BUAK-AIS: Efficient Bayesian Updating with Active Learning Kriging-based Adaptive Importance Sampling

Chaolin Song[1,2], Zeyu Wang[1,3], Abdollah Shafieezadeh[1]*, Rucheng Xiao[2]
[1]Risk Assessment and Management of Structural and Infrastructure Systems (RAMSIS) Lab, Department of Civil, Environmental, and Geodetic Engineering, The Ohio State University, Columbus, OH, 43210, United States
[2]Department of Bridge Engineering, Tongji University, Shanghai, 200092, China
[3]Department of Civil Engineering, Tsinghua University, Beijing, 100084, China

ABSTRACT

Bayesian updating provides a sound mathematical framework for probabilistic calibration as new information emerges. Bayesian Updating with Structural reliability methods (BUS) reformulates the acceptance domain in rejection sampling as a failure domain in reliability analysis, offering considerable potential for higher efficiency and accuracy. Kriging-based Monte Carlo Simulation has been studied to facilitate the application of BUS for problems with expensive-to-evaluate likelihood functions. Nevertheless, as the implementation of BUS often involves a rare event, the number of required Monte Carlo samples can become unaffordable. This gap is addressed here through Bayesian Updating with Active learning Kriging-based Adaptive Importance Sampling (BUAK-AIS). An importance sampling density based on Gaussian mixture distribution is introduced, and the discrepancy between the adopted and theoretically best sampling densities is measured through the Kullback–Leibler cross entropy. The proposed method includes an active learning framework that adaptively extends the training set and optimizes the parameters of the Gaussian mixture distribution based on the cross entropy and the current Kriging model. As BUS uses accepted samples to estimate the posterior distribution, the present work discusses the estimate for the first moment of the posterior distribution, and proposes a criterion to check the sufficiency of the number of accepted samples to guarantee robust estimations. A new stopping criterion is also developed by quantifying the error introduced by Kriging. Three numerical examples and an engineering application concerning model updating of cable-stayed bridges in the construction process are investigated, demonstrating the efficiency and accuracy of the proposed method.

Key words: *Bayesian updating, Calibration, Reliability Analysis, Importance sampling, Active learning Kriging*

Nomenclature

c	A constant required to formulate the acceptance domain in rejection sampling
d	The dimension of the variable space
$\mathbb{E}_{f_{IS}}(\cdot)$	The expectation with $(\boldsymbol{X}, P)$ following f_{IS}
$f'(\cdot)$	The prior joint PDF of $\boldsymbol{X}$
$f''(\cdot)$	The posterior joint PDF of $\boldsymbol{X}$
$f_{IS}(\cdot)$	The importance sampling distribution for $(\boldsymbol{X}, P)$
$f_{\boldsymbol{\varepsilon}}(\cdot)$	The joint PDF of $\boldsymbol{\varepsilon}$
$g(\cdot)$	The limit state function
$g_K(\cdot)$	The limit state function predicted by Kriging
$I_{g\le 0}(\cdot)$	The indicator function that equals one when the realization satisfies $g \le 0$ and zero otherwise
$I_w(\cdot)$	The indicator of the wrong classification event $I_w(\boldsymbol{x}, p) = \left\vert I_{g\le 0}(\boldsymbol{x}, p) - I_{g_K\le 0}(\boldsymbol{x}, p)\right\vert$
K	The number of Gaussian distributions for the quasi-optimal Gaussian mixture distribution
$L(\cdot)$	The likelihood function
m	The number of independent observations
N_{Call}	The number of calls of the performance function
N_{CE}	The number of samples to update the IS distribution
N_F	The number of low-discrepancy samples in the design space
N_I	The number of initial training samples for Kriging
N_{IS}	The number of IS samples
N_{MCS}	The number of MCS samples
N_{SS}	The number of samples in each subset for Subset Simulation
N_T	The number of training points for Kriging
P	The augmented variable introduced by rejection sampling
p	A realization of the variable P
$\boldsymbol{X}$	The random vector representing the system uncertainty
$\boldsymbol{x}$	A realization of the random vector $\boldsymbol{X}$
x^i	$\boldsymbol{x} = [x^1, \cdots, x^d]$, the superscript i denotes the dimension number
$\boldsymbol{Y}$	The vector representing the observation
$\boldsymbol{y}$	A realization of the observation $\boldsymbol{Y}$
$\boldsymbol{\varepsilon}$	The deviation of the observation
Ω	The probabilistic domain of $\boldsymbol{X}$
Ω_{acc}	The acceptance domain in rejection sampling
Ω_f	The failure domain in BUS
$\boldsymbol{\mu}_{f''}$	The first moment of the posterior distribution
$\boldsymbol{\mu}_{f'',K}$	The first moment of the posterior distribution estimated with Kriging
$\hat{\boldsymbol{\mu}}_{f''}$	The estimation of the first moment of the posterior distribution with realizations
$\boldsymbol{\Sigma}_{f''}$	The second moment of the posterior distribution
$\hat{\boldsymbol{\Sigma}}_{f''}$	The estimation of the second moment of the posterior distribution with realizations
$\mathcal{A}^i$	Denoting $x^i \frac{f'(\boldsymbol{x})}{f_{IS}(\boldsymbol{x},p)} I_{g\le 0}(\boldsymbol{x}, p)$
$\mathcal{B}$	Denoting $\frac{f'(\boldsymbol{x})}{f_{IS}(\boldsymbol{x},p)} I_{g\le 0}(\boldsymbol{x}, p)$
$\mathcal{J}_k$	Denoting $I_w(\boldsymbol{x}_k, p_k) \frac{f'(\boldsymbol{x}_k)}{f_{IS}(\boldsymbol{x}_k,p_k)}$
$\mathcal{P}_k^i$	Denoting $x_k^i I_w(\boldsymbol{x}_k, p_k) \frac{f'(\boldsymbol{x}_k)}{f_{IS}(\boldsymbol{x}_k,p_k)}$
$\Xi_{\mathcal{J}}$	The sum of $\mathcal{J}_k$ for the IS realizations, i.e., $\sum_{k=1}^{N_{IS}} \mathcal{J}_k$, also denoted as D_d

$\Xi_{\mathcal{P}^i}$	The sum of $\mathcal{P}_k^i$ for the IS realizations, i.e., $\sum_{k=1}^{N_{IS}} \mathcal{P}_k^i$, also denoted as D_n^i
$\boldsymbol{\epsilon}_{\boldsymbol{\mu}_{f''}}$	The vector that equals $\left[COV_{\hat{\mu}_{\mathcal{A}^1}}, \cdots, COV_{\hat{\mu}_{\mathcal{A}^d}}, COV_{\hat{\mu}_{\mathcal{B}}}\right]$
$\boldsymbol{\epsilon}_{\boldsymbol{D}_n}$	The vector representing the maximum relative error for estimating $\mathbb{E}_{f_{IS}}\left[\boldsymbol{x}\frac{f'(\boldsymbol{x}_k)}{f_{IS}(\boldsymbol{x},p)} I_{g\leq 0}(\boldsymbol{x}, p)\right]$ with the Kriging surrogate model
ϵ_{D_d}	The maximum relative error for estimating $\mathbb{E}_{f_{IS}}\left[\frac{f'(\boldsymbol{x}_k)}{f_{IS}(\boldsymbol{x},p)} I_{g\leq 0}(\boldsymbol{x}, p)\right]$ with the Kriging surrogate model

1. Introduction

In the design and management of structures and infrastructure systems, a multitude of uncertainties (e.g., various environmental or load conditions, workmanship, human error and occurrence of future events [1-3]) exist that must be characterized and considered when designing or maintaining systems. Collecting new information for model calibration and system identification can reduce these uncertainties and facilitate effective decisions [4-7]. Bayesian updating provides a coherent framework for probabilistic calibration, where a posterior joint probability density function (PDF) is derived using the prior joint PDF and the observed data. With the advancements in sensing and monitoring technologies, Bayesian updating has recently received much attention and been successfully applied in various fields [8-15].

The Kalman filter [16] is one of the most popular algorithms for Bayesian updating. As the Kalman filter adopts the linear Gaussian assumption, only the first and second moments are required to describe a distribution. However, this assumption is not applicable for nonlinear and non-Gaussian problems. Many studies [17-22] have attempted relaxing these assumptions to extend the Kalman filter. For instance, Extended Kalman Filter, the Assumed Density Filter and the Unscented Kalman Filter consist of linearizing the state-space using Taylor series expansion, moments matching and weighted statistical linear regression process, respectively [23]. On the other hand, approximating posterior distribution via sampling realizations has received much attention, as a parametric assumption (e.g., the Gaussian assumption) is not required a priori. The rejection sampling [24] is a basic technique to generate samples from the posterior distribution considering the likelihood function as a filter. However, the acceptance rate of the simple rejection sampling can be very low in high-dimensional problems or with multiple observations, resulting in low sampling efficiency. Markov chain Monte Carlo (MCMC), which allows sampling directly from the posterior distribution, has been widely investigated for Bayesian updating. To avoid the convergence issue, Beck and Au [25] proposed an adaptive Metropolis–Hastings method (AMH), which utilized a series of intermediate PDF to approach the target posterior PDF. Subsequently, Ching and Chen [26] proposed a TMCMC method, which adopted a resampling technique instead of the kernel density in AMH to improve the efficiency. However, the gained efficiency may not be significant with the increase of the dimension of the variable space [27].

As an alternative, Straub and Papaioannou [28] proposed performing Bayesian Updating with Structural reliability methods (BUS), by transforming the acceptance domain in rejection sampling as the failure domain in reliability analysis. The significant advantage of BUS lies in exploring the possibility of using reliability methods to solve the Bayesian updating problem. The performance of BUS highly depends on the adopted reliability method. The subset simulation, which calculates the failure probability with MCMC through a series of intermediate events, has been applied to BUS in [28]. On the other hand, surrogate model-based reliability analysis, which uses the surrogate model as a substitute of original performance function has become popular in recent years due to its high efficiency. Compared with traditional sampling methods, the computational cost can be reduced by several orders of magnitude, as noticed in reliability analysis with Artificial Neural Networks (ANN) [27], Polynomial Chaos Expansion [29, 30], and Support Vector Regression [31, 32]. Among these, Kriging-based adaptive reliability analysis [33] has become one of the most popular approaches, as Kriging can provide the uncertainty of prediction to guide the selection of the training point and quantify the accuracy of the surrogate model. Wang and Shafieezadeh [34] proposed implementing adaptive Kriging with Monte Carlo Simulation (MCS) in the framework of BUS, called BUAK, which shows high efficiency, especially with expensive-to-evaluate likelihood functions. However, as the implementation of BUS often involves a rare event estimation, the BUAK-MCS may not be computationally affordable in some applications due to the low convergence rate of MCS.

To address this gap, this paper proposes an efficient Bayesian updating method with active learning Kriging-based adaptive importance sampling, called BUAK-AIS. The main contributions of the paper can be summarized as follows. First, this work proposes an active learning Kriging-based adaptive importance sampling framework for efficient Bayesian updating. In the framework, the training database is expanded in each iteration and the importance sampling distribution is optimized sequentially based on the cross entropy. Moreover, the estimate for the first moment of the posterior distribution is discussed. A criterion is proposed and checked in the framework to guarantee that the realizations in the accepted domain can provide a robust estimate for the posterior distribution. A new stopping criterion for the active learning of Kriging is also proposed by quantifying the error caused by the Kriging surrogate model in the estimation of the posterior distribution. Three numerical examples and an engineering application regarding model updating of cable-stayed bridges in the construction process are selected to illustrate the accuracy and efficiency of the proposed methods.

