dang@irz.uni-hannover.de (C. Dang).

Abstract: This work proposes a Bayesian updating approach, called parallel Bayesian optimization and quadrature (PBOQ). It is rooted in Bayesian updating with structural reliability methods (BUS) and offers a coherent Bayesian approach for the BUS analysis by assuming Gaussian process priors. The first step of the method, i.e., parallel Bayesian optimization, effectively explores a constant c in BUS by a novel parallel infill sampling strategy. The second step (parallel Bayesian quadrature) then infers the posterior distribution by another parallel infill sampling strategy using subset simulation. The proposed approach enables to make the fullest use of prior knowledge and parallel computing, resulting in a substantial reduction of the computational burden of model updating. Four numerical examples with varying complexity are investigated for demonstrating the proposed method against several existing methods. The results show the potential benefits by advocating a coherent Bayesian fashion to the BUS analysis.

Keywords:
Bayesian model updating; Bayesian optimization; Bayesian quadrature; Gaussian process; Parallel computing

1. Introduction

In a number of engineering fields, numerical models are a popular tool for assessing the response of a physical system. However, there are inevitable discrepancies between model predictions and the actual behavior of the system. These discrepancies are mainly caused by a multitude of uncertainties, such as the modeling errors, measurement errors, and unknown or varying model inputs, that must be appropriately considered in the models. Bayesian model updating provides a robust and coherent probabilistic framework for calibrating the current model and reducing epistemic uncertainty on the inputs, given new system observations [1,2].

In Bayesian model updating, uncertainties are represnted by a prior distribution over the model inputs, and then updated to a posterior distribution using the likelihood function that quantifies the discrepancy between the model predictions and observations. In this context, the computation of the posterior distribution is a major task of Bayesian model updating, and the Markov chain Monte Carlo (MCMC) methods have constituted a widely used class of sampling methods to estimate the posterior [3-5]. In particular, to avoid the convergence issue in MCMC, Beck and Au [4] proposed the adaptive Metropolis-Hasting (AMH) algorithm, that gradually pushes samples from the prior to posterior by means of a sequence of the intermediate distributions which converge to the posterior. Subsequently, Ching and Chen [6] proposed the transitional Markov chain Monte Carlo (TMCMC) algorithm, which adopts a resampling scheme to improve the efficiency of the AMH algorithm. TMCMC has been used in numerous engineering applications attributed its capability in inferring large number of inputs at one time (i.e., up to 24) [7] and sampling from complex-shaped distributions [8,9].

More recently, another novel class of the sampling methods has been introduced by Straub and Papaioannou [10], called Bayesian updating with structural reliability methods (BUS). The principal idea behind BUS is reformulating the Bayesian updating problem into a rare event estimation; hence, it explores the possibility of using reliability analysis methods to draw samples from the posterior. In particular, the subset simulation techniques [11,12] have constituted a widely used reliability analysis method enabling efficient estimation of the probability of the rare event (i.e., failure probability), and have been incorporated into BUS [13,14]. The combination of BUS with subset simulation has shown great efficiency in estimating the posterior distribution; however, it still requires a significant number of likelihood evaluations which can be infeasible if the likelihood function involves time-consuming models, such as finite element (FE) models.

On the contrary, surrogate model-based methods have become more popular for estimating the failure probability. These methods aim to substitute the expensive-to-evaluate performance function for an inexpensive-to-evaluate surrogate using a limited number of observations of the performance function. Typical surrogate models include, e.g., response surfaces [15], polynomial chaos expansions [16,17], artificial neural networks [18], and Gaussian process regression (GPR, also known as Kriging) [19,20]. In order to avoid non-informative observations, there has been growing attention to the infill sampling criteria that effectively suggest more informative points which contribute to improving the accuracy of the surrogate model. In this context, the Kriging model is of particular interest due to its capability of quantifying uncertainty of the prediction, which can be used to define an infill sampling criterion and assess the accuracy of the surrogate [20]. Wang and Shafieezadeh [21] recently proposed a method, termed Bayesian updating using adaptive Kriging (BUAK), within the BUS framework, in which the so-called U learning function [20] with the Kriging model is employed as the infill sampling criterion. Moreover, the similar infill sampling criterion is adopted in Kitahara et al. [22] with subset simulation and in Song et al. [23] with sequential importance sampling.

Another challenge in BUS is the choice of a constant c in the context of the rejection principle. To guarantee the theoretical correctness of the analogy, on the one hand, its value must be less than the reciprocal of the maximum value of the likelihood function, which is generally unknown a priori. On the other hand, selecting its value conservatively small decreases the efficiency of the method. While its optimal choice can be achieved by solving an optimization problem that maximizes the likelihood function, it could be intractable if the likelihood function involves computationally expensive models. Alternatively, adaptive approaches have been also developed for the combination of BUS with subset simulation, where the constant c is adaptively learnt during the subset simulation procedure [13,14]. However, as already mentioned, these approaches can still render high computational burden. Rossat et al. [24] hence recently proposed to use the polynomial chaos Kriging (PCK) as the surrogate of the likelihood function within the adaptive approach combining BUS with subset simulation. Moreover, Liu et al. [25] developed a two-step approach, where the first step finds the constant c by constructing the Kriging surrogate of the likelihood function, whereas the second-step aims to sample the posterior distribution by constructing the Kriging surrogate of the performance function in BUS.

In summary, BUS can be interpreted to comprise two different tasks, i.e., the optimization of the likelihood function to explore the constant c and the quadrature of the failure probability to estimate the posterior distribution. So as to deal with these two tasks within a unified and efficient framework, this study proposes a novel method, termed two-step parallel Bayesian optimization and quadrature (PBOQ). Bayesian optimization [26] has been widely employed for the optimization of expensive-to-evaluate objective functions by treating the discretization errors due to a limited number of function evaluations as uncertainty, which is quantified and reduced within a Bayesian framework. Bayesian quadrature (also known as probabilistic integration) [27] similarly quantifies and reduces uncertainty which prevents us from inferring the true integral value according to Bayes' theorem. In recent years, it has been intensively investigated also to the rare event estimation by the second author and his co-workers [28,29].

The Bayesian optimization step starts to place a Gaussian process (GP) prior over the likelihood function, and then the prior is updated to a posterior over the likelihood function by observations of the likelihood function. The posterior, in turn, is employed to design an infill sampling criterion, and the expected improvement (EI) criterion [26] has been widely used in the field of global optimization. The EI criterion quantifies the expectation that any point in the search space will give a better solution than the current best solution within the maximization process. Hence, its convergence provides the maxima of the likelihood function, from which the constant c can be obtained. It should be noted that this step is similar to the first-step in Ref. [25], where the Kriging surrogate of the likelihood function is constructed by the EI criterion. However, this procedure is sequential in nature, hindering the use of ever-growing parallel computing facilities. To tackle the limitation, we propose a novel multi-point infill sampling criterion combining the k-means clustering with the EI criterion to enable identifying a batch of informative and diverse points at each iteration, and hence parallel distributed processing. After the constant c is determined, the Bayesian quadrature step then similarly places a GP prior over the performance function in BUS, and then it is updated to a posterior over the performance function by observations of the performance function. This, in turn, arrives at a posterior distribution over the failure probability, which is not achieved in existing surrogate methods. The posterior over the failure probability leads to an infill sampling criterion that enables better uncertainty reduction on the failure probability estimation, compared to the U learning function used in Refs. [21-25], since the latter gives only an indirect measure of the uncertainty on the failure probability estimation. Specifically, we use a multi-point infill sampling criterion which combines the k-means clustering with the upper-bound posterior variance contribution (UPVC) learning function [28]. We further develop a novel numerical integrator for the above Bayesian quadrature process by subset simulation to assess very small failure probabilities without generating a tremendously large number of samples. In this way, the proposed method provides a coherent Bayesian approach to the BUS analysis and its implementation supports fully parallel distributed processing.

The rest of this paper is organized as follows. In Section 2, we review the fundamental theory of Bayesian model updating and BUS. Section 3 outlines the details of the proposed Bayesian updating method: two-step parallel Bayesian optimization and quadrature (PBOQ). Then, the performance of the method is illustrated in Section 4 upon four numerical examples of increasing complexity. Finally, concluding remarks are presented in Section 5.

2. Bayesian updating

A key advantage of Bayesian model updating lies in its ability to combine the prior knowledge on the model with some new observations to yield a stochastic characterization of model inputs to be inferred.

Let $\boldsymbol{x} \in \mathcal{D}_{\boldsymbol{x}}$ mean the model inputs of dimension p and $\boldsymbol{y}$ be m newly available observations that are gathered in a vector. The prior belief on the inputs $\boldsymbol{x}$ represented by a probability density function (PDF) is updated using the well-known Bayes' theorem as [1]:

$$P(\boldsymbol{x}|\boldsymbol{y}) = \frac{L(\boldsymbol{y}|\boldsymbol{x})P(\boldsymbol{x})}{c_E} \tag{1}$$

where $P(\boldsymbol{x})$ denotes the prior distribution of $\boldsymbol{x}$, reflecting one's initial belief on $\boldsymbol{x}$; $P(\boldsymbol{x}|\boldsymbol{y})$ indicates the posterior distribution which represents the posterior state of knowledge on $\boldsymbol{x}$; $L(\boldsymbol{y}|\boldsymbol{x})$ is the likelihood function that is theoretically defined as the probability density of $\boldsymbol{y}$ given $\boldsymbol{x}$; c_E refers to the so-called evidence that normalizes the posterior distribution.

In order to link observations $\boldsymbol{y}$ to model predictions $M(\boldsymbol{x})$, the deviation, $\boldsymbol{\varepsilon} = \boldsymbol{y} - M(\boldsymbol{x})$, between them that is caused by modeling errors and measurement errors is modeled by the PDF $f_{\boldsymbol{\varepsilon}}$. This leads to the likelihood function formulated as:

$$L(\boldsymbol{y}|\boldsymbol{x}) = f_{\boldsymbol{\varepsilon}}\big(\boldsymbol{y} - M(\boldsymbol{x})\big) \tag{2}$$

While $f_{\boldsymbol{\varepsilon}}$ is typically assumed to be a multivariate Gaussian distribution with zero mean, it can be any other unbiased distribution. For m independent observations, the likelihood function can be written as:

$$L(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^{m} L_i(y_i|\boldsymbol{x}) = \prod_{i=1}^{m} f_{\varepsilon_i}\big(y_i - M_i(\boldsymbol{x})\big) \tag{3}$$

where L_i means the likelihood function of the ith observation y_i; f_{ε_i} refers to the PDF of the deviation, ε_i, between y_i and the corresponding model prediction $M_i(\boldsymbol{x})$.

Another key component in Eq. (1) is the evidence c_E (also known as marginal likelihood), whose definition is expressed as:

$$c_E = \int_{\mathcal{D}_x} L(\boldsymbol{y}|\boldsymbol{x})P(\boldsymbol{x})d\boldsymbol{x} \tag{4}$$

The evidence c_E is a measure of the plausibility of the assumed model class. In the context of Bayesian model class selection [30], it allows to evaluate the posterior plausibility of each model class to decide the most probable model. Hence, it is advantageous if c_E can be estimated as a by-product of Bayesian updating methods.

2.1. Bayesian updating with structural reliability methods (BUS)

The main idea of BUS is based on the simple rejection principle, introducing an auxiliary random variable which is uniformly distributed, $\pi \in \mathcal{D}_\pi = [0,1]$, to the input space $\mathcal{D}_{\boldsymbol{x}}$. The Bayesian updating problem in Eq. (1) is then regarded as a reliability analysis problem in the augmented space $\mathcal{D}_{\boldsymbol{x}} \times \mathcal{D}_\pi$, where the failure domain Ω and corresponding performance function $g(\boldsymbol{x},\pi)$ are defined as:

$$\Omega = \{[\boldsymbol{x},\pi] \in \mathcal{D}_{\boldsymbol{x}} \times \mathcal{D}_\pi | g(\boldsymbol{x},\pi) \leq 0\} \tag{5}$$

$$g(\boldsymbol{x},\pi) = \pi - c \cdot L(\boldsymbol{y}|\boldsymbol{x}) \tag{6}$$

Where c is a constant chosen such that $c \cdot L(\boldsymbol{y}|\boldsymbol{x}) \leq 1$ holds true for any $\boldsymbol{x}$. In this context, the quantity $c \cdot L(\boldsymbol{y}|\boldsymbol{x})$ can be expressed as [14]:

$$c \cdot L(\boldsymbol{y}|\boldsymbol{x}) = \int_{0 \leq \pi \leq c \cdot L(\boldsymbol{y}|\boldsymbol{x})} d\pi \tag{7}$$

Consequently, Eq. (1) is written as:

$$P(\boldsymbol{x}|\boldsymbol{y}) = c_E^{-1} c^{-1} \int_{0 \leq \pi \leq c \cdot L(\boldsymbol{y}|\boldsymbol{x})} P(\boldsymbol{x}) d\pi \tag{8}$$

The integral in the right-hand side of Eq. (8) can be performed by sampling in the failure domain Ω according to the prior distribution $P(\boldsymbol{x})$. Hence, sampling from the posterior distribution is converted into sampling from the failure domain Ω. Moreover, the evidence in Eq. (4) can be expressed as:

$$c_E = c^{-1} \int_{\mathcal{D}_x} \int_{0 \leq \pi \leq c \cdot L(\boldsymbol{y}|\boldsymbol{x})} P(\boldsymbol{x}) d\pi \, d\boldsymbol{x} = c^{-1} P_f \tag{9}$$

where P_f denotes the failure probability corresponding to the failure domain Ω.