The rest of the paper is organized as follows. Section 2 presents an overview of the fundamental theory of Bayesian updating, BUS and Kriging. Section 3 presents the details of the proposed method, BUAK-AIS. Section 4 provides three numerical examples and an engineering application to illustrate the performance of the proposed method. Concluding remarks are presented in Section 5.

2. Preliminaries

This section provides a brief overview of Bayesian updating, Bayesian Updating with Structural reliability methods (BUS), and Kriging. More information about these techniques can be found in [28, 34], [28], and [35], respectively.

2.1 Bayesian updating

Uncertainties stemming from external environments and internal conditions may pose significant challenge to design and maintaining the functionality and integrity of systems in their service life. Following the classical reliability framework, let the random vector $\boldsymbol{X}$ represent the uncertainty of the system associated with the phenomena of interest. Given the probabilistic model of $\boldsymbol{X}$ and the formulation of the system performance function, risk and reliability analysis can help with the assessment of the system state and facilitate risk-informed decision making. However, the background knowledge, where the prior probabilistic model of $\boldsymbol{X}$ is established, may not perfectly describe the actual system. This deviation can yield inaccurate assessments and predictions, therefore resulting in decisions that are not most cost-effective and even threatening the safety of the system. Thus, collecting new information through Structural Health Monitoring (SHM) becomes essential in the management of critical structures and infrastructure systems, as the observed data can be used for system identification and probabilistic calibration. For this purpose, Bayesian updating provides a sound mathematical framework.

Let $f'(\boldsymbol{x})$ and $\boldsymbol{Y}$ denote the prior PDF of $\boldsymbol{X}$ and the vector of observations, respectively. $L(\boldsymbol{x})$ denotes the likelihood function, which is proportional to the conditional probability of receiving a realization of $\boldsymbol{Y}$ given a realization of $\boldsymbol{X}$, i.e., $L(\boldsymbol{x}) \propto \Pr(\boldsymbol{Y} = \boldsymbol{y} | \boldsymbol{X} = \boldsymbol{x})$. The Bayesian updating can be formulated as:

$$f''(\boldsymbol{x}) = \frac{L(\boldsymbol{x}) f'(\boldsymbol{x})}{\int_{\Omega} L(\boldsymbol{x}) f'(\boldsymbol{x}) d\boldsymbol{x}} \tag{1}$$

where Ω is the probabilistic domain of the random variables $\boldsymbol{X}$, and $f''(\boldsymbol{x})$ is the posterior distribution which is the aim of Bayesian updating.

Let $h(\cdot)$ denote the function describing the responses of observed phenomena. Considering a realization of observations $\boldsymbol{y}$ of equality type and a realization of the random vector $\boldsymbol{x}$, the deviation between the observation and the prediction can be expressed as $\boldsymbol{y} - h(\boldsymbol{x})$. The deviation can be caused by measurement errors or model errors. Let $\boldsymbol{\varepsilon}$ denote the deviation of the observation, and $f_{\boldsymbol{\varepsilon}}$ denote the joint PDF of $\boldsymbol{\varepsilon}$. The likelihood function can be formulated as:

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

With m mutually independent observations ($\boldsymbol{y} = [y_1, \cdots, y_m]$), the likelihood function can be decomposed as:

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

where L_i denotes the likelihood function of receiving the observation y_i, $h_i(\cdot)$ denotes the function describing the prediction of the observation y_i, ε_i denotes the deviation of y_i, and f_{ε_i} denotes the PDF of ε_i.

2.2 Bayesian updating with structural reliability methods (BUS)

The rejection sampling is a basic sampling method to estimate the posterior distribution. The idea of rejection sampling consists of extending the space of random variables to $[\boldsymbol{X}, P]$ by introducing an augmented variable P, and defining the accepted domain as:

$$\Omega_{acc} = [p \leq cL(\boldsymbol{x})] \tag{4}$$

where p is a realization of P and c is a constant. The distribution of P can be defined as a standard uniform distribution, and c can be therefore defined as $1/\max(L(\boldsymbol{x}))$ to satisfy $cL(\boldsymbol{x}) \leq 1$. An adaptive approach for determining c is discussed in [36, 37]. DiazDelaO et al. [8] also presented a fundamental discussion on the role of c and developed an adaptive BUS-based formulation to approach the posterior distribution without c a prior. When sampling P and $\boldsymbol{X}$ from the standard uniform distribution and prior distribution, respectively, the PDF for the realizations of $\boldsymbol{X}$ in the acceptance domain can approach the posterior distribution. This can be guaranteed by:

$$\frac{\int_{P\in\Omega_{acc}} f'(\boldsymbol{x})\,\mathrm{d}p}{\int_{[\boldsymbol{X},P]\in\Omega_{acc}} f'(\boldsymbol{x})\,\mathrm{d}p\mathrm{d}\boldsymbol{x}} = \frac{\int_0^{cL(x)} f'(\boldsymbol{x})\,\mathrm{d}p}{\int_\Omega \int_0^{cL(x)} f'(\boldsymbol{x})\,\mathrm{d}p\mathrm{d}\boldsymbol{x}} = \frac{cL(\boldsymbol{x})f'(\boldsymbol{x})}{\int_\Omega cL(\boldsymbol{x})f'(\boldsymbol{x})d\boldsymbol{x}} = f''(\boldsymbol{x}) \tag{5}$$

However, as mentioned in [28], the acceptance rate for the simple rejection sampling can be extremely low. With m independent and identically distributed measurements, the average acceptance ratio is proportional to $1/\sqrt{m}$. This limits the application of rejection sampling for complex engineering problems. To address this limitation, the Bayesian updating with structural reliability methods (BUS) is proposed by Straub and Papaioannou [28]. The core idea is to reformulate the rejection sampling as a reliability problem. The acceptance domain Ω_{acc} can also be considered as the failure domain in reliability analysis, as shown by Eq. (6).

$$\Omega_f = [g(\boldsymbol{x}, p) \leq 0] \tag{6}$$

where $g(\boldsymbol{x}, p) = p - cL(\boldsymbol{x})$, which is known as the limit state function in reliability analysis. Thus, existing reliability analysis methods, e.g., MCS and subset simulation, can be implemented in the framework of BUS to solve this reliability problem. However, unlike reliability analysis which focuses on the probability of failure, BUS concentrates on estimating the posterior distribution with the samples in the failure domain. The estimation of the Cumulative Distribution Function (CDF), and the first and second moments of the posterior distribution with MCS-based BUS can be formulated as follows:

$$\begin{aligned} CDF(\boldsymbol{\psi}) &= \int_\Omega I(\boldsymbol{x} \leq \boldsymbol{\psi}) f''(\boldsymbol{x}) d\boldsymbol{x} \\ &= \frac{\int_{[\boldsymbol{X},P]\in\Omega_f} I(\boldsymbol{x} \leq \boldsymbol{\psi}) f'(\boldsymbol{x}) dp d\boldsymbol{x}}{\int_{[\boldsymbol{X},P]\in\Omega_f} f'(\boldsymbol{x}) dp d\boldsymbol{x}} \approx \frac{\sum_{k=1}^{N_{MCS}} I(\boldsymbol{x}_k \leq \boldsymbol{\psi}) \cdot I_{g\leq 0}(\boldsymbol{x}_k, p_k)}{\sum_{k=1}^{N_{MCS}} I_{g\leq 0}(\boldsymbol{x}_k, p_k)} \end{aligned} \tag{7}$$

$$\boldsymbol{\mu}_{f''} = \int_\Omega \boldsymbol{x} f''(\boldsymbol{x}) d\boldsymbol{x} = \frac{\int_{[\boldsymbol{X},P]\in\Omega_f} \boldsymbol{x} f'(\boldsymbol{x}) dp d\boldsymbol{x}}{\int_{[\boldsymbol{X},P]\in\Omega_f} f'(\boldsymbol{x}) dp d\boldsymbol{x}} \approx \frac{\sum_{k=1}^{N_{MCS}} \boldsymbol{x}_k \cdot I_{g\leq 0}(\boldsymbol{x}_k, p_k)}{\sum_{k=1}^{N_{MCS}} I_{g\leq 0}(\boldsymbol{x}_k, p_k)} \tag{8}$$

$$\boldsymbol{\Sigma}_{f''} = \int_\Omega (\boldsymbol{x} - \boldsymbol{\mu}_{f''})^{\mathrm{T}} (\boldsymbol{x} - \boldsymbol{\mu}_{f''}) f''(\boldsymbol{x}) d\boldsymbol{x} = \frac{\int_{[\boldsymbol{X},P]\in\Omega_f} (\boldsymbol{x} - \boldsymbol{\mu}_{f''})^{\mathrm{T}} (\boldsymbol{x} - \boldsymbol{\mu}_{f''}) f'(\boldsymbol{x}) dp d\boldsymbol{x}}{\int_{[\boldsymbol{X},P]\in\Omega_f} f'(\boldsymbol{x}) dp d\boldsymbol{x}} \tag{9}$$

$$\approx \frac{\sum_{k=1}^{N_{MCS}} \left(\boldsymbol{x}_k - \widehat{\boldsymbol{\mu}}_{f''}\right)^{\mathrm{T}} \left(\boldsymbol{x}_k - \widehat{\boldsymbol{\mu}}_{f''}\right) \cdot I_{g\le 0}(\boldsymbol{x}_k, p_k)}{\sum_{k=1}^{N_{MCS}} I_{g\le 0}(\boldsymbol{x}_k, p_k)}$$

where N_{MCS} is the number of MCS samples, $I_{g\le 0}(\cdot)$ is an indicator function that equals one when the realization satisfies $g \le 0$ and zero otherwise, and $I(\boldsymbol{x} \le \boldsymbol{\psi})$ is also an indicator function that equals one when the realization satisfies $\boldsymbol{x} \le \boldsymbol{\psi}$ and zero otherwise.

2.3 Kriging

As an exact interpolation method that can provide the uncertainty information based on the Gaussian process assumption, Kriging has received much attention recently. A Kriging surrogate model assumes that the true prediction of a point consists of a regression part and a random part:

$$\hat{g}(\boldsymbol{x}) = S(\boldsymbol{x}, \boldsymbol{\beta}) + z(\boldsymbol{x}) \tag{10}$$

where $\hat{g}(\boldsymbol{x})$ is the prediction of the performance, $S(\boldsymbol{x}, \boldsymbol{\beta}) = \boldsymbol{s}(\boldsymbol{x})^T \boldsymbol{\beta}$ is the regression part and $z(\boldsymbol{x})$ is the zero-mean Gaussian process. $\boldsymbol{s}(\boldsymbol{x})$ is identical to $\{s_1(\boldsymbol{x}), \cdots, s_{N_S}(\boldsymbol{x})\}^T$, and $\boldsymbol{\beta} = \{\beta_1, \cdots, \beta_{N_S}\}^T$ is the set of regression parameters. For instance, when the trend has a constant yet unknown value, the Kriging model is known as the Ordinary Kriging, and in this case, both N_S and $s_1(\boldsymbol{x})$ are identical to 1.

Subsequently, with N_T training points $\boldsymbol{\mathcal{T}} = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{N_T}\}^{\mathrm{T}}$ and their noise-free responses $\boldsymbol{\mathcal{M}} = \{g(\boldsymbol{x}_1), \cdots, g(\boldsymbol{x}_{N_T})\}^T$, the prediction $\hat{g}(\boldsymbol{x}_t)$ of an arbitrary test point $\boldsymbol{x}_t$ and $\boldsymbol{\mathcal{M}}$ are assumed to follow a joint Gaussian distribution:

$$\begin{Bmatrix} \hat{g}(\boldsymbol{x}_t) \\ \boldsymbol{\mathcal{M}} \end{Bmatrix} = \mathcal{N}_{N_T+1}\left(\begin{Bmatrix} S(\boldsymbol{x}_t, \boldsymbol{\beta}) \\ S(\boldsymbol{\mathcal{T}}, \boldsymbol{\beta}) \end{Bmatrix}, \sigma^2 \begin{Bmatrix} 1 & \boldsymbol{r}^{\mathrm{T}}(\boldsymbol{x}_t) \\ \boldsymbol{r}(\boldsymbol{x}_t) & \boldsymbol{R} \end{Bmatrix}\right) \tag{11}$$

where $S(\boldsymbol{x}_t, \boldsymbol{\beta})$ has been explained above, $S(\boldsymbol{\mathcal{T}}, \boldsymbol{\beta}) = \boldsymbol{S\beta}$, and $\boldsymbol{S}$ is a matrix with the element in the row i and column j being $s_j(\boldsymbol{x}_i)$, $i \in [1, \cdots, N_T]$, $j \in [1, \cdots, N_S]$. Thus, the size of matrix $\boldsymbol{S}$ is $N_T \times N_S$. $\boldsymbol{r}(\boldsymbol{x}_t)$ is the vector representing the correlation between $\boldsymbol{x}_t$ and $\boldsymbol{\mathcal{T}}$, $\boldsymbol{R}$ is the covariance matrix with the elements R_{ij} in the row i and column j describing the correlation between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ ($i, j \in [1, \cdots, N_T]$), and σ^2 represents the generalized mean square error (MSE) from the regression.

Linear, spline or Gaussian correlation models, among others, can be applied to define R_{ij} and $\boldsymbol{r}(\boldsymbol{x}_t)$. For instance, with Gaussian correlation model, the R_{ij} and $\boldsymbol{r}(\boldsymbol{x}_t)$ will have the following form:

$$R_{ij} = R(\boldsymbol{x}_i, \boldsymbol{x}_j) = \prod_{k=1}^{d} exp\left[-\theta_k (x_i^k - x_j^k)^2\right] \tag{12}$$

$$\boldsymbol{r}(\boldsymbol{x}_t) = \{R(\boldsymbol{x}_1, \boldsymbol{x}_t), \cdots, R(\boldsymbol{x}_{N_T}, \boldsymbol{x}_t)\}^T \tag{13}$$

where d denotes the dimension of the input variable, i.e., $\boldsymbol{x}_k = [x_k^1, \cdots, x_k^d]$ with the subscript k denoting the sample number and the superscript denoting the dimension number, and $\boldsymbol{\theta} = [\theta_1, \cdots, \theta_d]$ is the set of correlation parameters.

As pointed out in [38], $\boldsymbol{\theta}$ can significantly influence the performance of Kriging in reliability analysis. Approaches such as Maximum Likelihood Estimation (MLE) and Cross-validation Estimation have been proposed to determine the optimal $\boldsymbol{\theta}$, called $\widehat{\boldsymbol{\theta}}^*$. Note that as $\widehat{\boldsymbol{\theta}}^*$ is determined based on the training database, $\widehat{\boldsymbol{\theta}}^*$ can typically change in the active learning process, and with limited samples, the parameters that provide the most accurate reliability result may not be the ones obtained by optimization [38]. To keep consistent with the existing work [34], MLE is adopted in this paper:

$$\widehat{\boldsymbol{\theta}}^* = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\left(|\boldsymbol{S}|^{\frac{1}{N_T}} \sigma^2\right) \tag{14}$$

where σ^2 can be formulated as:

$$\sigma^2 = \frac{1}{n_t} (\boldsymbol{\mathcal{M}} - \boldsymbol{S\beta}^*)^T \boldsymbol{R}^{-1} (\boldsymbol{\mathcal{M}} - \boldsymbol{S\beta}^*) \tag{15}$$

$\boldsymbol{\beta}^*$ is generalized least-squares estimate of $\boldsymbol{\beta}$:

$$\boldsymbol{\beta}^* = \left(\boldsymbol{S}^T\boldsymbol{R}^{-1}\boldsymbol{S}\right)^{-1}\boldsymbol{S}^T\boldsymbol{R}^{-1}\boldsymbol{\mathcal{M}} \tag{16}$$

Based on the above equations, the mean and variance of the prediction $\hat{g}(\boldsymbol{x}_t)$ conditional on $\boldsymbol{\mathcal{T}}$ and $\boldsymbol{\mathcal{M}}$ can be formulated as:

$$\mu_K(\boldsymbol{x}_t) = \boldsymbol{s}(\boldsymbol{x}_t)^T\boldsymbol{\beta}^* + \boldsymbol{r}(\boldsymbol{x}_t)^T\boldsymbol{R}^{-1}(\boldsymbol{\mathcal{M}} - \boldsymbol{S}\boldsymbol{\beta}^*) \tag{17}$$

$$\sigma_K^2(\boldsymbol{x}_t) = \sigma^2\left(1 + \boldsymbol{u}(\boldsymbol{x}_t)^{\boldsymbol{T}}(\boldsymbol{S}^T\boldsymbol{R}\boldsymbol{S})^{-1}\boldsymbol{u}(\boldsymbol{x}_t) - \boldsymbol{r}(\boldsymbol{x}_t)^T\boldsymbol{R}^{-1}\boldsymbol{r}(\boldsymbol{x}_t)\right) \tag{18}$$

where $\boldsymbol{u}(\boldsymbol{x}_t)$ can be formulated as:

$$\boldsymbol{u}(\boldsymbol{x}_t) = \boldsymbol{S}^T\boldsymbol{R}^{-1}\boldsymbol{r}(\boldsymbol{x}_t) - \boldsymbol{s}(\boldsymbol{x}) \tag{19}$$

3. Bayesian Updating with Active learning Kriging-based Adaptive Importance Sampling (BUAK-AIS)

This section presents the details of the proposed method. Section 3.1 presents the derivation of the BUS with Importance Sampling (IS) for Bayesian updating, and Section 3.2 proposes a criterion to measure whether the number of accepted samples is enough for an accurate posterior estimate. The maximum of the error introduced by the use of the Kriging surrogate model is derived and discussed in Section 3.3. The framework of the proposed method, Bayesian updating with Active learning Kriging-based Adaptive Importance Sampling, called BUAK-AIS, is presented in Section 3.4.

3.1 BUS with Importance Sampling

A classical approach of reliability analysis is MCS, which is based on repeated random sampling to obtain statistical results. The application of MCS to the reliability problem formulated in BUS has been illustrated in Section 2.2. However, in MCS, realizations of $\boldsymbol{X}$ and P are randomly generated from the prior joint PDF $f(\boldsymbol{x})$ and the uniform distribution. For a rare event estimation, only a very small proportion of points falls in the failure domain and contributes to the estimation of the posterior distribution. The importance sampling, which samples from an alternative proposal distribution, can be introduced in the BUS framework to improve the sampling efficiency.

In the following, the statistical characteristics (e.g., Cumulative Distribution Function (CDF) and the first and second moments) of the posterior distribution with importance sampling-based BUS are derived. The rest of the paper uses f_{IS} to represent the proposal importance sampling distribution for $(\boldsymbol{X}, P)$. According to Eq. (5), $f''(\boldsymbol{x})$ can be formulated as:

$$f''(\boldsymbol{x}) = \frac{\int_{P\in\Omega_{acc}} f'(\boldsymbol{x})\,\mathrm{d}p}{\int_{[\boldsymbol{X},P]\in\Omega_{acc}} f'(\boldsymbol{x})\,\mathrm{d}p\mathrm{d}\boldsymbol{x}} = \frac{\int_{P\in\Omega_{acc}} \frac{f'(\boldsymbol{x})}{f_{IS}(\boldsymbol{x},p)} f_{IS}(\boldsymbol{x},p)\,\mathrm{d}p}{\int_{[\boldsymbol{X},P]\in\Omega_{acc}} \frac{f'(\boldsymbol{x})}{f_{IS}(\boldsymbol{x},p)} f_{IS}(\boldsymbol{x},p)\,\mathrm{d}p\mathrm{d}\boldsymbol{x}} \tag{20}$$

Subsequently, the estimation of the CDF of the posterior distribution can be formulated as:

$$\begin{aligned} CDF(\boldsymbol{\psi}) &= \int_{\Omega} I(\boldsymbol{x} \le \boldsymbol{\psi}) f''(\boldsymbol{x}) d\boldsymbol{x} = \frac{\int_{\Omega}\int_{P\in\Omega_{acc}} I(\boldsymbol{x} \le \boldsymbol{\psi}) \frac{f'(\boldsymbol{x})}{f_{IS}(\boldsymbol{x},p)} f_{IS}(\boldsymbol{x},p)\, dp d\boldsymbol{x}}{\int_{[\boldsymbol{X},P]\in\Omega_{acc}} \frac{f'(\boldsymbol{x})}{f_{IS}(\boldsymbol{x},p)} f_{IS}(\boldsymbol{x},p)\, dp d\boldsymbol{x}} \\ &= \frac{\int_{[\boldsymbol{X},P]\in\Omega_{acc}} \frac{I(\boldsymbol{x} \le \boldsymbol{\psi}) f'(\boldsymbol{x})}{f_{IS}(\boldsymbol{x},p)} f_{IS}(\boldsymbol{x},p)\,\mathrm{d}p\mathrm{d}\boldsymbol{x}}{\int_{[\boldsymbol{X},P]\in\Omega_{acc}} \frac{f'(\boldsymbol{x})}{f_{IS}(\boldsymbol{x},p)} f_{IS}(\boldsymbol{x},p)\, dp d\boldsymbol{x}} \\ &\approx \frac{\sum_{k=1}^{N_{IS}} I(\boldsymbol{x}_k \le \boldsymbol{\psi}) I_{g\le 0}(\boldsymbol{x}_k, p_k) \frac{f'(\boldsymbol{x}_k)}{f_{IS}(\boldsymbol{x}_k, p_k)}}{\sum_{k=1}^{N_{IS}} I_{g\le 0}(\boldsymbol{x}_k, p_k) \frac{f'(\boldsymbol{x}_k)}{f_{IS}(\boldsymbol{x}_k, p_k)}} \end{aligned} \tag{21}$$

where N_{IS} is the number of IS samples. The estimates of the first and the second moments of the posterior distribution with IS-based BUS can be derived as follows:

$$\begin{aligned}\boldsymbol{\mu}_{f''} &= \frac{\int_{[\boldsymbol{X},P]\in\Omega_f} \boldsymbol{x} f'(\boldsymbol{x}) dp d\boldsymbol{x}}{\int_{[\boldsymbol{X},P]\in\Omega_f} f'(\boldsymbol{x}) dp d\boldsymbol{x}} = \frac{\int_{[\boldsymbol{X},P]\in\Omega_f} \boldsymbol{x} \frac{f'(\boldsymbol{x})}{f_{IS}(\boldsymbol{x},p)} f_{IS}(\boldsymbol{x},p) dp d\boldsymbol{x}}{\int_{[\boldsymbol{X},P]\in\Omega_f} \frac{f'(\boldsymbol{x})}{f_{IS}(\boldsymbol{x},p)} f_{IS}(\boldsymbol{x},p) dp d\boldsymbol{x}} \\ &\approx \frac{\sum_{k=1}^{N_{IS}} \boldsymbol{x}_k \cdot I_{g\le 0}(\boldsymbol{x}_k,p_k) \cdot \frac{f'(\boldsymbol{x}_k)}{f_{IS}(\boldsymbol{x}_k,p_k)}}{\sum_{k=1}^{N_{IS}} I_{g\le 0}(\boldsymbol{x}_k,p_k) \cdot \frac{f'(\boldsymbol{x}_k)}{f_{IS}(\boldsymbol{x}_k,p_k)}}\end{aligned} \tag{22}$$

$$\begin{aligned}\boldsymbol{\Sigma}_{f''} &= \frac{\int_{[\boldsymbol{X},P]\in\Omega_f} (\boldsymbol{x}-\boldsymbol{\mu}_{f''})^T (\boldsymbol{x}-\boldsymbol{\mu}_{f''}) f'(\boldsymbol{x}) dp d\boldsymbol{x}}{\int_{[\boldsymbol{X},P]\in\Omega_f} f'(\boldsymbol{x}) dp d\boldsymbol{x}} \\ &= \frac{\int_{[\boldsymbol{X},P]\in\Omega_f} (\boldsymbol{x}-\boldsymbol{\mu}_{f''})^T (\boldsymbol{x}-\boldsymbol{\mu}_{f''}) \frac{f'(\boldsymbol{x})}{f_{IS}(\boldsymbol{x},p)} f_{IS}(\boldsymbol{x},p) dp d\boldsymbol{x}}{\int_{[\boldsymbol{X},P]\in\Omega_f} \frac{f'(\boldsymbol{x})}{f_{IS}(\boldsymbol{x},p)} f_{IS}(\boldsymbol{x},p) dp d\boldsymbol{x}} \\ &\approx \frac{\sum_{k=1}^{N_{IS}} (\boldsymbol{x}-\hat{\boldsymbol{\mu}}_{f''})^T (\boldsymbol{x}-\hat{\boldsymbol{\mu}}_{f''}) I_{g\le 0}(\boldsymbol{x}_k,p_k) \frac{f'(\boldsymbol{x}_k)}{f_{IS}(\boldsymbol{x}_k,p_k)}}{\sum_{k=1}^{N_{IS}} I_{g\le 0}(\boldsymbol{x}_k,p_k) \frac{f'(\boldsymbol{x}_k)}{f_{IS}(\boldsymbol{x}_k,p_k)}}\end{aligned} \tag{23}$$

3.2 On the Accuracy of the Estimates of the Posterior Distribution

As Eq. (21) - (23) show, the posterior distribution can be estimated with realizations in IS-based BUS. The number of realizations in the acceptance domain is therefore critical to achieve a reliable estimation. Thus, this paper proposes a new criterion to check the sufficiency of the number of IS realizations and therefore guarantee the quality of the estimate for the posterior distribution.

First, Eq. (22) can be reformulated as:

$$\begin{aligned}\boldsymbol{\mu}_{f''} &= \frac{\int_{[\boldsymbol{X},P]\in\Omega_f} \boldsymbol{x} f'(\boldsymbol{x}) dp d\boldsymbol{x}}{\int_{[\boldsymbol{X},P]\in\Omega_f} f'(\boldsymbol{x}) dp d\boldsymbol{x}} = \frac{\int_{[\boldsymbol{X},P]\in\Omega_f} \boldsymbol{x} \frac{f'(\boldsymbol{x})}{f_{IS}(\boldsymbol{x},p)} f_{IS}(\boldsymbol{x},p) dp d\boldsymbol{x}}{\int_{[\boldsymbol{X},P]\in\Omega_f} \frac{f'(\boldsymbol{x})}{f_{IS}(\boldsymbol{x},p)} f_{IS}(\boldsymbol{x},p) dp d\boldsymbol{x}} \\ &= \frac{\mathbb{E}_{f_{IS}}\left[\boldsymbol{x} \frac{f'(\boldsymbol{x})}{f_{IS}(\boldsymbol{x},p)} I_{g\le 0}(\boldsymbol{x},p)\right]}{\mathbb{E}_{f_{IS}}\left[\frac{f'(\boldsymbol{x})}{f_{IS}(\boldsymbol{x},p)} I_{g\le 0}(\boldsymbol{x},p)\right]}\end{aligned} \tag{24}$$

where $\mathbb{E}_{f_{IS}}$ represents the expectation with $(\boldsymbol{X}, P)$ following the proposal distribution f_{IS}.

Considering that d is the dimension of $\boldsymbol{X}$, $\boldsymbol{\mu}_{f''}$ can be defined as the vector $[\mu_{f''}{}^1, \cdots, \mu_{f''}{}^d]$, where $\mu_{f''}{}^i$ denotes the mean of the posterior distribution in the dimension i ($i \in [1, \cdots, d]$). The following equation can be therefore obtained:

$$\mu_{f''}{}^i = \frac{\mathbb{E}_{f_{IS}}\left[x^i \frac{f'(\boldsymbol{x})}{f_{IS}(\boldsymbol{x},p)} I_{g\le 0}(\boldsymbol{x},p)\right]}{\mathbb{E}_{f_{IS}}\left[\frac{f'(\boldsymbol{x})}{f_{IS}(\boldsymbol{x},p)} I_{g\le 0}(\boldsymbol{x},p)\right]} \tag{25}$$

where $\boldsymbol{x} = [x^1, \cdots, x^d]$ and the superscript denotes the dimension number.

Let $\mathcal{A}^i$ denote the function $x^i \frac{f'(x)}{f_{IS}(x,p)} I_{g\le 0}(\boldsymbol{x},p)$ and $\mathcal{B}$ denote $\frac{f'(x)}{f_{IS}(x,p)} I_{g\le 0}(\boldsymbol{x},p)$. With N_{IS} realizations, the mean and the variance of $\mathcal{A}^i$ and $\mathcal{B}$ can be estimated as:

$$\hat{\mu}_{\mathcal{A}^i} \approx \frac{1}{N_{IS}} \sum_{k=1}^{N_{IS}} x_k^i \frac{f'(\boldsymbol{x}_k)}{f_{IS}(\boldsymbol{x}_k, p_k)} I_{g\leq 0}(\boldsymbol{x}_k, p_k) \tag{26}$$

$$\hat{\mu}_{\mathcal{B}} \approx \frac{1}{N_{IS}} \sum_{k=1}^{N_{IS}} \frac{f'(\boldsymbol{x}_k)}{f_{IS}(\boldsymbol{x}_k, p_k)} I_{g\leq 0}(\boldsymbol{x}_k, p_k) \tag{27}$$

$$\widehat{var}_{\mathcal{A}^i} \approx \frac{1}{N_{IS}-1} \sum_{k=1}^{N_{IS}} \left(x_k^i \frac{f'(\boldsymbol{x}_k)}{f_{IS}(\boldsymbol{x}_k, p_k)} I_{g\leq 0}(\boldsymbol{x}_k, p_k) - \hat{\mu}_{\mathcal{A}^i} \right)^2 \tag{28}$$

$$\widehat{var}_{\mathcal{B}} \approx \frac{1}{N_{IS}-1} \sum_{k=1}^{N_{IS}} \left(\frac{f'(\boldsymbol{x}_k)}{f_{IS}(\boldsymbol{x}_k, p_k)} I_{g\leq 0}(\boldsymbol{x}_k, p_k) - \hat{\mu}_{\mathcal{B}} \right)^2 \tag{29}$$

According to the Central Limit Theorem, the variances for the estimates of $\hat{\mu}_{\mathcal{A}^i}$ and $\hat{\mu}_{\mathcal{B}}$ can be obtained as $\frac{var_{\mathcal{A}^i}}{N_{IS}}$ and $\frac{var_{\mathcal{B}}}{N_{IS}}$. Here, $\widehat{var}_{\mathcal{A}^i}$ and $\widehat{var}_{\mathcal{B}}$ are considered to approximate $var_{\mathcal{A}^i}$ and $var_{\mathcal{B}}$; and $\hat{\mu}_{\mathcal{A}^i}$ and $\hat{\mu}_{\mathcal{B}}$ are considered to approximate $\mu_{\mathcal{A}^i}$ and $\mu_{\mathcal{B}}$. The variances for the estimates of $\hat{\mu}_{\mathcal{A}^i}$ and $\hat{\mu}_{\mathcal{B}}$ can thus be determined as $\frac{\widehat{var}_{\mathcal{A}^i}}{N_{IS}}$ and $\frac{\widehat{var}_{\mathcal{B}}}{N_{IS}}$; and the coefficients of variation for the mean estimates can be determined as $COV_{\hat{\mu}_{\mathcal{A}^i}} = \sqrt{\frac{\widehat{var}_{\mathcal{A}^i}}{N_{IS}}} / \hat{\mu}_{\mathcal{A}^i}$ and $COV_{\hat{\mu}_{\mathcal{B}}} = \sqrt{\frac{\widehat{var}_{\mathcal{B}}}{N_{IS}}} / \hat{\mu}_{\mathcal{B}}$.

Given a significance level α and the corresponding z-value γ_α, the estimates $\hat{\mu}_{\mathcal{A}^i}$ and $\hat{\mu}_{\mathcal{B}}$ will be inside the following intervals according to the Central Limit Theorem:

$$\mu_{\mathcal{A}^i} \in [\hat{\mu}_{\mathcal{A}^i} - \gamma_\alpha \sqrt{\frac{\widehat{var}_{\mathcal{A}^i}}{N_{IS}}}, \hat{\mu}_{\mathcal{A}^i} + \gamma_\alpha \sqrt{\frac{\widehat{var}_{\mathcal{A}^i}}{N_{IS}}}] \tag{30}$$

$$\mu_{\mathcal{B}} \in [\hat{\mu}_{\mathcal{B}} - \gamma_\alpha \sqrt{\frac{\widehat{var}_{\mathcal{B}}}{N_{IS}}}, \hat{\mu}_{\mathcal{B}} + \gamma_\alpha \sqrt{\frac{\widehat{var}_{\mathcal{B}}}{N_{IS}}}] \tag{31}$$

Therefore, controlling the values of $COV_{\hat{\mu}_{\mathcal{A}^i}}$ and $COV_{\hat{\mu}_{\mathcal{B}}}$ can help to narrow the bounds for $\mu_{\mathcal{A}^i}$ and $\mu_{\mathcal{B}}$, therefore yielding a more reliable estimate of $\mu_{f''}{}^i$ with one simulation. Let $\boldsymbol{\epsilon}_{\boldsymbol{\mu}_{f''}}$ denote the vector $\left[COV_{\hat{\mu}_{\mathcal{A}^1}}, \cdots, COV_{\hat{\mu}_{\mathcal{A}^d}}, COV_{\hat{\mu}_{\mathcal{B}}}\right]$. In this paper, the maximum value of entries of vector $\boldsymbol{\epsilon}_{\boldsymbol{\mu}_{f''}}$ is adopted as the stopping criterion in the framework to guarantee that the number of realizations is enough to achieve a reliable estimate of the posterior distribution.