A key component in the BUS formulation is the constant c. As already mentioned, its reciprocal needs to be selected as close as possible to the maximum value of the likelihood function, while such value is generally not known a priori. The task to find the constant c can be regarded as the following optimization problem:

$$c^{-1} = \max_{x \in \mathcal{D}_x} L(\boldsymbol{y}|\boldsymbol{x}) \tag{10}$$

where the likelihood function $L(\boldsymbol{y}|\boldsymbol{x})$ is treated as a black-box model. In general, global optimization algorithms, such as the genetic algorithm, need a large number of function evaluations to solve such black-box problem; and thus can be computationally intractable especially if each such evaluation is expensive. In order to alleviate the computational burden, a parallel Bayesian optimization approach is developed in the following section as the first step of the proposed PAOQ.

After the constant c is chosen, the remaining task is the quadrature of the integrals given in Eqs. (8) and (9). Analytical solutions to these integrals are unavailable for the performance function in BUS which is generally given as black-box. Thus, numerical approximation techniques are necessary, and subset simulation and surrogate model-based methods have been successfully utilized in this context. Existing methods, however, still suffers from there respective limitations as already discussed, which motivates us to develop a parallel Bayesian quadrature approach as the second step of the proposed PBOQ.

3. Two-step parallel Bayesian optimization and quadrature (PBOQ)

3.1. Preliminary

For convenience, let us first transform the performance function in Eq. (6) to the standard normal space $\mathcal{D}_{\boldsymbol{u}}$:

$$g(\boldsymbol{u}) = \Phi(u_1) - c \cdot L\big(\boldsymbol{y}|T(\boldsymbol{u}^*)\big) \tag{11}$$

whereby $\boldsymbol{u} = [u_1, \boldsymbol{u}^*] \in \mathcal{D}_{\boldsymbol{u}} \subseteq \mathbb{R}^d$ with $u_1 = \Phi^{-1}(\pi)$ and $\boldsymbol{u}^* = T^{-1}(\boldsymbol{x})$ indicates a vector of $d(= p + 1)$ independent standard normal variables that follow the joint PDF $f_{\boldsymbol{U}}(\boldsymbol{u})$; $\Phi(\cdot)$ indicates the cumulative distribution function (CDF) of a standard normal variable; $T^{-1}(\cdot)$ is the inverse transformation, such as Nataf or Rosenblatt transformation.

By doing so, the optimization problem in Eq. (10) can be expressed as:

$$c^{-1} = \max_{\boldsymbol{u}^* \in \mathcal{D}_{\boldsymbol{u}}} L(\boldsymbol{y}|\boldsymbol{u}^*) \tag{12}$$

The likelihood function $L(\boldsymbol{y}|\boldsymbol{u}^*)$ is often strongly nonlinear, which makes it inappropriate to place a GP prior over the likelihood function. Thus, we use the logarithm $\mathcal{L}\big(\boldsymbol{y}|T(\boldsymbol{u}^*)\big) = \log L\big(\boldsymbol{y}|T(\boldsymbol{u}^*)\big)$ of the likelihood function instead, leading to Eq. (12) being rewritten as: (leaving dependence on $\boldsymbol{y}$ implicit)

$$c = \exp\left(- \max_{\boldsymbol{u}^* \in \mathcal{D}_{\boldsymbol{u}}} \mathcal{L}(\boldsymbol{u}^*)\right) \tag{13}$$

Similarly, Eq. (11) is rewritten as:

$$\mathcal{g}(\boldsymbol{u}) = \log \Phi(u_1) - \log c - \mathcal{L}(\boldsymbol{u}^*) \tag{14}$$

Its failure indicator function can be then given by:

$$I(\boldsymbol{u}) = \begin{cases} 1, \mathcal{g}(\boldsymbol{u}) \leq 0 \\ 0, \text{otherwise} \end{cases} \tag{15}$$

where $Z = \{\boldsymbol{u} \in \mathcal{D}_{\boldsymbol{u}} | \mathcal{g}(\boldsymbol{u}) \leq 0\}$ refers to the failure domain in the standard normal space. Accordingly, the failure probability is expressed as:

$$P_f = \int_{\mathcal{D}_{\boldsymbol{u}}} I(\boldsymbol{u}) f_{\boldsymbol{U}}(\boldsymbol{u}) d\boldsymbol{u} \tag{16}$$

which will be inferred by the Bayesian quadrature procedure.

3.2. First-step: Parallel Bayesain optimization

Under the black-box assumption, no knowledge on the inner structure of the likelihood function, e.g., concavity and linearity, is available, while we can evaluate the function at some points to observe its values. According to the Bayes' theorem, our prior beliefs on the likelihood function before seeing any observations can be modeled by placing a prior distribution over the likelihood function. In this study, we employ a GP prior over the log-likelihood function $\mathcal{L}(\boldsymbol{u}^*)$ instead of the original likelihood function:

$$\mathcal{L}_0 \sim \mathcal{GP}\left(m_{\mathcal{L}_0}(\boldsymbol{u}^*), k_{\mathcal{L}_0}(\boldsymbol{u}^*, \boldsymbol{u}^{*\prime})\right) \tag{17}$$

where $\mathcal{L}_0$ means the prior distribution of $\mathcal{L}$ prior to seeing any observations; $m_{\mathcal{L}_0}(\boldsymbol{u}^*)$ and $k_{\mathcal{L}_0}(\boldsymbol{u}^*, \boldsymbol{u}^{*\prime})$ denote the prior mean and covariance functions, respectively. Among many options available in the literature, we utilize the widely used constant mean and squared exponential kernel functions:

$$m_{\mathcal{L}_0}(\boldsymbol{u}^*) = b \tag{18}$$

$$k_{\mathcal{L}_0}(\boldsymbol{u}^*, \boldsymbol{u}^{*\prime}) = \sigma_0^2 \exp\left(-\frac{1}{2}\sum_{i=2}^{d}\left(\frac{u_i - u_i'}{l_i}\right)^2\right) \tag{19}$$

where σ_0^2 means the process variance, l_i is the correlation length in the ith direction. The GP prior is thus parametrized by a set of $d + 1$ hyperparameters, i.e., $\boldsymbol{\theta} = [b, \sigma_0, l_2, \dots, l_d]$.

Now assume that we have obtained some observations by evaluating the log-likelihood function at some points. Let $\widehat{\boldsymbol{u}}^* = \{\boldsymbol{u}^{*(j)}\}_{j=1}^{n}$ be an $n \times d$ matrix with its jth row being jth observed point $\boldsymbol{u}^{*(j)}$, and $\widehat{\boldsymbol{z}}_1 = \left\{z_1^{(j)}\right\}_{j=1}^{n}$ be an $n \times 1$ vector with its jth element being jth observed value, $z_1^{(j)} = \mathcal{L}\left(\boldsymbol{u}^{*(j)}\right)$, of the log-likelihood function. The hyperparameters $\boldsymbol{\theta}$ involved in the GP prior can be learned from the observations $\boldsymbol{\mathcal{D}}_1 = \{\widehat{\boldsymbol{u}}^*, \widehat{\boldsymbol{z}}_1\}$ by obtaining the maximum likelihood estimate, see, e.g., Ref. [31].

Conditioning on the observations $\boldsymbol{\mathcal{D}}_1$, the posterior distribution over the log-likelihood function turns out to be a GP:

$$\mathcal{L}_n \sim \mathcal{GP}\left(m_{\mathcal{L}_n}(\boldsymbol{u}^*), k_{\mathcal{L}_n}(\boldsymbol{u}^*, \boldsymbol{u}^{*\prime})\right) \tag{20}$$

where $\mathcal{L}_n$ means the posterior distribution of $\mathcal{L}$ after seeing n observations; $m_{\mathcal{L}_n}(\boldsymbol{u}^*)$ and $k_{\mathcal{L}_n}(\boldsymbol{u}^*, \boldsymbol{u}^{*\prime})$ denote the posterior mean and covariance functions, closed-form expressions of which are available as:

$$m_{\mathcal{L}_n}(\boldsymbol{u}^*) = m_{\mathcal{L}_0}(\boldsymbol{u}^*) + k_{\mathcal{L}_0}(\boldsymbol{u}^*, \widehat{\boldsymbol{u}}^*)^T \boldsymbol{K}_{\mathcal{L}_0}^{-1}(\widehat{\boldsymbol{u}}^*, \widehat{\boldsymbol{u}}^*)\left(\widehat{\boldsymbol{z}}_1 - m_{\mathcal{L}_0}(\widehat{\boldsymbol{u}}^*)\right) \tag{21}$$

$$k_{\mathcal{L}_n}(\boldsymbol{u}^*, \boldsymbol{u}^{*\prime}) = k_{\mathcal{L}_0}(\boldsymbol{u}^*, \boldsymbol{u}^{*\prime}) - k_{\mathcal{L}_0}(\boldsymbol{u}^{*\prime}, \widehat{\boldsymbol{u}}^*)^T \boldsymbol{K}_{\mathcal{L}_0}^{-1}(\widehat{\boldsymbol{u}}^*, \widehat{\boldsymbol{u}}^*) k_{\mathcal{L}_0}(\widehat{\boldsymbol{u}}^*, \boldsymbol{u}^{*\prime}) \tag{22}$$

where $m_{\mathcal{L}_0}(\widehat{\boldsymbol{u}}^*)$ is an $n \times 1$ mean vector, whose jth element is $m_{\mathcal{L}_0}\left(\boldsymbol{u}^{*(j)}\right)$, $k_{\mathcal{L}_0}(\boldsymbol{u}^*, \widehat{\boldsymbol{u}}^*)$ means an $n \times 1$ covariance vector between $\boldsymbol{u}^*$ and $\widehat{\boldsymbol{u}}^*$, with its jth element being $k_{\mathcal{L}_0}\left(\boldsymbol{u}^*, \boldsymbol{u}^{*(j)}\right)$; $k_{\mathcal{L}_0}(\boldsymbol{u}^{*\prime}, \widehat{\boldsymbol{u}}^*)$ is defined in a similar way to $k_{\mathcal{L}_0}(\boldsymbol{u}^*, \widehat{\boldsymbol{u}}^*)$; $\boldsymbol{K}_{\mathcal{L}_0}(\widehat{\boldsymbol{u}}^*, \widehat{\boldsymbol{u}}^*)$ represents an $n \times n$ covariance matrix between $\widehat{\boldsymbol{u}}^*$ and $\widehat{\boldsymbol{u}}^*$, whose (j, l)th entry being $k_{\mathcal{L}_0}\left(\boldsymbol{u}^{*(j)}, \boldsymbol{u}^{*(l)}\right)$. In this context, the posterior mean function $m_{\mathcal{L}_n}(\boldsymbol{u}^*)$ can be used as a predictor, whereas the posterior variance function $\sigma_{\mathcal{L}_n}^2(\boldsymbol{u}^*) = k_{\mathcal{L}_n}(\boldsymbol{u}^*, \boldsymbol{u}^*)$ enables to measure the prediction uncertainty.