As this section focuses on whether the number of accepted samples can accurately describe the posterior distribution, which is considered as a deterministic result from the BUS method, the concept of confidence intervals is adopted. However, note that when the bound for the true parameters of the distribution is of interest, the credible interval should be employed according to Bayesian statistics, and typically more observed samples can help to reduce uncertainties and narrow the credible interval.

3.3 Quantification for the Accuracy of Kriging

The Kriging surrogate model has recently received much attention for reliability analysis problems as it has offered a viable substitute for expensive-to-evaluate limit state functions. Recent studies on Kriging-based reliability analysis include, among others, the development of learning functions [39-43] and quantification of the maximum error, which is introduced by the use of Kriging, for the estimation of the failure probability [44-46]. However, using this maximum error as the stopping criterion for the Kriging surrogate model when estimating posterior distributions may not be adequate [47], as the probability of failure is not identical to the posterior distribution, which BUS aims to estimate. Studies in [34] found that the use of U learning function [40], which measures the probability of wrong classification for each realization, as a stopping criterion in Bayesian updating problems can lead to a large number of unnecessary calls of the limit state function. Thus, a more effective stopping criterion is needed for the active learning of Kriging in the proposed Bayesian updating framework. To address this gap, the rest of this section discusses the maximum error for the estimation of the first moment of the posterior distribution and proposes a new stopping criterion.

When Kriging is applied as a substitute of the original limit state function $g(\cdot)$, the Kriging predictor can be denoted as $g_K(\cdot)$. As mentioned in Section 2.3, Kriging assumes that the response for a test point follows a Gaussian distribution with mean μ_K and σ_K^2. In this setting, μ_K is often used as g_K. The failure domain with Kriging model can be defined as follows:

$$\Omega_{f,K} = [g_K(\boldsymbol{x}, p) \leq 0] \tag{32}$$

The first order moment of the posterior distribution can therefore be reformulated as:

$$\boldsymbol{\mu}_{f'',K} = \frac{\mathbb{E}_{f_{IS}}\left[\boldsymbol{x}\frac{f'(\boldsymbol{x})}{f_{IS}(\boldsymbol{x},p)} I_{g_K\leq 0}(\boldsymbol{x},p)\right]}{\mathbb{E}_{f_{IS}}\left[\frac{f'(\boldsymbol{x})}{f_{IS}(\boldsymbol{x},p)} I_{g_K\leq 0}(\boldsymbol{x},p)\right]} \approx \frac{\sum_{k=1}^{N_{IS}} \boldsymbol{x}_k I_{g_K\leq 0}(\boldsymbol{x}_k,p_k)\frac{f'(\boldsymbol{x}_k)}{f_{IS}(\boldsymbol{x}_k,p_k)}}{\sum_{k=1}^{N_{IS}} I_{g_K\leq 0}(\boldsymbol{x}_k,p_k)\frac{f'(\boldsymbol{x}_k)}{f_{IS}(\boldsymbol{x}_k,p_k)}} \tag{33}$$

Note that the error caused by approximating the original model with Kriging stems from the wrong classification of samples, namely $I_{g_K\leq 0}(\boldsymbol{x}_k,p_k) \neq I_{g\leq 0}(\boldsymbol{x}_k,p_k)$ for the realizations with wrong sign estimation. For the numerator of Eq. (33), the sum of the absolute difference between each realization estimated by the original limit state function and that by the Kriging surrogate model can be expressed as:

$$\boldsymbol{D}_n = \sum_{k=1}^{N_{IS}} \boldsymbol{x}_k \frac{f'(\boldsymbol{x}_k)}{f_{IS}(\boldsymbol{x}_k,p_k)} \left|I_{g\leq 0}(\boldsymbol{x}_k,p_k) - I_{g_K\leq 0}(\boldsymbol{x}_k,p_k)\right| \tag{34}$$

where $\boldsymbol{D}_n = [D_n^1, \cdots, D_n^d]$ is a vector containing the absolute error caused by Kriging in each dimension and d denotes the number of dimensions of $\boldsymbol{X}$. D_n^i ($i \in [1, \cdots, d]$) can be expressed as:

$$D_n^i = \sum_{k=1}^{N_{IS}} x_k^i \frac{f'(\boldsymbol{x}_k)}{f_{IS}(\boldsymbol{x}_k,p_k)} \left|I_{g\leq 0}(\boldsymbol{x}_k,p_k) - I_{g_K\leq 0}(\boldsymbol{x}_k,p_k)\right| \tag{35}$$

The same applies to the denominator of Eq. (33); therefore, the sum of the absolute error can be formulated as:

$$D_d = \sum_{k=1}^{N_{IS}} \frac{f'(\boldsymbol{x}_k)}{f_{IS}(\boldsymbol{x}_k,p_k)} \left|I_{g\leq 0}(\boldsymbol{x}_k,p_k) - I_{g_K\leq 0}(\boldsymbol{x}_k,p_k)\right| \tag{36}$$

However, the exact value of $\left|I_{g\leq 0}(\boldsymbol{x}_k,p_k) - I_{g_K\leq 0}(\boldsymbol{x}_k,p_k)\right|$ is unknown, as Kriging aims at replacing the original limit state function $g(\cdot)$, which can be expensive-to-evaluate. Thus, in the following, the estimation of D_n^i and D_d based on the Gaussian process assumption of Kriging is presented.

For an arbitrary test point $(\boldsymbol{x}_k, p_k)$ with predicted mean μ_K and variance σ_K^2 provided by Eq. (17) and Eq. (18), the probability of providing a wrong sign estimation for this point can be formulated as follows:

$$P_k^{wse} = \Phi\left(-\frac{|\mu_K(\boldsymbol{x}_k,p_k)|}{\sigma_K(\boldsymbol{x}_k,p_k)}\right) \tag{37}$$

Let $I_w(\cdot)$ denote the function $\left|I_{g\leq 0}(\boldsymbol{x}_k,p_k) - I_{g_K\leq 0}(\boldsymbol{x}_k,p_k)\right|$, which is the indicator of the wrong classification event. When Kriging provides a wrong sign estimation for $(\boldsymbol{x}_k,p_k)$, $I_w(\boldsymbol{x}_k,p_k)$ is 1; otherwise, $I_w(\boldsymbol{x}_k,p_k)$ is 0. Therefore, $I_w(\boldsymbol{x}_k,p_k)$ follows a Binomial distribution. The same applies to $x_k^i I_w(\boldsymbol{x}_k,p_k)\frac{f'(\boldsymbol{x}_k)}{f_{IS}(\boldsymbol{x}_k,p_k)}$, denoted as $\mathcal{P}_k^i$, and $I_w(\boldsymbol{x}_k,p_k)\frac{f'(\boldsymbol{x}_k)}{f_{IS}(\boldsymbol{x}_k,p_k)}$, denoted as $\mathcal{J}_k$, where $k \in [1, \cdots, N_{IS}]$ and $i \in [1, \cdots, d]$. The means for $\mathcal{P}_k^i$ and $\mathcal{J}_k$ can be formulated as:

$$\mu_{\mathcal{P}_k^i} = P_k^{wse}\, x_k^i \frac{f'(\boldsymbol{x}_k)}{f_{IS}(\boldsymbol{x}_k,p_k)} \tag{38}$$

$$\mu_{\mathcal{J}_k} = P_k^{wse}\, \frac{f'(\boldsymbol{x}_k)}{f_{IS}(\boldsymbol{x}_k,p_k)} \tag{39}$$

The variances for $\mathcal{P}_k^i$ and $\mathcal{J}_k$ can be formulated as:

$$var_{\mathcal{P}_k^i} = P_k^{wse}(1 - P_k^{wse})\left[x_k^i \frac{f'(\boldsymbol{x}_k)}{f_{IS}(\boldsymbol{x}_k, p_k)}\right]^2 \tag{40}$$

$$var_{\mathcal{J}_k} = P_k^{wse}(1 - P_k^{wse})\left[\frac{f'(\boldsymbol{x}_k)}{f_{IS}(\boldsymbol{x}_k, p_k)}\right]^2 \tag{41}$$

Assuming that the wrong classification events are independent for these realizations, the sum of independent Binomial distributions follows the Poisson Binomial distribution [44-46]. Let $\Xi_{\mathcal{P}i}$ denote $\sum_{k=1}^{N_{IS}} \mathcal{P}_k^i$ and $\Xi_{\mathcal{J}}$ denote $\sum_{k=1}^{N_{IS}} \mathcal{J}_k$. The mean and variance for $\Xi_{\mathcal{P}i}$ can be expressed as:

$$\mu_{\Xi_{\mathcal{P}i}} = \sum_{k=1}^{N_{IS}} \mu_{\mathcal{P}_k^i} \tag{42}$$

$$var_{\Xi_{\mathcal{P}i}} = \sum_{k=1}^{N_{IS}} var_{\mathcal{P}_k^i} \tag{43}$$

The mean and variance for $\Xi_{\mathcal{J}}$ can be formulated as:

$$\mu_{\Xi_{\mathcal{J}}} = \sum_{k=1}^{N_{IS}} \mu_{\mathcal{J}_k} \tag{44}$$

$$var_{\Xi_{\mathcal{J}}} = \sum_{k=1}^{N_{IS}} var_{\mathcal{J}_k} \tag{45}$$

Given the significance level α, the confidence intervals for $\Xi_{\mathcal{P}i}$ and $\Xi_{\mathcal{J}}$ can be estimated with the Central Limit Theorem as:

$$\Xi_{\mathcal{P}i} \in \left[\mu_{\Xi_{\mathcal{P}i}} - \gamma_\alpha \sqrt{var_{\Xi_{\mathcal{P}i}}}, \mu_{\Xi_{\mathcal{P}i}} + \gamma_\alpha \sqrt{var_{\Xi_{\mathcal{P}i}}}\right] \tag{46}$$

$$\Xi_{\mathcal{J}} \in \left[\mu_{\Xi_{\mathcal{J}}} - \gamma_\alpha \sqrt{var_{\Xi_{\mathcal{J}}}}, \mu_{\Xi_{\mathcal{J}}} + \gamma_\alpha \sqrt{var_{\Xi_{\mathcal{J}}}}\right] \tag{47}$$

Thus, the confidence interval for D_n^i is estimated as:

$$D_n^i = \Xi_{\mathcal{P}i} \in \left[\mu_{\Xi_{\mathcal{P}i}} - \gamma_\alpha \sqrt{var_{\Xi_{\mathcal{P}i}}}, \mu_{\Xi_{\mathcal{P}i}} + \gamma_\alpha \sqrt{var_{\Xi_{\mathcal{P}i}}}\right] \tag{48}$$

And the confidence interval for D_d is estimated as:

$$D_d = \Xi_{\mathcal{J}} \in \left[\mu_{\Xi_{\mathcal{J}}} - \gamma_\alpha \sqrt{var_{\Xi_{\mathcal{J}}}}, \mu_{\Xi_{\mathcal{J}}} + \gamma_\alpha \sqrt{var_{\Xi_{\mathcal{J}}}}\right] \tag{49}$$

Subsequently, the maximum relative error for estimating the numerator of Eq. (33) with the Kriging surrogate model can be approximated as:

$$\epsilon_{D_n^i} = max\left(\left|\frac{\mu_{\Xi_{\mathcal{P}i}} - \gamma_\alpha \sqrt{var_{\Xi_{\mathcal{P}i}}}}{\sum_{k=1}^{N_{IS}} x_k^i I_{g_K \le 0}(\boldsymbol{x}_k, p_k) \frac{f'(\boldsymbol{x}_k)}{f_{IS}(\boldsymbol{x}_k, p_k)}}\right|, \left|\frac{\mu_{\Xi_{\mathcal{P}i}} + \gamma_\alpha \sqrt{var_{\Xi_{\mathcal{P}i}}}}{\sum_{k=1}^{N_{IS}} x_k^i I_{g_K \le 0}(\boldsymbol{x}_k, p_k) \frac{f'(\boldsymbol{x}_k)}{f_{IS}(\boldsymbol{x}_k, p_k)}}\right|\right) \tag{50}$$

The same applies for D_d:

$$\epsilon_{D_d} = max\left(\left|\frac{\mu_{\Xi_{\mathcal{J}}} - \gamma_\alpha \sqrt{var_{\Xi_{\mathcal{J}}}}}{\sum_{k=1}^{N_{IS}} I_{g_K \le 0}(\boldsymbol{x}_k, p_k) \frac{f'(\boldsymbol{x}_k)}{f_{IS}(\boldsymbol{x}_k, p_k)}}\right|, \left|\frac{\mu_{\Xi_{\mathcal{J}}} + \gamma_\alpha \sqrt{var_{\Xi_{\mathcal{J}}}}}{\sum_{k=1}^{N_{IS}} I_{g_K \le 0}(\boldsymbol{x}_k, p_k) \frac{f'(\boldsymbol{x}_k)}{f_{IS}(\boldsymbol{x}_k, p_k)}}\right|\right) \tag{51}$$

Note that for accurately estimating the first moment of the posterior distribution with the Kriging surrogate model, both the numerator and the denominator should be accurately quantified. Thus, the maximum value among the vector $\boldsymbol{\epsilon}_{\boldsymbol{D}_n} = [\epsilon_{D_n^1}, \cdots, \epsilon_{D_n^d}]$ and ϵ_{D_d} is checked as the stopping criterion for the active learning of Kriging in the proposed method.

3.4 The Proposed Method: BUAK-AIS

Based on the above formulations in Section 3.1 to 3.3, this section presents the proposed method: Bayesian Updating with adaptive Kriging-based Adaptive Importance Sampling, BUAK-AIS. The flowchart of the proposed method is shown in Fig. 1. In the first stage, N_{F} low-discrepancy samples are generated within the sampling region with a sampling technique, such as Sobol Sequence [48]. A fraction of these generated samples are selected as the initial training data for the Kriging surrogate model in the second step. Let N_I denote the number of initial samples. Then, in the third step of the proposed method, the Kriging surrogate model is established based on the current training data. Subsequently, the fourth step of the proposed method concentrates on determining the importance sampling distribution. The optimal proposal distribution in importance sampling can be theoretically derived as [49]:

$$f_{IS}^*(\boldsymbol{x}, p) = \frac{I_{g\leq 0}(\boldsymbol{x}, p) f'(\boldsymbol{x})}{\int_\Omega \int_0^1 I_{g\leq 0}(\boldsymbol{x}, p) f'(\boldsymbol{x})\, \mathrm{d}p \mathrm{d}\boldsymbol{x}} \tag{52}$$

As directly sampling from f_{IS}^* is challenging, a Gaussian mixture distribution, which consists of multiple multivariate Gaussian distribution components, is introduced as the quasi-optimal importance sampling distribution in this paper. Let K denote the number of Gaussian distributions; $\boldsymbol{\mu}_i$, $\boldsymbol{\Sigma}_i$ and π_i denote the mean, covariance matrix and the likelihood of selecting the ith Gaussian distribution ($i \in [1, \cdots, K]$). Thus, a Gaussian mixture distribution can be parametrized by $\boldsymbol{u} = \{\pi_1, \cdots, \pi_K, \boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_K, \boldsymbol{\Sigma}_1, \cdots, \boldsymbol{\Sigma}_K\}$. The optimality of the established Gaussian mixture distribution can be measured by the Kullback–Leibler-based cross-entropy as follows [50]:

$$\mathrm{KL}(\boldsymbol{u}) = \int_\Omega \int_0^1 f_{IS}^*(\boldsymbol{x}, p) In f_{IS}^*(\boldsymbol{x}, p) dp d\boldsymbol{x} - \int_\Omega \int_0^1 f_{IS}^*(\boldsymbol{x}, p) In h(\boldsymbol{x}, p, \boldsymbol{u}) dp d\boldsymbol{x} \tag{53}$$

where $h(\boldsymbol{x}, p, \boldsymbol{u})$ is the Gaussian mixture distribution with the set of parameters $\boldsymbol{u}$. As only the second part of KL depends on $\boldsymbol{u}$, the determination of $\boldsymbol{u}$ can be considered as an optimization problem to minimize the discrepancy:

$$\boldsymbol{u}^* = \underset{\boldsymbol{u}}{\operatorname{argmax}} \int_\Omega \int_0^1 I_{g\leq 0}(\boldsymbol{x}, p) f'(\boldsymbol{x}) In h(\boldsymbol{x}, p, \boldsymbol{u}) dp d\boldsymbol{x} = \underset{\boldsymbol{u}}{\operatorname{argmax}}\, \mathbb{E}_{f'}[I_{g\leq 0}(\boldsymbol{x}, p) In h(\boldsymbol{x}, p, \boldsymbol{u})] \tag{54}$$

where $\mathbb{E}_{f'}[\cdot]$ denotes the mathematical expectation with respect to the original joint PDF of $(\boldsymbol{X}, P)$.

For more efficient estimation, an alternative density can be introduced to reformulate the above equation as follows:

$$\begin{aligned} \boldsymbol{u}^* &= \underset{\boldsymbol{u}}{\operatorname{argmax}} \int_\Omega \int_0^1 I_{g\leq 0}(\boldsymbol{x}, p) f'(\boldsymbol{x}) In h(\boldsymbol{x}, p, \boldsymbol{u}) \frac{h(\boldsymbol{x}, p, \boldsymbol{w})}{h(\boldsymbol{x}, p, \boldsymbol{w})} dp d\boldsymbol{x} \\ &= \underset{\boldsymbol{u}}{\operatorname{argmax}}\, \mathbb{E}_w[I_{g\leq 0}(\boldsymbol{x}, p) In h(\boldsymbol{x}, p, \boldsymbol{u}) \frac{f'(\boldsymbol{x})}{h(\boldsymbol{x}, p, \boldsymbol{w})}] \\ &\approx \underset{\boldsymbol{u}}{\operatorname{argmax}} \sum_{i=1}^{N_{CE}} \frac{1}{N_E} I_{g\leq 0}(\boldsymbol{x}_i, p_i) In h(\boldsymbol{x}_i, p_i, \boldsymbol{u}) \frac{f'(\boldsymbol{x}_i)}{h(\boldsymbol{x}_i, p_i, \boldsymbol{w})} \end{aligned} \tag{55}$$

where N_{CE} denotes the number of realizations to estimate expectation, and $h(\boldsymbol{x}, p, \boldsymbol{w})$ represents the alternative sampling distribution. $\boldsymbol{w}$ is not optimal but facilitates more efficient sampling compared with the original PDF of $(\boldsymbol{X}, P)$.

Then, in the fifth step of the proposed method, N_{IS} realizations are sampled from the obtained distribution $h(\boldsymbol{x}, p, \boldsymbol{u}^*)$ for estimating the posterior distribution with the current Kriging model, shown by Eq. (21) to Eq. (23). The accuracy of the Kriging surrogate model is checked in the sixth step. As discussed in Section 3.3, if the maximum value among the vector $\boldsymbol{\epsilon}_{\boldsymbol{D}_n} = [\epsilon_{D_n^1}, \cdots, \epsilon_{D_n^d}]$ and ϵ_{D_d} does not exceed the threshold (ϵ_K, set as 0.05 in this paper), the algorithm enters the seventh step where the next training point is selected from the sampling pool with the U learning function [40]. Otherwise, the algorithm goes to the eighth step where the maximum value of $\boldsymbol{\epsilon}_{\boldsymbol{\mu}_{f''}}$ is checked to guarantee that the number of realizations is enough for a robust estimation. If the maximum value of $\boldsymbol{\epsilon}_{\boldsymbol{\mu}_{f''}}$ does not exceed the threshold (ϵ_T, set as 0.05 in this paper), the algorithm ends. Otherwise, the algorithm goes to ninth step to increase the value of N_{IS}, and then enters the fifth step again.