In order to infer the maximum of the log-likelihood function using as few function evaluations as possible, our main concern is to design an infill sampling criterion to effectively select points where the log-likelihood function should be observed. This task can be achieved by combining the k-means clustering with the EI criterion. Let $\mathcal{L}_{\max} = \max_{1 \le j \le n} z_1^{(j)}$ be the current best solution observed so far. The improvement at the point $\boldsymbol{u}^*$ over $\mathcal{L}_{\max}$ can be defined as [32]:

$$I_{\max}(\boldsymbol{u}^*) = \max\left(m_{\mathcal{L}_n}(\boldsymbol{u}^*) - \mathcal{L}_{\max}, 0\right) = \begin{cases} m_{\mathcal{L}_n}(\boldsymbol{u}^*) - \mathcal{L}_{\max}, \text{if } m_{\mathcal{L}_n}(\boldsymbol{u}^*) > \mathcal{L}_{\max} \\ 0, \text{otherwise} \end{cases} \tag{23}$$

The EI criterion over the current maximum consists of taking expectation of $I_{\max}(\boldsymbol{u}^*)$, and is derived in a closed-form expression as:

$$EI_{\max}(\boldsymbol{u}^*) = \left(m_{\mathcal{L}_n}(\boldsymbol{u}^*) - \mathcal{L}_{\max}\right)\Phi\left(\frac{m_{\mathcal{L}_n}(\boldsymbol{u}^*) - \mathcal{L}_{\max}}{\sigma_{\mathcal{L}_n}(\boldsymbol{u}^*)}\right) + \sigma_{\mathcal{L}_n}(\boldsymbol{u}^*)\phi\left(\frac{m_{\mathcal{L}_n}(\boldsymbol{u}^*) - \mathcal{L}_{\max}}{\sigma_{\mathcal{L}_n}(\boldsymbol{u}^*)}\right) \tag{24}$$

where $\phi(\cdot)$ indicates the PDF of a standard normal variable. This criterion measures the improvement of the current best solution at the point $\boldsymbol{u}^*$.

On the other hand, the k-means clustering [33] enables to partition a dataset into k clusters that are given by k centroids. However, the conventional k-means clustering algorithm does not consider weight information of the data. To tackle the limitation, we propose a weighted clustering algorithm, termed EI-weighted k-means clustering, which identify k centroids by using N_1 samples $\{\boldsymbol{u}^{*(j)}\}_{j=1}^{N_1}$ of $\boldsymbol{u}^*$ while considering their EI values as weights. Therefore, the k centroids correspond to the batch of points we wish to select. Once the k points are obtained, evaluation of the true log-likelihood function on these points can be distributed in parallel. A pseudocode of the weighted clustering algorithm is shown in Algorithm 1. The above infill sampling process is continued until our improvement will be sufficiently small. In view of this, we also propose a stopping criterion as:

$$\frac{\max\left(EI_{\max}(\boldsymbol{u}^*)\right)}{\mathcal{L}_{\max} - \mathcal{L}_{\min}} < \epsilon_1 \tag{25}$$

where $\mathcal{L}_{\min}$ is the current worst solution; ϵ_1 is a pre-determined tolerance. If the stopping criterion is satisfied twice in succession, the constant c can be obtained as $\hat{c} = \exp(-\mathcal{L}_{\max})$.

Algorithm 1 Weighted k-means clustering algorithm

Input: The weight function (i.e., EI criterion), number of clusters k, and dataset $\{\boldsymbol{u}^{*(j)}\}_{j=1}^{N_1}$

1. **Initialization.** Randomly select k points from the dataset $\{\boldsymbol{u}^{*(j)}\}_{j=1}^{N_1}$ as the initial centroids, denoted by $\boldsymbol{\mathcal{S}} = \{\boldsymbol{s}^{(i)}\}_{i=1}^{k}$;
2. **Assignment step.** Assign each point among $\{\boldsymbol{u}^{*(j)}\}_{j=1}^{N_1}$ to the nearest cluster by the least squared Euclidian distance. The ith cluster is denoted as $\boldsymbol{\mathcal{C}}^{(i)} = \left\{\boldsymbol{c}_j^{(i)}\right\}_{j=1}^{N^{(i)}}$, where $\boldsymbol{c}_j^{(i)}$ indicates the jth point in the ith cluster; $N^{(i)}$ is the number of points in the ith cluster;
3. **Update step.** The ith centroid is updated by the weighted mean of the points belonging to the ith cluster:
$$\boldsymbol{s}^{(i)} = \frac{\sum_{j=1}^{N^{(i)}} EI_{\max}\left(\boldsymbol{c}_j^{(i)}\right) \times \boldsymbol{c}_j^{(i)}}{\sum_{j=1}^{N^{(i)}} EI_{\max}\left(\boldsymbol{c}_j^{(i)}\right)}$$
4. **Iteration.** Repeat Steps 2 and 3 until the centroids do not change or the pre-specified number (e.g., 100) of iteration is reached.

Output: k centroids

Lastly, it should be noted that the best solution $\mathcal{L}_{\max}$ identified by the aforementioned Bayesian optimization procedure will be very likely smaller than the true maximum value of the log-likelihood function. Nevertheless, this does not prevent us from producing samples which follow the posterior distribution, provided that the number of samples N_1 is selected large enough. The readers can refer to Ref. [13] for the detailed discussion on the asymptotic validity of the constant c chosen through the simulation.

3.3. Second-step: Parallel Bayesain quadrature

Due to large discontinuity of the failure indicator function in Eq. (16), it is challenging to directly place a GP prior over it. Alternatively, the parallel Bayesian quadrature step starts to place a GP prior over the performance function in Eq. (14):

$$g_0 \sim \mathcal{GP}\left(m_{g_0}(\boldsymbol{u}), k_{g_0}(\boldsymbol{u}, \boldsymbol{u}')\right) \tag{26}$$

where g_0 indicates the prior distribution of g prior to seeing any observations; $m_{g_0}(\boldsymbol{u})$ and $k_{g_0}(\boldsymbol{u}, \boldsymbol{u}')$ denote the prior mean and covariance functions, respectively. As in the previous step, the prior mean function is assumed to be an unknown constant and the prior covariance function utilizes the squared exponential kernel.

Suppose that we have obtained some observations $\boldsymbol{\mathcal{D}}_2 = \{\hat{\boldsymbol{u}}, \hat{\boldsymbol{z}}_2\}$, where $\hat{\boldsymbol{z}}_2 = \left\{z_2^{(j)}\right\}_{j=1}^{n}$ denotes an $n \times 1$ vector with its ith element being ith observed performance function value $z_2^{(j)} = g\left(\boldsymbol{u}^{(j)}\right)$. Note that, we can reuse the observations of the log-likelihood function in the previous step to obtain initial observations of g as $\hat{\boldsymbol{z}}_2 = \left\{z_2^{(j)}\right\}_{j=1}^{n_1}$, where $z_2^{(j)} = \log \Phi\left(u_1^{(j)}\right) - \log c - z_1^{(j)}$; n_1 means the total number of samples in which the log-likelihood function is evaluated in the previous step; $u_1^{(j)}$ is the ith newly generated standard normal sample. According to Bayes' theorem, the posterior distribution of g that is conditional on $\boldsymbol{\mathcal{D}}_2$ turns out to be a GP:

$$g_n \sim \mathcal{GP}\left(m_{g_n}(\boldsymbol{u}), k_{g_n}(\boldsymbol{u}, \boldsymbol{u}')\right) \tag{27}$$

where g_n indicates the posterior distribution of g after seeing n observations; $m_{g_n}(\boldsymbol{u})$ and $k_{g_n}(\boldsymbol{u}, \boldsymbol{u}')$ denote the posterior mean and covariance functions, closed-form expressions of which are available as similar to Eqs. (21) and (22); thus, we do not repeat them for the sake of brevity.

This then follows that the posterior distribution of the failure indicator function I conditional on $\boldsymbol{\mathcal{D}}_2$ admits a generalized Bernoulli process [28]:

$$I_n \sim \mathcal{GBP}\left(m_{I_n}(\boldsymbol{u}), k_{I_n}(\boldsymbol{u}, \boldsymbol{u}')\right) \tag{28}$$

where I_n denotes the posterior distribution of I after obtaining n observations; $m_{I_n}(\boldsymbol{u})$ and $k_{I_n}(\boldsymbol{u},\boldsymbol{u}')$ represent the posterior mean and covariance functions of I, respectively. The posterior mean function $m_{I_n}(\boldsymbol{u})$ is given by:

$$m_{I_n}(\boldsymbol{u}) = \Phi\left(-\frac{m_{\mathscr{g}_n}(\boldsymbol{u})}{\sigma_{\mathscr{g}_n}(\boldsymbol{u})}\right) \tag{29}$$

where $\sigma_{\mathscr{g}_n}(\boldsymbol{u}) = \sqrt{k_{\mathscr{g}_n}(\boldsymbol{u},\boldsymbol{u})}$ is the posterior standard deviation function of $\mathscr{g}$. Instead of the posterior covariance function, $k_{I_n}(\boldsymbol{u},\boldsymbol{u}')$, no closed-form expression of which is available, the posterior variance function $\sigma_{I_n}^2(\boldsymbol{u})$ is given by:

$$\sigma_{I_n}^2(\boldsymbol{u}) = \Phi\left(-\frac{m_{\mathscr{g}_n}(\boldsymbol{u})}{\sigma_{\mathscr{g}_n}(\boldsymbol{u})}\right)\Phi\left(\frac{m_{\mathscr{g}_n}(\boldsymbol{u})}{\sigma_{\mathscr{g}_n}(\boldsymbol{u})}\right) \tag{30}$$

The induced posterior distribution $P_{f,n}$ of the failure probability P_f should hence follow a certain random variable. Whilst its exact distribution type is not known yet, its posterior mean $m_{P_{f,n}}$ and an upper-bound of its posterior variance $\bar{\sigma}_{P_{f,n}}^2$ can be given as [28]:

$$m_{P_{f,n}} = \int_{\mathcal{D}_{\boldsymbol{u}}} m_{I_n}(\boldsymbol{u}) f_{\boldsymbol{U}}(\boldsymbol{u}) d\boldsymbol{u} = \int_{\mathcal{D}_{\boldsymbol{u}}} \Phi\left(-\frac{m_{\mathscr{g}_n}(\boldsymbol{u})}{\sigma_{\mathscr{g}_n}(\boldsymbol{u})}\right) f_{\boldsymbol{U}}(\boldsymbol{u}) d\boldsymbol{u} \tag{31}$$

$$\begin{aligned}\bar{\sigma}_{P_{f,n}}^2 &= \int_{\mathcal{D}_{\boldsymbol{u}}}\int_{\mathcal{D}_{\boldsymbol{u}}} \sigma_{I_n}(\boldsymbol{u})\sigma_{I_n}(\boldsymbol{u}') f_{\boldsymbol{U}}(\boldsymbol{u}) f_{\boldsymbol{U}}(\boldsymbol{u}') d\boldsymbol{u} d\boldsymbol{u}' \\ &= \left(\int_{\mathcal{D}_{\boldsymbol{u}}} \sqrt{\Phi\left(-\frac{m_{\mathscr{g}_n}(\boldsymbol{u})}{\sigma_{\mathscr{g}_n}(\boldsymbol{u})}\right)\Phi\left(\frac{m_{\mathscr{g}_n}(\boldsymbol{u})}{\sigma_{\mathscr{g}_n}(\boldsymbol{u})}\right)} f_{\boldsymbol{U}}(\boldsymbol{u}) d\boldsymbol{u}\right)^2\end{aligned} \tag{32}$$

In this context, the posterior mean $m_{P_{f,n}}$ can be employed as the failure probability estimator, and the upper-bound of the posterior variance $\bar{\sigma}_{P_{f,n}}^2$ measures the maximum possible prediction uncertainty. It is noted that the exact posterior variance of the failure probability can be also derived [29], whereas its computation is quite expensive. Hence, we use its upper-bound instead to measure the prediction uncertainty. The interested readers refer to Ref. [29] for the derivation of the posterior variance of the failure probability and Refs. [34,35] for its Monte Carlo simulation (MCS) estimator.

This further brings an open task to approximate the analytically intractable integrals in Eqs. (31) and (32). The most straightforward solution is to use the crude MCS. However, a prohibitively large number of samples is necessary to achieve a satisfactory accuracy if the true failure probability is very small (e.g., $P_f \leq 10^{-4}$). This can make the Bayesian quadrature procedure time-consuming and even cause the memory problem. Since samples located in the failure domain Z contribute the most to the integrals, methods which allow more efficient exploration of the failure domain can provide effective numerical integrators for these integrals, and subset simulation is one of such methods that have been widely employed within the BUS framework.

Subset simulation [11,12] is an adaptive MCMC approach and its principal idea is to express the failure domain Z by means of a sequence of intermediate nested domains $\mathcal{D}_{\boldsymbol{u}} = Z_0 \supset Z_1 \supset \cdots \supset Z_r = Z$. The intermediate domain is defined as $Z_i = \{\mathscr{g}(\boldsymbol{u}) \leq b_i\}$, where b_i is the threshold that holds $+\infty = b_0 > b_1 > \cdots > b_r = 0$. As such, the failure probability can be written as:

$$P_f = P\left(\bigcap_{i=0}^{r} Z_j\right) = \prod_{i=1}^{r} P(Z_i|Z_{i-1}) \tag{33}$$

As similar to Eqs. (31) and (32), the posterior distribution $\mathscr{g}_n$ of the g-function results in the posterior distribution of the conditional probability $P(Z_i|Z_{i-1})$, the posterior mean and an upper-bound of the posterior variance of which are expressed by:

$$m_{P(Z_i|Z_{i-1})_n} = \int_{\mathcal{D}_{\boldsymbol{u}}} \Phi\left(-\frac{m_{\mathscr{g}_n}(\boldsymbol{u}) - b_i}{\sigma_{\mathscr{g}_n}(\boldsymbol{u})}\right) f_{\boldsymbol{U}}(\boldsymbol{u}) d\boldsymbol{u} \tag{34}$$

$$\bar{\sigma}^2_{P(Z_i|Z_{i-1})_n} = \left(\int_{\mathcal{D}_u} \sqrt{\Phi\left(-\frac{m_{g_n}(\boldsymbol{u}) - b_i}{\sigma_{g_n}(\boldsymbol{u})}\right) \Phi\left(\frac{m_{g_n}(\boldsymbol{u}) - b_i}{\sigma_{g_n}(\boldsymbol{u})}\right)} f_{\boldsymbol{U}}(\boldsymbol{u}) d\boldsymbol{u} \right)^2 \tag{35}$$

Their numerical integrators are then given by:

$$\tilde{m}_{P(Z_i|Z_{i-1})_n} = \frac{1}{N_2} \sum_{j=1}^{N_2} \Phi\left(-\frac{m_{g_n}\left(\boldsymbol{u}_{i-1}^{(j)}\right) - b_i}{\sigma_{g_n}\left(\boldsymbol{u}_{i-1}^{(j)}\right)}\right) \tag{36}$$

$$\tilde{\bar{\sigma}}^2_{P(Z_i|Z_{i-1})_n} = \left(\frac{1}{N_2} \sum_{j=1}^{N_2} \sqrt{\Phi\left(-\frac{m_{g_n}\left(\boldsymbol{u}_{i-1}^{(j)}\right) - b_i}{\sigma_{g_n}\left(\boldsymbol{u}_{i-1}^{(j)}\right)}\right) \Phi\left(\frac{m_{g_n}\left(\boldsymbol{u}_{i-1}^{(j)}\right) - b_i}{\sigma_{g_n}\left(\boldsymbol{u}_{i-1}^{(j)}\right)}\right)} \right)^2 \tag{37}$$

where $\left\{\boldsymbol{u}_{i-1}^{(j)}\right\}_{j=1}^{N_2}$ represents N_2 samples of $\boldsymbol{u}$ drawn in Z_{i-1}. Samples in $Z_0 (= \mathcal{D}_u)$ can be simply drawn from the joint PDF $f_{\boldsymbol{U}}(\boldsymbol{u})$ by MCS, whereas for each $i \in \{1, \cdots, r-1\}$, samples in Z_i are generated by MCMC procedure. The threshold values $\{b_i: 1 \le i \le r-1\}$ are adaptively chosen as the P_t percentile of $m_{g_n}(\boldsymbol{u}_{i-1})$ so that the corresponding conditional probabilities $\{P(Z_i|Z_{i-1}): 1 \le i \le r-1\}$ are equal to a pre-defined target probability P_t. By doing so, the posterior estimate of the failure probability is given by $\tilde{m}_{P_{f,n}} = \prod_{i=1}^{r} \tilde{m}_{P(Z_i|Z_{i-1})_n}$. This, in turn, implies the posterior estimate of the evidence as $\tilde{c}_E = c^{-1}\tilde{m}_{P_{f,n}}$.

Another issue to be solved within the Bayesian quadrature framework is how to design an infill sampling criterion to effectively select points where the performance function is observed. As in the previous step, we propose a parallel infill sampling criterion combining the k-means clustering with the UPVC learning function [28], which attempts to make the fullest possible use of the posterior GP and parallel computing simultaneously. The UPVC learning function for the conditional probability $P(Z_i|Z_{i-1})$ can be defined as:

$$\mathrm{UPVC}(\boldsymbol{u}_{i-1}) = \sqrt{\Phi\left(-\frac{m_{g_n}\left(\boldsymbol{u}_{i-1}^{(j)}\right) - b_i}{\sigma_{g_n}\left(\boldsymbol{u}_{i-1}^{(j)}\right)}\right) \Phi\left(\frac{m_{g_n}\left(\boldsymbol{u}_{i-1}^{(j)}\right) - b_i}{\sigma_{g_n}\left(\boldsymbol{u}_{i-1}^{(j)}\right)}\right)} f_{\boldsymbol{U}}(\boldsymbol{u}_{i-1}) \tag{38}$$

It should be noted that, the UPVC learning function holds that $\bar{\sigma}_{P(Z_i|Z_{i-1})_n} = \int_{\mathcal{D}_u} \mathrm{UPVC}(\boldsymbol{u}_{i-1}) d\boldsymbol{u}$. Thus, it measures the contribution of our uncertainty on the conditional probability estimation at the point $\boldsymbol{u}_{i-1}$. By taking the UPVC values as weights, a weighted k-means clustering algorithm, that is similar to the one in Algorithm 1, determines a batch of k points using the dataset $\left\{\boldsymbol{u}_{i-1}^{(j)}\right\}_{j=1}^{N_2}$. For convenience, the number of samples k in a batch is assumed to be same as in the previous step, while it should not to be. Once the k points are given, evaluation of the true g-function on these points can be distributed in parallel. This, in turn, updates the posterior GP in Eq. (27), and hence the threshold value b_i. The above infill sampling process is continued until our uncertainty will be sufficiently small. In view of this, this study also propose a stopping criterion as:

$$\frac{\tilde{\bar{\sigma}}_{P(Z_i|Z_{i-1})_n}}{\tilde{m}_{P(Z_i|Z_{i-1})_n}} < \epsilon_2 \tag{39}$$

where the left-hand side of the inequality is the estimated upper-bound of the posterior coefficient of variation (COV) of the conditional probability; ϵ_2 denotes a pre-determined tolerance. Note that, the true estimation COV would be certainly small compared to the upper-bound; thus, a relatively large value can be selected as the tolerance. If the stopping criterion is satisfied, b_i is obtained and samples in Z_i can be generated by the current posterior GP of g.

3.4. Numerical implementation procedure

The numerical implementation procedure of the proposed PBOQ, which is also summarized in Fig. 1, consists of the following main steps:

Step 1.1: Generate standard normal samples

To check the stopping criterion and enrich the dataset, a set of N_1 standard normal samples need to be generated, which are denoted as $\boldsymbol{\mathcal{U}}^* = \{\boldsymbol{u}^{*(j)}\}_{j=1}^{N_1}$.

Step 1.2: Obtain an initial dataset $\boldsymbol{\mathcal{D}}_1$ from the log-likelihood function

In order to start the first step, i.e., parallel Bayesian optimization, of the proposed PBOQ , a set of n_0 initial samples of $\hat{\boldsymbol{u}}^* = \{\boldsymbol{u}^{*(j)}\}_{j=1}^{n_0}$ needs to be drawn by Latin hypercube sampling (LHS). In this study, the number of initial samples n_0 is selected as $n_0 = 10$. These n_0 samples are evaluated on the log-likelihood function in parallel, and the corresponding observations are denoted by $\hat{\boldsymbol{z}}_1 = \left\{z_1^{(j)}\right\}_{j=1}^{n_0}$. The initial dataset is then constructed by $\boldsymbol{\mathcal{D}}_1 = \{\hat{\boldsymbol{u}}^*, \hat{\boldsymbol{z}}_1\}$. Let $n = n_0$.

Step 1.3: Infer the GP posterior of $\mathcal{L}$

The GP posterior of $\mathcal{L}$ conditional on $\boldsymbol{\mathcal{D}}_1$ can be inferred as Eq. (20). This step mainly consists of learning the hyper-parameters using maximum likelihood estimation. All the numerical examples in this study are performed by the *fitrgp* function in MATLAB Statistics and Machine Learning Toolbox.

Step 1.4: Check the stopping criterion

If the stopping criterion given in Eq. (25) is satisfied twice in succession, go to **Step 1.6**; else, go to **Step 1.5**. In this study, the tolerance ϵ_1 is specified as $\epsilon_1 = 0.01$.

Step 1.5: Enrich the dataset by the EI-weighted k-means clustering

This step consists of identifying k new points $\hat{\boldsymbol{u}}_+^* = \left\{\boldsymbol{u}_+^{*(j)}\right\}_{j=1}^{k}$ from $\boldsymbol{\mathcal{U}}^*$ using the EI-weighted k-means clustering. In this study, the number of samples k in a batch is assumed to be $k = 4$. Then, the corresponding observations of $\mathcal{L}(\boldsymbol{u}^*)$ at these k points are obtained in parallel, which are denoted by $\hat{\boldsymbol{z}}_{1+} = \left\{z_{1+}^{(j)}\right\}_{j=1}^{k}$. The dataset $\boldsymbol{\mathcal{D}}_1$ is enriched with $\boldsymbol{\mathcal{D}}_{1+} = \{\hat{\boldsymbol{u}}_+^*, \hat{\boldsymbol{z}}_{1+}\}$, i.e., $\boldsymbol{\mathcal{D}}_1 = \boldsymbol{\mathcal{D}}_1 \cup \boldsymbol{\mathcal{D}}_{1+}$. Let $n = n_0 + k$, and go to **Step 1.3**.

Step 1.6: End of the first step

The first step stops and the constant c is obtained from the last best solution as $\hat{c} = \exp(-\mathcal{L}_{\max})$. Let $n = n_1$ and $i = 1$.

Step 2.1: Generate samples conditional on Z_{i-1}

To approximate $\widetilde{m}_{P(Z_i|Z_{i-1})_n}$ and $\tilde{\tilde{\sigma}}^2_{P(Z_i|Z_{i-1})_n}$, check the stopping criterion, and enrich the dataset, a set of N_2 samples, $\boldsymbol{\mathcal{U}}_{i-1} = \left\{\boldsymbol{u}_{i-1}^{(j)}\right\}_{j=1}^{N_2}$, conditional on Z_{i-1} need to be generated by MCS for $i = 1$; else, by MCMC using $\mathcal{g}_n$ obtained in **Step 2.3**. If $i = 1$, go to **Step 2.2**; else, go to **Step 2.4**.

Step 2.2: Obtain an initial dataset $\boldsymbol{\mathcal{D}}_2$ from the g-function

In order to start the second step, i.e., parallel Bayesian quadrature, of the proposed PBOQ, a set of n_1 standard normal samples of $\hat{\boldsymbol{u}}_1 = \left\{u_1^{(j)}\right\}_{j=1}^{n_1}$ needs to be produced. The initial sample points are then denoted as $\hat{\boldsymbol{u}} = \{\hat{\boldsymbol{u}}_1, \hat{\boldsymbol{u}}^*\}$. The corresponding observations are obtained using $\hat{\boldsymbol{z}}_1$ and are denoted by $\hat{\boldsymbol{z}}_2 = \left\{z_2^{(j)}\right\}_{j=1}^{n_1}$. The initial dataset is then constructed by $\boldsymbol{\mathcal{D}}_2 = \{\hat{\boldsymbol{u}}, \hat{\boldsymbol{z}}_2\}$.

Step 2.3: Infer the GP posterior of $\mathcal{g}$

The GP posterior of $\mathcal{g}$ conditional on $\boldsymbol{\mathcal{D}}_2$ can be inferred as Eq. (27).

Step 2.4: Choose the threshold b_i

The threshold b_i is chosen as the P_t percentile of $m_{\mathcal{g}_n}(\boldsymbol{\mathcal{U}}_{i-1})$. In this study, the target probability P_t is selected as $P_t = 0.1$.

Step 2.5: Infer the GP posterior of $P(Z_i|Z_{i-1})$

The GP posterior of $P(Z_i|Z_{i-1})$ conditional on $\boldsymbol{\mathcal{D}}_2$ can be inferred using $\mathcal{g}_n$ and b_i. The posterior mean $m_{P(Z_i|Z_{i-1})_n}$ and an upper-bound of the posterior variance $\bar{\sigma}^2_{P(Z_i|Z_{i-1})_n}$ are approximated by Eqs. (36) and (37).

Step 2.6: Check the stopping criterion

If the stopping criterion given in Eq. (39) is not satisfied, go to **Step 2.7**. Otherwise, if $b_i \leq 0$, go to **Step 2.8**; else, let $i = i + 1$ and go to **Step 2.1**. In this study, the tolerance ϵ_2 is specified as $\epsilon_2 = 0.1$.