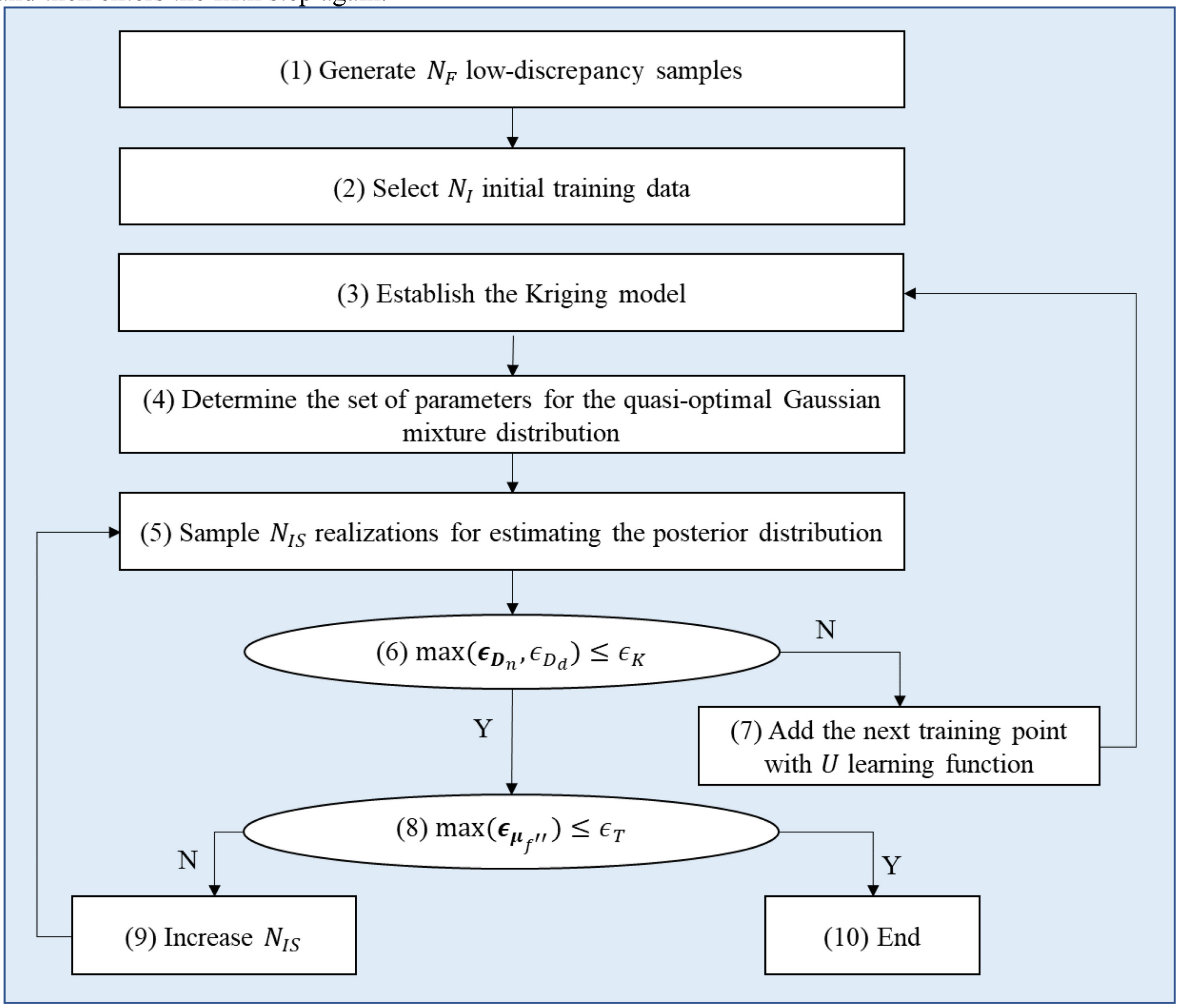


Fig. 1 Flowchart of the proposed method: BUAK-AIS

A more detailed algorithm table is also provided as follows for the readability:

Algorithm 1. The proposed BUAK-AIS algorithm

---

1. Generate N_F low-discrepancy samples within the sampling space. This paper generates these samples with the Sobol sequence.
2. A fraction of the generated samples are selected as the initial training data for the Kriging surrogate model. In this paper, $\frac{(d+1)(d+2)}{2} + 1$ initial samples are selected from the pool as the initial database, where d denotes the dimension of problems.
3. Establish the Kriging surrogate model. The present work adopts the DACE package embedded in the platform of MATLAB software. The leave-one-out cross validation is used to select the trend that provides the minimum error (from the constant, linear and quadratic trends).

4. Determine the IS distribution based on the current Kriging model. The following steps are executed:
(a) Initialize the IS distribution (defining $\boldsymbol{w}$): K samples are randomly selected from the failure samples predicted by the current Kriging model. These K samples are adopted as the initial center of the Gaussian mixture distribution with uniform weights with unit-standard deviations and zero correlations. If the number of failure points is smaller than K, the sample that minimizes the performance function is added to the training data and used as the center of a standard Gaussian distribution to generate more realizations for enriching the pool until K failure samples are found.
(b) Generate IS samples: N_{CE} realizations are generated from the initial IS distribution.
(c) Update IS distribution (estimating $\boldsymbol{u}^*$): introduce $\gamma_{i,j} = \frac{\pi_j^0 N(\boldsymbol{x}_i|\boldsymbol{\mu}_j^0,\boldsymbol{\Sigma}_j^0)}{\sum_{k=1}^{K}\pi_k^0 N(\boldsymbol{x}_i|\boldsymbol{\mu}_k^0,\boldsymbol{\Sigma}_k^0)}$ and $W(\boldsymbol{x},p,\boldsymbol{w}) = \frac{f'(x)}{h(\boldsymbol{x},p,\boldsymbol{w})}$, where $h(\boldsymbol{x},p,\boldsymbol{w})$ is the joint PDF for the Gaussian mixture distribution parametrized by $\boldsymbol{w}$, and $\boldsymbol{w} = \{\pi_1^0,\cdots,\pi_K^0,\boldsymbol{\mu}_1^0,\cdots,\boldsymbol{\mu}_K^0,\boldsymbol{\Sigma}_1^0,\cdots,\boldsymbol{\Sigma}_K^0\}$ is defined in part (a) of this step, where $i \in [1,\cdots,N_{CE}]$, $j \in [1,\cdots,K]$. The following equations are used to update the IS distribution according to [50]:

$$\boldsymbol{\mu}_j^* = \frac{\sum_{i=1}^{N_{CE}} I_{g_K\le 0}(\boldsymbol{x}_i,p_i)W(\boldsymbol{x}_i,p_i,\boldsymbol{w})\gamma_{i,j}[\boldsymbol{x}_i,p_i]}{\sum_{i=1}^{N_{CE}} I_{g_K\le 0}(\boldsymbol{x}_i,p_i)W(\boldsymbol{x}_i,p_i,\boldsymbol{w})\gamma_{i,j}} \tag{56}$$

$$\boldsymbol{\Sigma}_j^* = \frac{\sum_{i=1}^{N_{CE}} I_{g_K\le 0}(\boldsymbol{x}_i,p_i)W(\boldsymbol{x}_i,p_i,\boldsymbol{w})\gamma_{i,j}\left([\boldsymbol{x}_i,p_i]-\boldsymbol{\mu}_j^*\right)\left([\boldsymbol{x}_i,p_i]-\boldsymbol{\mu}_j^*\right)^{\mathrm{T}}}{\sum_{i=1}^{N_{CE}} I_{g_K\le 0}(\boldsymbol{x}_i,p_i)W(\boldsymbol{x}_i,p_i,\boldsymbol{w})\gamma_{i,j}} \tag{57}$$

$$\pi_j^* = \frac{\sum_{i=1}^{N_{CE}} I_{g_K\le 0}(\boldsymbol{x}_i,p_i)W(\boldsymbol{x}_i,p_i,\boldsymbol{w})\gamma_{i,j}}{\sum_{i=1}^{N_{CE}} I_{g_K\le 0}(\boldsymbol{x}_i,p_i)W(\boldsymbol{x}_i,p_i,\boldsymbol{w})} \tag{58}$$

Note that $[\boldsymbol{x}_i,p_i]$ denotes the vector in the augmented space, and N_{CE} is the number of samples used to estimate the optimal IS distribution. K is set as 40 for Example Three; and 10 otherwise.
5. Generate N_{IS} realizations from the updated IS distribution. This paper defines the initial N_{IS} as 10^4 for Case Two in Example Two and Example Three; and 10^3, otherwise.
6. Calculate $\boldsymbol{\epsilon}_{\boldsymbol{D}_{\boldsymbol{n}}}$ and ϵ_{D_d} according to Eq. (50) and Eq. (51), respectively. Check if $max\left(\boldsymbol{\epsilon}_{\boldsymbol{D}_{\boldsymbol{n}}},\epsilon_{D_d}\right) \le \epsilon_K$ (defined as 0.05 here):
(a) True, go to Step 8;
(b) False, go to Step 7.
7. Select the next training point based on the U learning function; then, go to Step 3.
8. Calculate $\boldsymbol{\epsilon}_{\boldsymbol{\mu}_{f''}}$ according to Section 3.2. Check if $\max\left(\boldsymbol{\epsilon}_{\boldsymbol{\mu}_{f''}}\right) \le \epsilon_T$ (defined as 0.05 here):
(a) True, go to Step 10;
(b) False, go to Step 9.
9. Increase the IS realizations N_{IS}. This work increases N_{IS} by 10^3 in this step.
10. End the calculation.

The proposed framework adaptively explores the vicinity of the limit state function by active learning, and refines the importance sampling distribution, i.e., optimizing $\boldsymbol{u}$, in each step based on the current Kriging model. As demonstrated in the next section, the proposed framework and stopping criteria can considerably improve both efficiency and accuracy of Bayesian updating. In particular, the sampling efficiency is significantly improved with the implementation of the adaptive importance sampling.

4. Numerical Examples

In this section, three numerical examples with increasing complexities and an engineering application are selected to demonstrate the efficiency and accuracy of the proposed method. The results are compared with other state-of-the-art Bayesian updating methods, and detailed discussions are presented.

4.1 Example One: One-dimensional Illustrative Example

A toy example from [28, 34] is selected as the first example for illustration purposes. The problem involves only one random variable, denoted by X. The prior distribution of X follows a standard normal distribution, and the likelihood function can be expressed by:

$$\mathrm{L}(x) = \frac{1}{0.5\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-2)^2}{0.5^2}} \tag{59}$$

According to Eq. (6), the limit state function is formulated as:

$$g(p, \boldsymbol{x}) = In(p) - In(c) - In(L(x)) \tag{60}$$

where p follows a uniform standard distribution, and c is determined as $\frac{1}{\max(\mathrm{L}(x))} = 0.5\sqrt{2\pi}$.

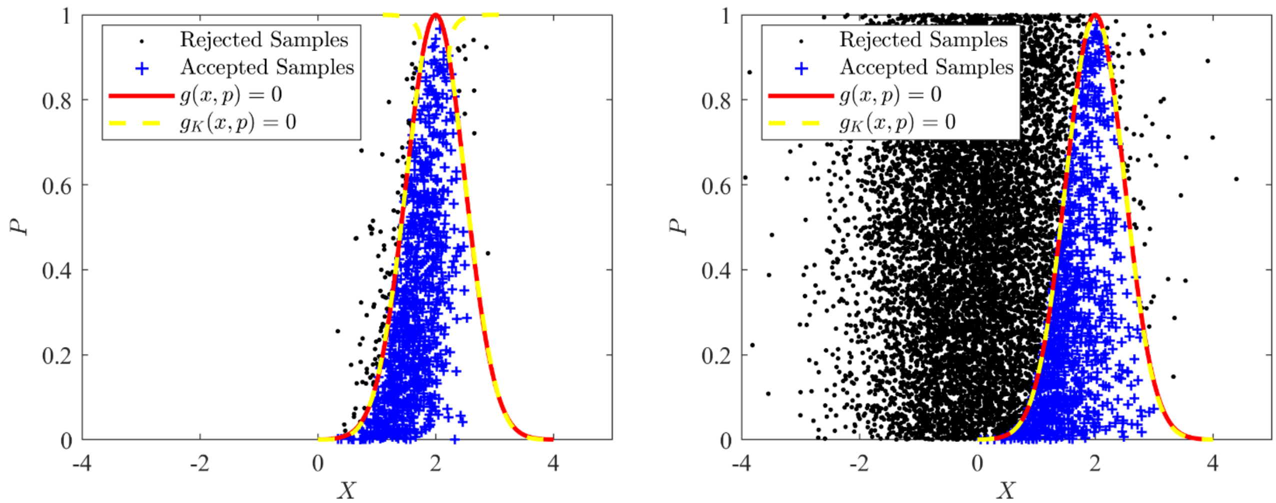
An analytical solution is derived and used as the benchmark. The posterior is calculated as a normal distribution with mean 1.6 and standard deviation $\sqrt{0.2}$. The proposed method, BUAK-AIS, is compared with two different BUS methods: BUS with Monte Carlo Simulation, called BUS-MCS, and BUS with Subset Simulation, called BUS-SuS. Two existing Kriging-based Bayesian updating methods including BUAK-MCS and BUAK-SuS [34] are also compared here. Table 1 lists the number of MCS samples (N_{MCS}), the number of IS samples (N_{IS}), the number of samples in each subset in Subset Simulation (N_{SS}), the required number of calls of the limit state function (N_{Call}), the required number of training points for the Kriging model (N_T) and the estimates of the posterior distribution ($\hat{\mu}$, $\hat{\sigma}$) for these methods.