Step 2.7: Enrich the dataset by the UPVC-weighted k-means clustering

This step consists of selecting k new points $\hat{\boldsymbol{u}}_+ = \left\{\boldsymbol{u}_+^{(j)}\right\}_{j=1}^{k}$ from $\boldsymbol{\mathcal{U}}_{i-1}$ using the UPVC-weighted k-means clustering. The corresponding observations of $\mathcal{g}(\boldsymbol{u})$ at these points are obtained in parallel, denoted by $\hat{\boldsymbol{z}}_{2+} = \left\{z_{2+}^{(j)}\right\}_{j=1}^{k}$. The dataset $\boldsymbol{\mathcal{D}}_2$ is enriched with $\boldsymbol{\mathcal{D}}_{2+} = \{\hat{\boldsymbol{u}}_+, \hat{\boldsymbol{z}}_{2+}\}$, i.e., $\boldsymbol{\mathcal{D}}_2 = \boldsymbol{\mathcal{D}}_2 \cup \boldsymbol{\mathcal{D}}_{2+}$. Let $n = n_1 + k$, and go to **Step 2.3**.

Step 2.8: End of the second step

The second step stops and the evidence is given as $\tilde{c}_E = \hat{c}^{-1}\tilde{m}_{P_{f,n}}$, with $\tilde{m}_{P_{f,n}} = \prod_{i=1}^{r} \tilde{m}_{P(Z_i|Z_{i-1})_n}$. Furthermore, the samples conditional on the failure domain can be drawn by the last GP posterior of $\mathcal{g}$, from which samples that follow the posterior distribution $P(\boldsymbol{x}|\boldsymbol{y})$ are obtained by using the inverse transformation.

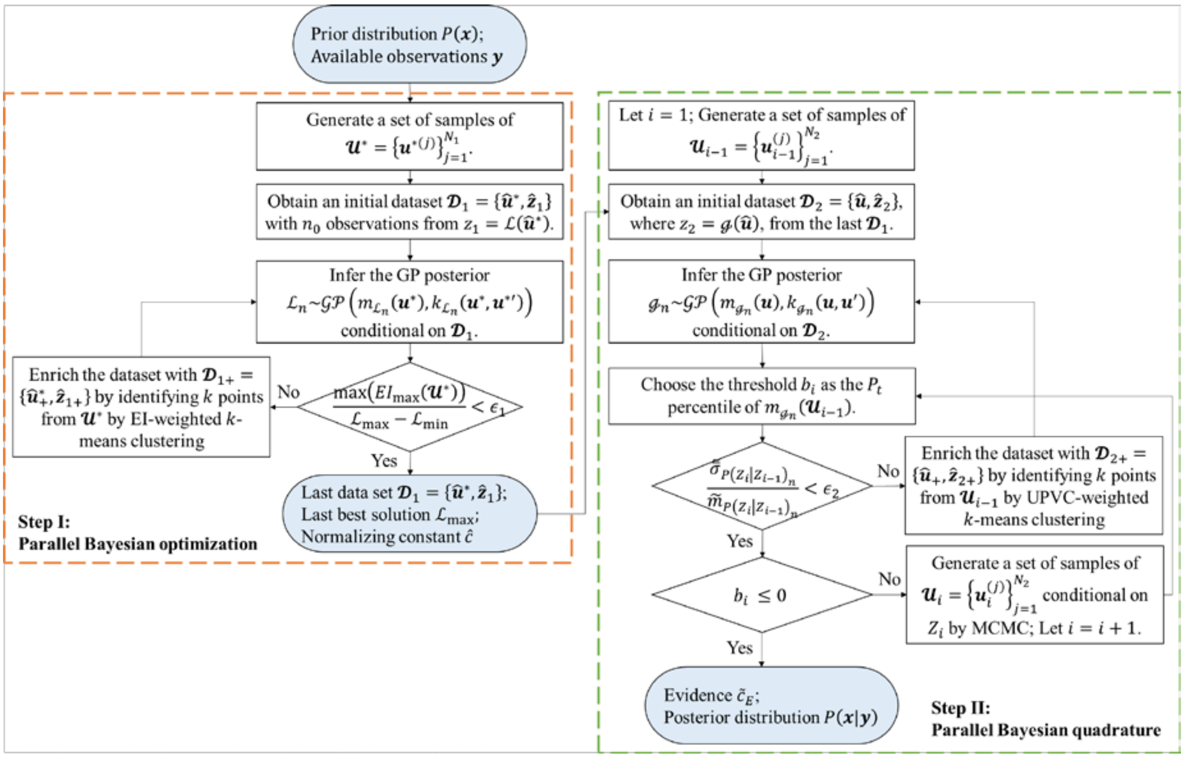

Fig. 1. Schematic of the proposed PBOQ method.

4. Numerical applications

In the following section, four numerical applications with varying complexities are investigated to demonstrate the efficiency and accuracy of the PBOQ method. For comparison, other state-of-the-art methods within the BUS framework, i.e., adaptive BUS (aBUS) [13], aBUS with PCK (aBUS-PCK) [24], BUS with two-step adaptive Kriging (BUS-AK²) [25], BUAK with subset simulation (BUAK-SuS) [21], BUAK with adaptive importance sampling (BUAK-AIS) [23], and BUS with adaptive Kriging

Markov chain Monte Carlo (BUS-AK-MCMC) [22] are implemented if applicable. Note that, the first three methods enable to infer the constant c within their procedures, whereas the others do not. Thus, the optimal choice of the constant c is directly employed in the BUAK-SuS, BUAK-AIS, and BUS-AK-MCMC methods.

4.1. One-dimensional illustrative application

A toy example proposed in Ref. [24] is investigated as the first example for illustrative purposes. The problem involves a one-dimensional random variable, x, the prior distribution of which follows a standard normal distribution. The likelihood function can be expressed as:

$$L(y|x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^2}(M(x)-y)^2\right], \text{with } M(x) = 1 + \cos\left(\frac{x}{2}\right) + 3\exp(-4(x-2)^2) \tag{40}$$

where $y = M(2)$ is an observation; $\sigma = 0.4$ denotes the standard deviation of the prediction error.

A semi-analytical solution of the problem is available and used as the benchmark. The posterior mean and standard deviation are $\mu = 1.942$ and $\sigma = 0.134$, respectively. In addition, the evidence is $c_E = 0.024$ and the optimal choice of the constant c is $c_{\text{opt}} = 1$. This means that the failure probability associated with the reliability problem in BUS approaches $P_f = 0.024$.

Table 1 summarizes the results of the proposed PBOQ method as well as aBUS, aBUS-PCK, and BUS-AK-MCMC. In the table, N denotes the number of samples per each subset in aBUS, aBUS-PCK, and BUS-AK-MCMC, N_{call} is the total number of model evaluations, N_{iter} is the number of iterations required for the infill sampling procedure in aBUS-PCK, BUS-AK-MCMC, and the proposed method, and $\hat{\mu}$ and $\hat{\sigma}$ mean the estimated posterior mean and standard deviation, respectively. The results of aBUS-PCK are directly taken from Ref. [24]. AK-BUS-MCMC and the PBOQ method are carried out 20 independent runs and the results are averaged. In aBUS and BUS-AK-MCMC, the parameters n_0 and P_t are set to be the same as those in the PBOQ . It is worth mentioning that, from an implementing viewpoint, BUS-AK-MCMC and the second step of the PBOQ are very similar procedures, where the differences are only the infill sampling criterion and stopping criterion employed.

Table 1. Bayesian updating results of the one-dimensional application.

Method		N, N_1/N_2	N_{call}	N_{iter}	$\hat{c}$	$\hat{\mu}/\mu$	$\hat{\sigma}/\sigma$
aBUS		3×10^3	9×10^3	–	1.0027	1.0001	1.0012
aBUS-PCK [24]	$k = 25$	5×10^3	$50+50=100$	3	–	1.0000	1.0299
BUS-AK-MCMC		1×10^4	$10+68.3=78.3$	69.3	–	0.9994	1.0010
Proposed PBOQ	$k = 4$	$1\times10^5/1\times10^4$	$10+14.4+13.2=29.5$	7.9	1.0044	0.9997	0.9833

Note: The results of aBUS-PCK are directly taken from Ref. [24] and averaged over 50 independent runs.

As can be seen in Table 1, with 9000 model evaluations, aBUS enables to provide accurate results with $\hat{c} = 1.0027$, $\hat{\mu}/\mu = 0.9695$, and $\hat{\sigma}/\sigma = 1.0228$. By employing the PCK or Kriging surrogate, N_{call} can be significantly reduced, such that aBUS-PCK and BUS-AK-MCMC provide accurate results with 100 and 78.3 model evaluations on average, respectively. Although aBUS-PCK enables inferring the the constant c during the subset simulation procedure, its identified value is not provided in Ref. [24]; thus it is not given in Table 1. Compared to these two methods, the PBOQ method requires much less model evaluations, while its accuracy ($\hat{c} = 1.0044$, $\hat{\mu}/\mu = 0.9997$, and $\hat{\sigma}/\sigma = 0.9833$) is sufficient. This indicates that the infill sampling criterion and stopping criterion by the UPVC learning function are more effective to select points on which the performance function is observed, compared to those by the conventional U learning function. Furthermore, aBUS-PCK and the proposed PBOQ can sensibly reduce N_{iter} compared to the non-parallel methods by the k-means clustering. Finally, the proposed PBOQ enables to provide an estimate for the evidence as $\hat{c}_E = 0.023$ as a byproduct, which is close to the true value $c_E = 0.024$.

Fig. 2 presents an illustration of the first step, i.e., parallel Bayesian optimization, of the proposed PBOQ. It can be observed that the parallel Bayesian optimization step gradually approaches the exact global maximum of the log-likelihood function as the infill sampling process goes on. Moreover, these added points are more densely distributed around the global maximum, and hence very informative for our purpose.

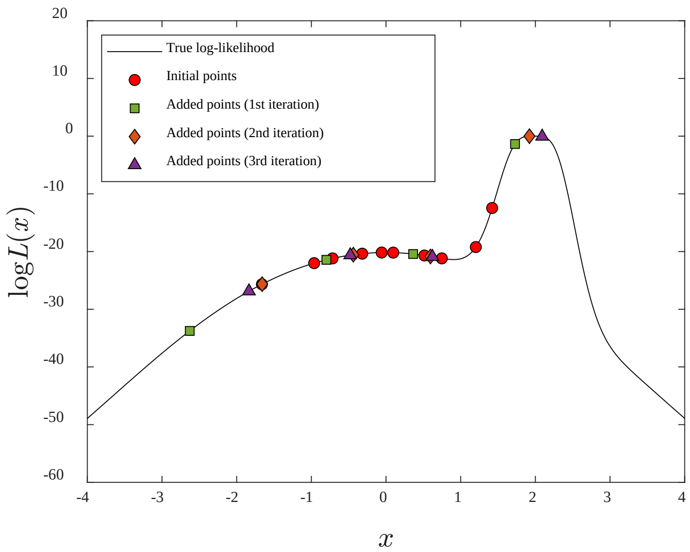

Fig. 2. Illustration of the first step of PBOQ.

Fig. 3 shows the posterior distribution estimated as a histogram by 1000 realizations of x that are drawn by the inverse transformation of u located in the failure domain. As can be seen, the posterior samples quite accurately approximate the analytical posterior distribution.

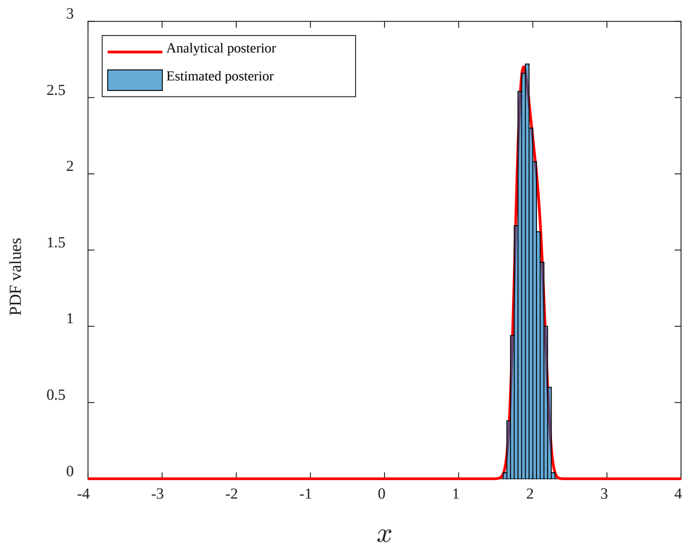

Fig. 3. Estimated posterior distribution.

4.2. Unimodal distribution application

As the second application, a unimodal distribution problem from Ref. [14,21,23] is investigated to demonstrate the proposed method for high-dimensional problems. The problem involves the prior distribution as the product of p independent standard normal distributions. The likelihood function can be expressed as:

$$P_L(\boldsymbol{x}) = \prod_{i=1}^{p} \frac{1}{\sigma_l} \phi\left(\frac{x_i - \mu_l}{\sigma_l}\right) \tag{41}$$

where σ_l is a constant value 0.2, and μ_l is given as:

$$\mu_l = \sqrt{-2(1+\sigma_l^2)\ln\left[c_E^{1/p}\sqrt{2\pi\sqrt{1+\sigma_l^2}}\right]} \tag{42}$$

where c_E represents the model evidence. In this study, two cases, i.e., Case I: $k = 2$ and $c_E = 10^{-4}$ and Case II: $k = 10$ and $c_E = 10^{-5}$ are considered.

An analytical solution is available for each case and is utilized as the benchmark. The mean and standard deviation of the posterior distribution for Case I are $\mu = 2.659$ and $\sigma = 0.1961$ while those for Case II are $\mu = 0.6542$ and $\sigma = 0.1961$. Furthermore, the optimal choice of the constant c is $c_{\text{opt}} = 0.0503$ for Case I and $c_{\text{opt}} = 2.00 \times 10^{-4}$ for Case II. By accounting for the chosen evidence, the failure probability approaches the order of 10^{-6} for Case I and 10^{-9} for Case II.

The results for Case I are shown in Table 2. The results of BUAK-SuS and BUAK-AIS are directly taken from Ref. [23]. BUS-AK-MCMC and the proposed method are performed 20 independent runs and the results are averaged. As can be seen, aBUS can provide accurate results but the computational burden is substantially large. By adopting the Kriging surrogate, N_{call} can be drastically reduced for BUAK-SuS, BUAK-AIS, and BUS-AK-MCMC. Among them, BUAK-AIS is the most efficient and only 12.9 g-function calls need on average. All these methods enable to provide accurate results. Similarly, the proposed PBOQ can also provide accurate estimates ($\hat{\mu}/\mu = 1.0032$ and $\hat{\sigma}/\sigma = 0.9649$). Note that, PBOQ requires more model evaluations than BUAK-AIS, because the former learns the constant c in its first step whilst the latter adopts its optimal choice. Nevertheless, PBOQ requires only 26.8 model evaluations on average and achieves the lowest N_{iter} by exploring parallel computing. Moreover, the estimated evidence by the PBOQ method is $\hat{c}_E = 2.68 \times 10^{-4}$, that is permissible compared to the true value $c_E = 10^{-4}$.

Table 2. Bayesian updating results of the unimodal distribution application (Case I).

Method	N, N_1/N_2	N_{call}	N_{iter}	$\hat{c}$	$\hat{\mu}/\mu$	$\hat{\sigma}/\sigma$
aBUS	3×10^3	2.1×10^4	–	0.0503	1.0011	1.0295
BUAK-SuS [23]	1×10^3	$7 + 24 = 31$	25	–	1.0271	1.0544
BUAK-AIS [23]	1×10^3	$7 + 5.9 = 12.9$	6.9	–	1.0030	1.0226
BUS-AK-MCMC	1×10^4	$10 + 34.7 = 44.7$	35.7	–	1.0044	0.9168
Proposed PBOQ	$1 \times 10^5/1 \times 10^4$	$10 + 16.4 + 0.4 = 26.8$	6.1	0.0611	1.0032	0.9649

Note: The results of BUAK-SuS and BUAK-AIS are taken from Ref. [23] and averaged over 20 runs.

Fig. 4 gives an illustration of the second step, i.e., parallel Bayesian quadrature, of the proposed PBOQ. Fig. 4 (a) shows realizations of the sample set $\{\boldsymbol{x} = [x_1, x_2]\}$ of each subset. In total, five subsets are utilized until convergence, implying that the failure probability reaches the order of 10^{-5}, that is very small to effectively explore by MCS. As can be seen, the subset gradually approaches the failure domain, and the satisfactory acceptance rate (i.e., 21.3 %) is achieved in the final subset. The accepted samples are also shown in Fig. 4 (b) as the posterior samples, which indicate a good agreement with the analytical posterior distribution. Moreover, the observed points are also shown in Fig. 4 (b). Note that, in this specific run, all the points are added in the first step, implying that these observations are sufficient to precisely infer the failure probability, and hence the posterior distribution. As such, the proposed method can reuse the observations in its first step to further reduce the computational cost.

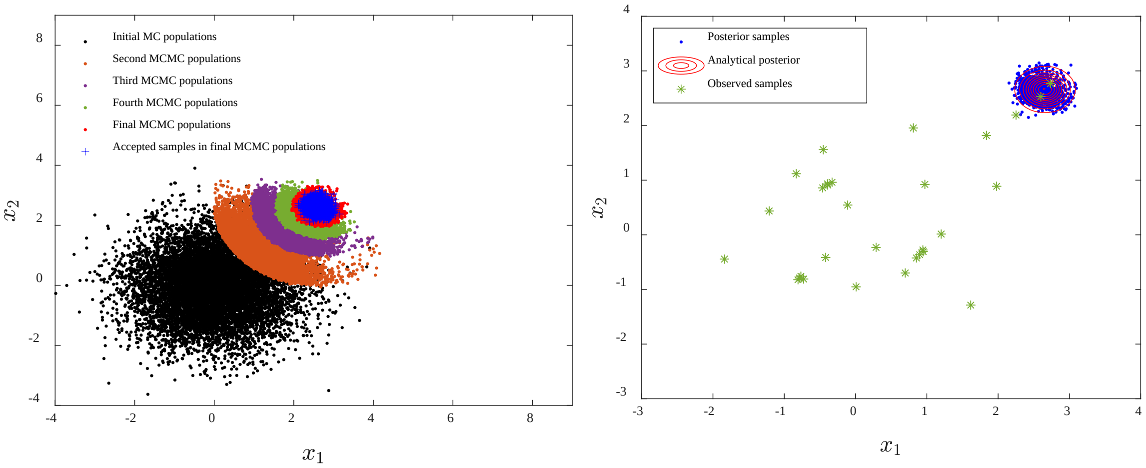

Fig. 4. Illustration of the second step of PBOQ: (a) realizations of each subset; (b) posterior samples.

Besides, the results for Case II are detailed in Table 3. aBUS can provide accurate results but the computational cost is even larger compared to Case I. By employing the Kriging surrogate, N_{call} can be significantly reduced for BUAK-SuS, BUAK-AIS, and BUS-AK-MCMC. Among them, BUAK-AIS is the most efficient and 90.5 g-function calls need on average. Compared to these methods, the PBOQ method is capable of providing accurate estimates ($\hat{\mu}/\mu = 1.0266$ and $\hat{\sigma}/\sigma = 0.9899$). The larger N_{call} compared to BUAK-SuS and BUAK-AIS is justified accounting for its ability to obtain the constant c. Until convergence, the second step of PBOQ adopts nine subsets, implying that the failure probability reaches the order of 10^{-9}. The estimated evidence by the proposed method is $\hat{c}_E = 3.20 \times 10^{-5}$, which is permissible compared to the true value $c_E = 10^{-5}$. Consequently, this example shows the capability of PBOQ addressing moderately high-dimensional problems.

Table 3. Bayesian updating results of the unimodal distribution application (Case II).

Method	N, N_1/N_2	N_{call}	N_{iter}	$\hat{c}$	$\hat{\mu}/\mu$	$\hat{\sigma}/\sigma$
aBUS	3×10^3	3×10^4	–	3.95×10^{-4}	0.9955	0.9886
BUAK-SuS [23]	1×10^4	$67 + 36 = 103$	37	–	1.0271	1.0544
BUAK-AIS [23]	1×10^4	$67 + 23.5 = 90.5$	24.5	–	1.0030	1.0226
BUS-AK-MCMC	1×10^4	$10 + 213.5 = 223.5$	214.5	–	0.9927	0.8607
Proposed PBOQ	$1 \times 10^6/1 \times 10^4$	$10 + 34.4 + 80.8 = 125.2$	38.8	1.78×10^{-3}	1.0266	0.9899

Note: The results of BUAK-SuS and BUAK-AIS are taken from Ref. [23] and averaged over 20 runs.

4.3. Two degree of freedom shear building model application

A two degree of freedom (DOF) structural dynamic problem [4,13,21] is investigated as the third application. The configuration of the system is shown in Fig. 5 (a). The first and second story masses are considered to be fixed values, i.e., $m_1 = 16.531 \times 10^3$ kg and $m_2 = 16.131 \times 10^3$ kg. Moreover, the first and second interstory stiffnesses are modeled as $k_1 = \bar{k}x_1$ and $k_2 = \bar{k}x_2$, where $\boldsymbol{x} = [x_1, x_2]$ is the inferred parameters, and $\bar{k} = 29.7 \times 10^6$ N/m denotes the nominal value. The prior distribution of $\boldsymbol{x}$ follows an uncorrelated log-normal distribution with the modes 1.3 and 0.8 for x_1 and x_2 respectively and the unit standard deviations. By adopting the first two natural frequencies $\hat{f}_1 = 3.13$ Hz and $\hat{f}_2 = 9.83$ Hz as the observation (i.e., $\boldsymbol{y} = [\hat{f}_1, \hat{f}_2]$), the likelihood function is expressed as:

$$L(\boldsymbol{y}|\boldsymbol{x}) \propto \exp\left[-\frac{M(\boldsymbol{x})}{2\sigma_\varepsilon^2}\right] \tag{43}$$

where $\sigma_\varepsilon = 1/16$; $M(\boldsymbol{x})$ is the modal measure-of-fit function given by:

$$M(\boldsymbol{x}) = \sum_{i=1}^{2} \lambda_i^2 \left[\frac{f_i^2(\boldsymbol{x})}{\hat{f}_i^2} - 1\right]^2 \tag{44}$$

where $\lambda_i = 1$ denotes the mean of the prediction error for ith natural frequency; $f_i(\boldsymbol{x})$ refers to the ith natural frequency evaluated as the model output. Fig. 5 (b) illustrates the posterior distribution of $\boldsymbol{x}$, which clearly indicates the non-uniqueness of the solution.

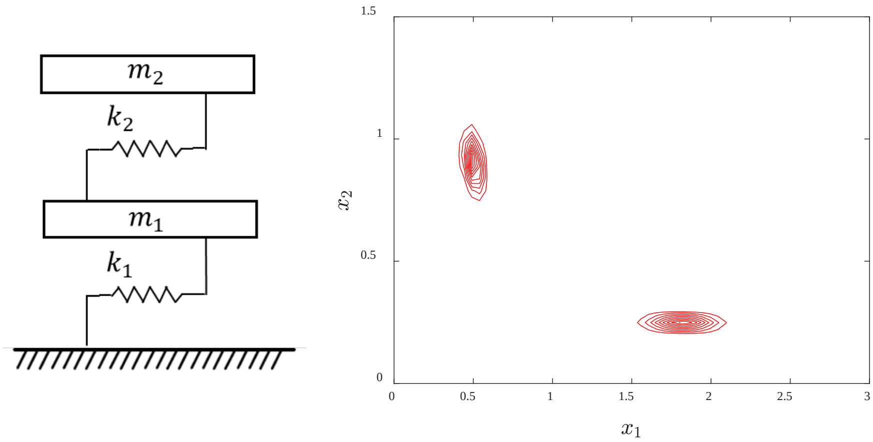

Fig. 5. (a) 2-DOF shear building model; (b) Posterior distribution of $\boldsymbol{x}$.

A semi-analytical solution of the problem is available and used as the benchmark. The posterior mean and standard deviation are $\mu = 1.125$ and $\sigma = 0.659$ for x_1 and $\mu = 0.583$ and $\sigma = 0.326$ for x_2, respectively. Furthermore, the evidence and the optimal choice of the constant c are $c_E = 0.0015$ and $c_{\text{opt}} = 1$, respectively; hence, the failure probability would be $P_f = 0.0015$.

The results are summarized in Tables 4 and 5. The results of aBUS-PCK and BUS-AK² are directly taken from Refs. [24] and [25], respectively, whilst the proposed PBOQ is carried out 20 independent runs and the results are averaged. As can be seen, aBUS enables to provide accurate estimates but the computational cost is substantially large. By using the PCK or Kriging surrogate, N_{call} is successfully reduced for aBUS-PCK and BUS-AK², whereas both methods can provide satisfactory accurate results. Note that, the posterior estimates are provided in a different form (i.e., for each mode identified) in Ref. [24], and thus the results are not shown in Table 5. The readers can refer to Ref. [24] for the results of aBUS-PCK. Compared to these two methods, the proposed PBOQ method achieves the lowest N_{call} and N_{iter}, while its accuracy ($\hat{\mu}/\mu = 0.9868$, $\hat{\sigma}/\sigma = 0.9985$ for x_1, $\hat{\mu}/\mu = 1.0232$, $\hat{\sigma}/\sigma = 1.0001$ for x_2) is sufficient. The obtained evidence is $\hat{c}_E = 0.00149$ and is very close to the true value $c_E = 0.0015$.

Table 4. Bayesian updating results of the 2-DOF model application (Efficiency).

Method		N, N_1/N_2	N_{call}	N_{iter}
aBUS		3×10^3	1.5×10^4	–
aBUS-PCK [24]	$k = 40$	5×10^3	320	4
BUS-AK² [25]		–	171	–
Proposed PBOQ	$k = 4$	$1 \times 10^5/1 \times 10^4$	$10 + 44.6 + 11.4 = 66.0$	24.0

Note: The results of aBUS-PCK are taken from Ref. [24] and averaged over 50 independent runs; The results of BUS-AK² are taken from Ref. [25] and are averaged over 10 independent runs.

Table 5. Bayesian updating results of the 2-DOF model application (Accuracy).

Method			x_1		x_2	
		$\hat{c}$	$\hat{\mu}/\mu$	$\hat{\sigma}/\sigma$	$\hat{\mu}/\mu$	$\hat{\sigma}/\sigma$
aBUS		1.0001	1.0758	1.0155	0.9428	1.0139
BUS-AK² [25]		1.0033	0.9996	1.0006	1.0002	0.9991
Proposed PBOQ	$k = 4$	1.0021	0.9868	0.9985	1.0232	1.0001

Note: The results of BUS-AK² can be found in Ref. [25].

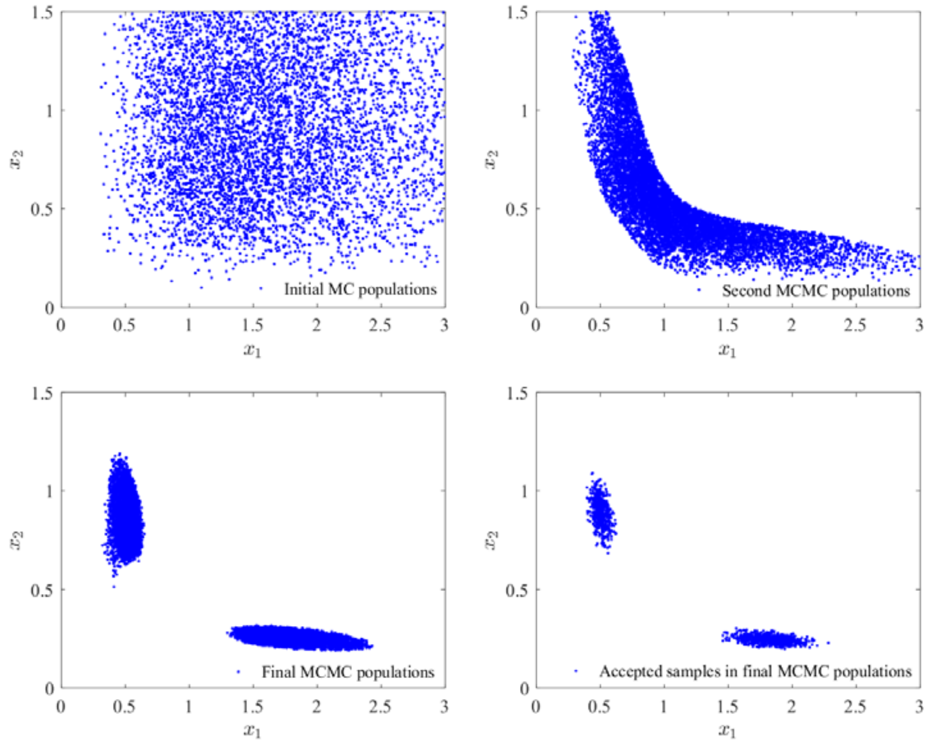

Fig. 6. Realizations of each subset by the proposed method.

Fig. 6 shows realizations of the sample set $\{\boldsymbol{x} = [x_1, x_2]\}$ for each subset in the proposed method. In total, three subsets are utilized until convergence, implying that the failure probability reaches the order of 10^{-3}. As can be observed, the subset gradually approaches the failure domain. The accepted samples in the final subset are also presented in Fig. 7 as the posterior samples, which demonstrate a favorable agreement with the analytical posterior distribution. Furthermore, the observed points are also illustrated in Fig. 7. It can be seen that, these observed points effectively reach the failure domain corresponding to the posterior distribution and are well distributed in both two modes.

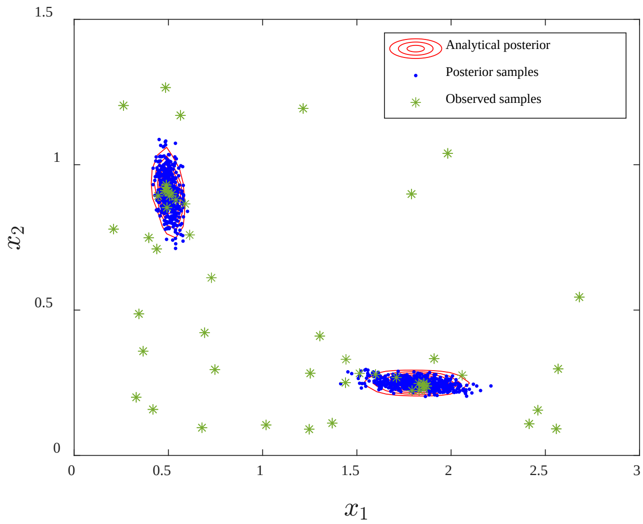

Fig. 7. Posterior samples of $\boldsymbol{x}$.

4.4. Seismic-isolated bridge pier model application

The FE model updating of a seismic-isolated bridge pier is investigated as the fourth example to demonstrate the applicability of the PBOQ method for complex applications. The target bridge shown in Fig. 8 is a five-span seismic-isolated bridge with lead rubber bearings and reinforced concrete (RC) piers, designed based on the specifications for highway bridges in Japan [36]. Structural descriptions of its isolated bridge pier are summarized in Table 6. The isolated bridge pier is numerically modeled as a FE model with 60 DOFs. In this model, the RC pier is represented by four Euler-Bernoulli beam elements and a rotational spring at the bottom. The shear force acting from the superstructure is taken into account as a lumped mass and is connected to the top of the pier using a horizontal spring of the rubber bearings. The footing is characterized as a lumped mass and is connected to the bottom of the pier by a beam element. The boundary condition is modeled by a pair of the sway and rocking springs to consider the soil-structure interaction effect.

Table 6. Descriptions of the seismic-isolated bridge pier.

	Structural parameter	Nominal value
Superstructure	Mass M_S (kg)	6040000
Rubber bearings	Yield stiffness K_B (N/m)	40000000
RC pier	Young's modulus of the concrete (N/m^2)	21000000
	Density of the concrete (N/m^3)	2400
	Cross-sectional area of the upper part (m^2)	26.4
	Cross-sectional area of the lower part (m^2)	11
	Flexural rigidity of the upper part EI_1 (Nm^2)	10000000000
	Flexural rigidity of the lower part EI_2 (Nm^2)	30614000000
	Yield bending moment (Nm)	34840960
	Yield rotational angle (m^2/s)	0.00309
Footing	Mass M_F (kg)	227500
	Moment of inertia I_F (kgm^2)	876800
	Sway spring stiffness (N/m)	1396500000
	Rocking spring stiffness (Ns/m)	17248000000

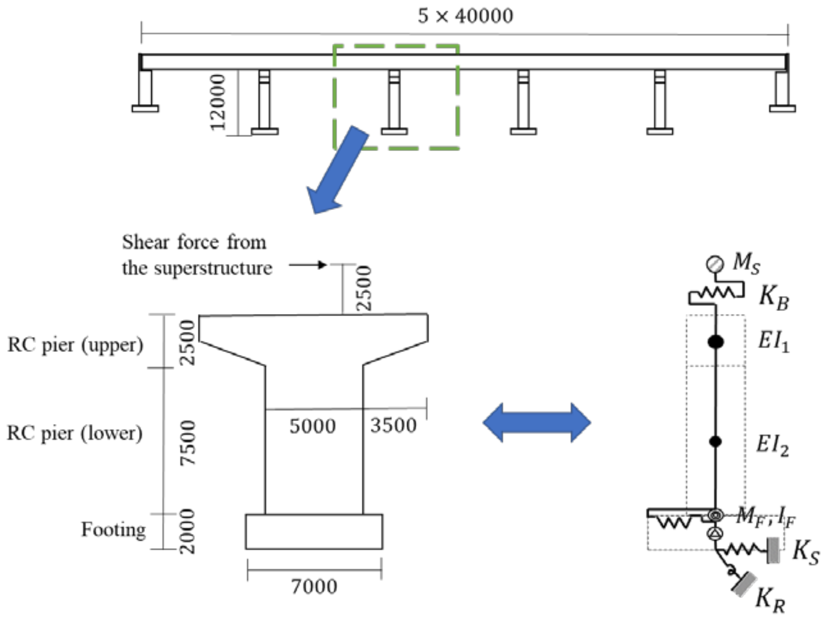

Fig. 8. Target seismic-isolated bridge (unit: mm) and its numerical modeling.

Among various structural parameters that are summarized in Table 6, three key parameters, i.e., the yield stiffness of the rubber bearings and the flexural rigidity of the upper and lower parts of the RC pier, are accounted for as random variables. The other parameters are considered to be constants. According to Adachi [37], these material properties can vary due to manufacturing tolerance, and the mean corresponds to their nominal values and the coefficient of variation (COV) is 0.07 for all of them. Meanwhile, these material properties can also vary due to aging deterioration. In particular, the yield stiffness of the rubber bearings is known to increase around 120 % due to the hardening of the rubbers [38]. Seismic-isolated bridges are typically designed to dissipate the earthquake energy by the rubber bearings so as to reduce the seismic force on the RC piers where the plastic deformation is undesirable. Hence, the change in the yield stiffness of the rubber bearings strongly affect the seismic capacity of the entire bridge system, and it is essential to precisely assess its current value by means of structural health monitoring.

In view of this, we consider the model updating of the target isolated bridge pier by measuring its natural frequencies up to fifth modes. The above three parameters are parameterized as $K_B = \overline{K}_B x_1$, $EI_1 = \overline{EI}_1 x_2$, and $EI_2 = \overline{EI}_2 x_3$, where $\overline{K}_B$, $\overline{EI}_1$, and $\overline{EI}_2$ indicate the nominal values given in Table 6; $\boldsymbol{x} = [x_1, x_2, x_3]$ denotes the inputs to be updated. The prior distribution of $\boldsymbol{x}$ is assumed as an independent normal distribution with the unit means and the standard deviation 0.14. It is noted that, the standard deviation is chosen as a larger value compared to the one to express the manufacturing tolerance (i.e., 0.07) so as to also consider the aging deterioration. Assigning $\boldsymbol{x} = [1.2, 1.0, 1.0]$, the natural frequencies up to fifth modes are obtained by the subspace method as $[f_1, f_2, f_3, f_4, f_5] = [1.03, 3.67, 5.08, 8.29, 8.67]$. Hereby, $x_1 = 1.2$ accounts for the hardening of the bearings. Among them, the three dominant modal frequencies, i.e., f_1, f_2, and f_5, are utilized as the features of interest. The frequency measurements of f_1, f_2, and f_5 are however inevitable to be corrupted with noises. The noises are assumed to follow a normal distribution with zero mean and the standard deviation set as 5 % of the above nominal value of f_1, f_2, and f_5, respectively. In this study, 100 independent realizations of these frequencies are hence generated by adding such noises and employed as the observations.

The likelihood function is modeled to follow a normal distribution and assuming independence between individual observations, it is expressed as follows:

$$L(\boldsymbol{y}|\boldsymbol{x}) = \prod_{j=1}^{100} \frac{1}{\sqrt{2\pi}\sigma_1\sigma_2\sigma_3} \exp\left[-\frac{\left(\hat{f}_1^{(j)} - f_1(\boldsymbol{x})\right)^2}{2\sigma_1^2} - \frac{\left(\hat{f}_2^{(j)} - f_2(\boldsymbol{x})\right)^2}{2\sigma_2^2} - \frac{\left(\hat{f}_3^{(j)} - f_3(\boldsymbol{x})\right)^2}{2\sigma_3^2}\right] \quad (44)$$

where σ_i indicates the standard deviation of the noise on the ith modal frequency; $\hat{f}_i^{(j)}$ means the jth realization of the ith modal frequency; $f_i(\boldsymbol{x})$ is the ith modal frequency obtained by assigning $\boldsymbol{x}$.

Table 7 shows the results of aBUS and the proposed PBOQ. PBOQ are performed 20 independent runs and the results are averaged. It can be seen that both aBUS and PBOQ enable to provide accurate posterior estimates that well correspond to the assigned target values $\boldsymbol{x} = [1.2,1.0,1.0]$. In particular, both methods successfully assess the change in the yield stiffness of the rubber bearings. While aBUS requires a substantially large number of model evaluations for convergence, the proposed PBOQ only requires 36.4 model evaluations on average. Finally, Fig. 9 shows the posterior distribution estimated by the proposed method as a histogram by 1000 realizations of $\boldsymbol{x}$. The horizontal axes correspond to the 99.7 % confidence interval of the prior distribution. As can be observed, the posterior distribution is well sharped and converged around the true target values.

Table 7. Bayesian updating results of the seismic-isolated bridge pier model application.

Method				x_1	x_2	x_3
	N, N_1/N_2	N_{call}	N_{iter}	μ	μ	μ
aBUS	3×10^3	1.5×10^4	–	1.1999	1.0270	1.0339
Proposed PBOQ	$1 \times 10^5/1 \times 10^4$	$10 + 16.4 + 10.0 = 36.4$	7.6	1.2019	1.0187	1.0290

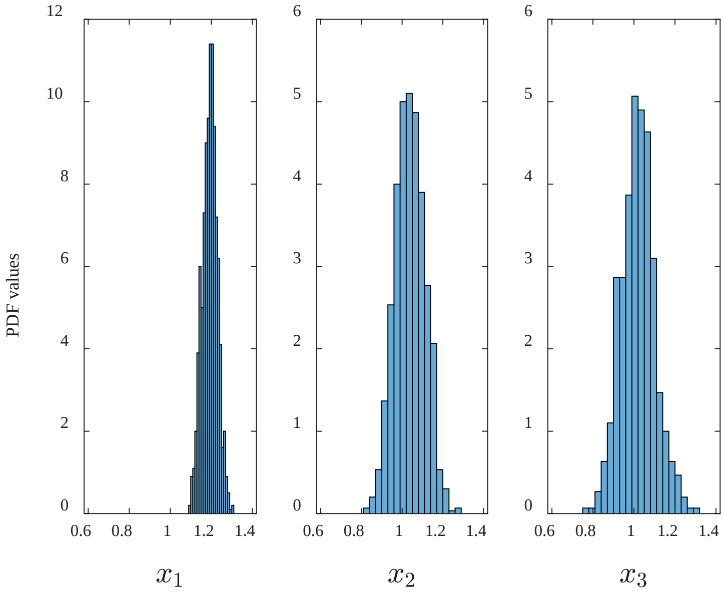

Fig. 9. Posterior distribution of $\boldsymbol{x}$.

5. conclusions

This paper presented a novel Bayesian updating method, termed parallel Bayesian optimization and quadrature (PBOQ), to provide a coherent and efficient approach for the BUS analysis. The BUS analysis comprises two different tasks. The first task is the optimization of the likelihood function to find the constant c that is employed to define a rare event estimation problem. The second task is the quadrature of the probability of the rare event (i.e., failure probability) that aims to infer the posterior distribution by means of the samples conditional on the failure domain. The PBOQ method offers a coherent framework to quantify, propagate, and reduce the numerical uncertainty in these two tasks in a Bayesian fashion by placing GP priors. This results in a significant reduction of the computational burden of model updating (i.e., the number of model evaluations).

Compared to other state-of-the-art methods within the BUS framework relying on GP modeling, the PBOQ method has several unique features. First, PBOQ can take advantage of parallel computing thanks to two new parallel infill sampling criteria, called EI-weighted k-means clustering and UPVC-weighted k-means clustering. These parallel infill sampling strategy can substantially reduce the total number of iteration compared to the conventional purely sequential infill sampling strategy. Second, the UPVC learning function enables the direct measuring of the uncertainty on the failure probability estimation; hence, it is more effective to suggest informative points than the conventional U learning function. Furthermore, PBOQ can properly address extremely small failure probabilities thanks to a new numerical integrator by subset simulation to estimate the posterior failure probability.

The performance of the proposed method is illustrated by means of four numerical applications with increasing complexity, involving a one-dimensional analytical problem, a unimodal distribution problem with 2- and 10-dimensional inputs, a 2-DOF dynamic problem, and the FE model updating of a seismic-isolated bridge pier. Compared to several existing methods, the proposed method shows improved performance for the BUS analysis in regards of efficiency. Nevertheless, the PBOQ method, in its current form, is only applicable to problems with up to medium-dimensional random variables, due to the limitation of GP modeling. For very high-dimensional applications, further research efforts are still needed in the future.

CRediT authorship contribution statement

Masaru Kitahara: Conceptualization, Methodology, Software, Validation, Visualization, Writing – original draft. **Chao Dang:** Conceptualization, Methodology, Software, Writing – review & editing. **Michael Beer:** Supervision, Funding acquisition.

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

The first author acknowledges the support of the Deutsche Forschungsgemensschaft (DFG, German Research Foundation) — SFB1463-434502799. The second author is mainly supported by China Scholarship Council (CSC). The second and third authors would also like to appreciate the support of Sino-German Mobility Program under grant number M-0175.

References

1. J.L. Beck, L.S. Katafygiotis, Updating models and their uncertainties. I: Bayesian statistical framework, J. Eng. Mech. 124 (1998) 455-461.
2. M.C. Kennedy, A. O'Hagan, Bayesian calibration of computer models, J. R. Stat. Soc. B (Statistical Methodology) 63(3) (2001) 425-464.
3. S. Brooks, Markov chain Monte Carlo method and its application, J. R. Stat. Soc. D (the Statistician) 47(1) (1998) 69-100.
4. J.L. Beck, S.K. Au, Bayesian updating of structural models and reliability using Markov chain Monte Carlo simulation, J. Eng. Mech. 128 (2002) 380-391.
5. S.H. Cheung, J.L. Beck, Baysian model updating using hybrid Monte Carlo simulation with application to structural dynamic models with many uncertain parameters, J. Eng. Mech. 135 (2009) 243-255.
6. J. Ching, Y.C. Chen, Transitional Markov chain Monte Carlo method for Bayesian updating, model class selection, and model averaging, J. Eng. Mech. 133 (2007) 816-832.
7. A. Lye, M. Kitahara, M. Broggi, E. Patelli, Robust optimization of a dynamic Black-box system under severe uncertainty: A distribution-free framework, Mech. Syst. Signal Process. 167 (2022) 108522.
8. G.A. Ortiz, D.A. Alvarez, D. Bedoya-Ruíz, Identification of Bouc–Wen type models using the Transitional Markov Chain Monte Carlo method, Compt. Struct. 146 (2015) 252-269.

9. E. Patelli, Y. Govers, M. Broggi, H.M. Gomes, M. Link, J.E. Mottershead, Sensitivity or Bayesian model updating: a comparison of techniques using the DLR AIRMOD test data, Arch. Appl. Mech. 87 (2017) 905-925.
10. D. Straub, I. Papaioannou, Bayesian updating with structural reliability methods, J. Eng. Mech. 141 (2015) 04014134.
11. S.K. Au, J.L. Beck, Estimation of small failure probabilities in high dimensions by subset simulation, Probabilist. Eng. Mech. 16 (2001) 263-277.
12. S.K. Au, E. Patelli, Rare event simulation in finite-infinite dimensional space, Reliab. Eng. Syst. Safety 148 (2016) 67-77.
13. W. Betz, I. Papaioannou, J.L. Beck, D. Straub, Bayesian inference with subset simulation: Strategies and improvements, Compt. Meth. Appl. Mech. Eng. 331 (2018) 72-93.
14. D. G. Giovanis, I. Papaioannou, D. Straub, V. Papadopoulos, Bayesian updating with subset simulation using artificial neural networks, Compt. Meth. Appl. Mech. Eng. 319 (2017) 124-145.
15. C.G. Bucher, U. Bourgund, A fast and efficient response surface approach for structural reliability problems, Struct. Safety 7 (1990) 57-66.
16. G. Blatman, B. Sudret, An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis, Probabilist. Eng. Mech. 25 (2010) 183-197.
17. B. Bhattacharyya, Structural reliability analysis by a Bayesian sparse polynomial chaos expansion, Struct. Safety 90 (2021) 102074.
18. J.E. Hurtado, A.A. Diego, Neural-network-based reliability analysis: a comparative study, Compt. Methods Appl. Mech. Eng. 191 (2001) 113-132.
19. I. Kaymaz, Application of Kriging method to structural reliability problems, Struct. Safety 27 (2005) 133-151.
20. B. Echard, N. Gayton, M. Lemaire, AK-MCS: An active learning reliability method combining Kriging and Monte Carlo Simulation, Struct. Safety 32 (2011) 145-154.
21. Z. Wang, A. Shafieezadeh, Highly efficient Bayesian updating using metamodels: An adaptive Kriging-based approach, Struct. Safety 84 (2020) 101915.
22. M. Kitahara, S. Bi, M. Broggi, M. Beer, Bayesian model updating in time domain with metamodel-based reliability method, J. Risk Uncertainty Eng. Syst. Part A: Civ. Eng. 7 (2021) 04021030.
23. C. Song, Z. Wang, A. Shafieezadeh, R. Xiao, BUAK-AIS: Efficient Bayesian Updating with Active learning Kriging-based Adaptive Importance Sampling, Compt. Methods Appl. Mech. Eng. 391 (2022) 114578.
24. D. Rossat, J. Baroth, M. Briffaut, F. Dufour, Bayesian inversion using adaptive Polynomial Chaos Kriging within Subset Simulation, J. Compt. Phys. 455 (2022) 110986.
25. Y. Liu, L. Li, S. Zhao, Efficient Bayesian updating with two-step adaptive Kriging, Struct. Safety 95 (2022) 102172.
26. D.R. Jones, M. Schonlau, W.J.. Welch, Efficient global optimization of expensive black-box functions, J. Golobal Optim. 13(4) (1998) 455-492.
27. A. O'Hagan, Bayes-Hermite quadrature, J. Statist. Plann. Inference 29(3) (1991) 245-260.
28. C. Dang, P. Wei, M. Faes, M.A. Valdebenito, M. Beer, Parallel adaptive Bayesian quadrature for rare event estimation, Reliab. Eng. Syst. Safety 225 (2022) 108621.

29. C. Dang, M.A. Valdebenito, M. Faes, P. Wei, M. Beer, Structural reliability analysis: A Bayesian perspective, Struct. Safety 99 (2022) 102259.
30. J.L. Beck, K.V. Yuen, Model selection using response measurements: Bayesian probabilistic approach, J. Eng. Mech. 130 (2004) 192-203.
31. E. Rasmussen, Gaussian processes in machine learning, Summer school on machine learning, Springer, Berlin, 2003.
32. C. Dang, P. Wei, M. Faes, M.A. Valdebenito, M. Beer, Interval uncertainty propagation by a parallel Bayesian global optimization method, Appl. Math. Model. 108 (2022) 220-235.
33. J. MacQueen, Some methods for classification and analysis of multivariate observations, Univ. California Press, Berkeley, 1967.
34. W. Jian, S. Yhili, Y. Qiang, L. Rui, Two accuracy measures of the Kriging model for structural reliability analysis, Reliab. Eng. Syst. Safety 167 (2017) 494-505.
35. Z. Wang, A. Shafiezadeh, On confidence intervals for failure probability estimates in Kriging-based reliability analysis, Reliab. Eng. Syst. Safety 196 (2020) 106758.
36. Japan Road Association, Design specifications of highway bridges V: Seismic design, Maruzen, Tokyo, 2016.
37. Y. Adachi, Reliability analysis and limit state design method of isolated bridges under extreme ground motions, Doctoral thesis, Dept. of Civil Engineering, Kyoto Univ. Kyoto, 2002. (In Japanese)
38. K. Hayashi, Y. Adachi, A. Igarashi, J. Dang,T. Higashide, Experimental evaluation of aging deterioration of rubber bearings in highway bridges, Proc. 2nd European Conf. Earthq, Eng. Seismology, Istanbul (2014) 25-29.