Table 1 Bayesian updating results of Example One

Method*	N_{MCS} / N_{IS} / N_{SS}	N_{Call} / N_T	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\mu}/\mu$	$\hat{\sigma}/\sigma$
BUS-MCS	1×10^5	1×10^5	1.6000	0.4471	1.0000	0.9996
BUS-SuS	1×10^3	2,543	1.5511	0.4574	0.9695	1.0228
BUAK-MCS	1×10^5	14.15	1.6066	0.4528	1.0041	1.0125
BUAK-SuS	1×10^3	14	1.5560	0.4699	0.9725	1.0508
BUAK-AIS	1×10^3	12.1	1.6003	0.4352	1.0002	0.9731

* Results are averaged over 20 runs.

As Table 1 shows, BUS-MCS can achieve high accuracy for estimating the posterior distribution, with $\hat{\mu}/\mu$=1.0000 and $\hat{\sigma}/\sigma$=0.9996. However, as the acceptance rate is relatively low and about 9%, the required number of samples is considerably larger than the other methods. BUS-SuS can generate around 1,000 accepted samples through two intermediate sets. With 2,543 calls of the limit state function, BUS-SuS can also provide accurate results with $\hat{\mu}/\mu$=0.9695 and $\hat{\sigma}/\sigma$=1.0228. Using the Kriging surrogate model, the required calls of the limit state function can be significantly reduced. With 14.15 calls of the limits state function on average, BUAK-MCS can provide accurate estimates with $\hat{\mu}/\mu$=1.0041 and $\hat{\sigma}/\sigma$=1.0125. Similarly, the required calls for BUAK-SuS are 14 with $\hat{\mu}/\mu$=0.9725 and $\hat{\sigma}/\sigma$=1.0508. It is noted that the proposed method, BUAK-AIS, requires fewer calls of the limit state function compared with BUAK-MCS and BUAK-SuS, while the accuracy of BUAK-AIS ($\hat{\mu}/\mu$=1.0002 and $\hat{\sigma}/\sigma$=0.9731) is higher than the results of BUAK-MCS and BUAK-SuS. The efficiency and accuracy of the proposed method are therefore evident.



(a) The proposed method: BUAK-AIS (b) The existing method: BUAK-MCS

Fig. 2 Illustration of the realizations in Example one

Fig. 2 shows the boundary of the limit state predicted by Kriging and realizations for the proposed method, BUAK-AIS, and the existing method, BUAK-MCS. The Kriging models in both methods achieve high accuracy and the predicted limit state boundary is close to the true limit state boundary in the design space. It is also noted that the acceptance rate for the proposed method BUAK-AIS (87.1%) is considerably higher than the rate for BUAK-MCS (8.8%). Thus, fewer realizations are required by the proposed method to achieve a robust estimation, therefore reducing the computational costs. Moreover, as only the importance area is focused by the proposed method, the number of required calls of the limit state function has reduced from 14.15 to 12.1 on average using the proposed method compared to BUAK-MCS.

4.2 Example Two: Unimodal Distribution Problem

The second example is selected to demonstrate the proposed methods for higher dimensional problems [28, 34, 51]. The prior distribution is constructed as the product of n independent standard normal distributions, denoted as $f'(\boldsymbol{x}) = \prod_{i=1}^{n} \varphi(x_i)$. The likelihood function can be expressed as:

$$\mathrm{L}(\boldsymbol{x}) = \prod_{i=1}^{n} \frac{1}{\sigma_l} \varphi(\frac{x_i - \mu_l}{\sigma_l}) \tag{61}$$

where σ_l is a constant value 0.2, and μ_l can be obtained by:

$$\mu_l = \sqrt{-2(1+\sigma_l{}^2)\mathrm{In}\left[c_E{}^{1/n}\sqrt{2\pi\sqrt{1+\sigma_l{}^2}}\right]} \tag{62}$$

where c_E is the model evidence. Two cases are considered in this paper: case one: $n = 2$ and $c_E = 10^{-4}$ and case two: $n = 10$ and $c_E = 10^{-5}$.

As both the prior and likelihood distributions are normal distributions, an analytical solution can be obtained, which is treated as the benchmark: for case one, the mean and standard deviation of the posterior distribution are 2.659 and 0.1961 and for case two, the mean and standard deviation of the posterior distribution are 0.6542 and 0.1961. The methods mentioned in example one are also applied here (MCS is not applied to case two because of the extremely low failure probability). Table 2 lists the number of MCS samples (N_{MCS}), the number of IS samples (N_{IS}), the number of samples in each subset in Subset Simulation (N_{SS}), the required number of calls of the limit state function (N_{Call}), the required number of training points for the Kriging model (N_T) and the estimates of the posterior distribution ($\hat{\mu}$, $\hat{\sigma}$) for case one. The results for case two are listed in Table 3.

Table 2 Bayesian updating results of Example Two (Case One)

Method*	N_{MCS} / N_{IS} / N_{SS}	N_{Call} / N_T	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\mu}/\mu$	$\hat{\sigma}/\sigma$
BUS-MCS	2×10^7	2×10^7	2.6612	0.1936	1.0008	0.9870
BUS-SuS	5×10^3	28,384	2.6765	0.2061	1.0066	1.0510
BUAK-MCS	2×10^7	31	2.6674	0.1853	1.0032	0.9446
BUAK-SuS	5×10^3	31	2.7310	0.2068	1.0271	1.0544
BUAK-AIS	1×10^3	12.9	2.6670	0.2004	1.0030	1.0226

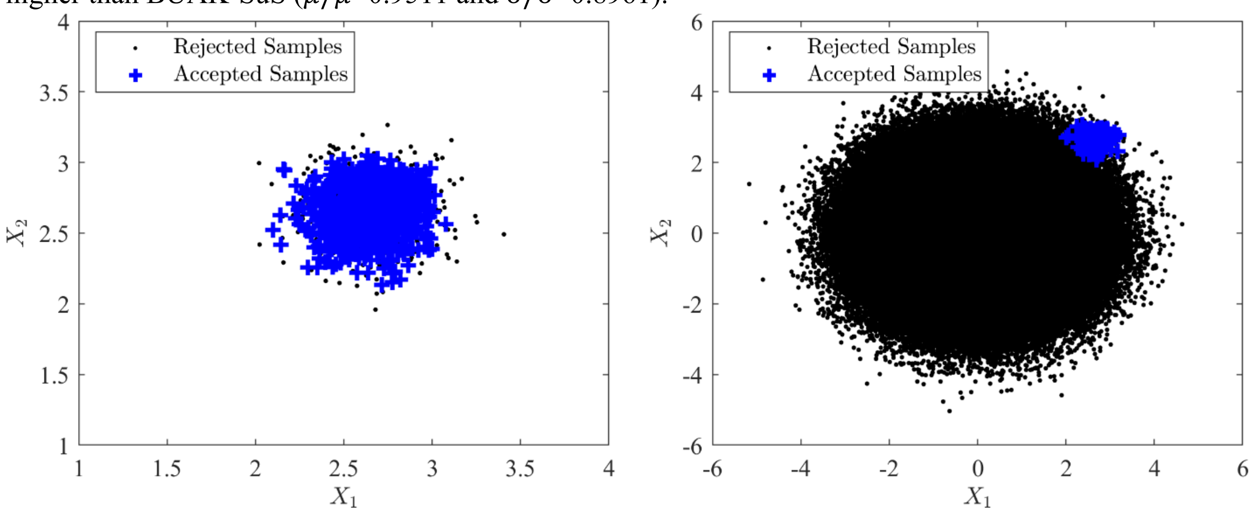
* Results for the proposed method are averaged over 20 runs. Other result can be found in [34].

Table 3 Bayesian updating results of Example Two (Case Two)

Method*	N_{IS} / N_{SS}	N_{Call} / N_T	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\mu}/\mu$	$\hat{\sigma}/\sigma$
BUS-SuS	1×10^4	64,886	0.6778	0.1811	1.0360	0.9236
BUAK-SuS	1×10^4	103	0.6222	0.1751	0.9511	0.8961
BUAK-AIS	1×10^4	90.5	0.6532	0.1965	0.9985	1.0026

* Results for the proposed method are averaged over 20 runs. Other result can be found in [34].

The results of the first case are summarized in Table 2. BUS-MCS and BUAK-SuS can both achieve high accuracy. However, the computational cost of these methods is substantially large. BUS-SuS requires 28,234 calls of the limit state function and BUS-MCS requires 2×10^7 calls on average. With the implementation of the Kriging surrogate model, the required number of calls shows a significant drop. BUAK-MCS and BUAK-SuS both require 31 calls of the limit state function on average. The proposed method, BUAK-AIS, which requires only 12.9 calls, is the most efficient approach. The estimates of the posterior distribution from BUAK-AIS are also more accurate ($\hat{\mu}/\mu$=1.0030 and $\hat{\sigma}/\sigma$=1.0226) compared with the ones from BUAK-MCS ($\hat{\mu}/\mu$=1.0032 and $\hat{\sigma}/\sigma$=0.9446) and BUAK-SuS ($\hat{\mu}/\mu$=1.0271 and $\hat{\sigma}/\sigma$=1.0544). Results in Table 3 for case two also point to the high efficiency and accuracy of the proposed method, BUAK-AIS. The number of required calls of the limit state function is reduced by 13% compared with BUAK-SuS, while the accuracy of the proposed method ($\hat{\mu}/\mu$=0.9985 and $\hat{\sigma}/\sigma$=1.0026) is noticeably higher than BUAK-SuS ($\hat{\mu}/\mu$=0.9511 and $\hat{\sigma}/\sigma$=0.8961).



(a) The proposed method: BUAK-AIS (b) The existing method: BUAK-MCS

Fig. 3 Illustration of the realizations in Example two

Fig. 3 illustrates the distribution of realizations for the proposed method, BUAK-AIS, and the existing method, BUAK-MCS, for case one. The acceptance rate for BUAK-AIS (73%) is substantially higher than the rate for BUAK-MCS (0.001%). Thus, the proposed method can provide higher sampling efficiency compared with the existing method, thus further reducing the computational costs.

4.3 Example Three: A Dynamic Problem

A two degree of freedom structural dynamic problem is selected as the third example [25, 28, 34]. The configuration of the system is shown in Fig. 4. The masses of the first and second story are considered as $m_1 = 16{,}531$ kg and $m_2 = 16{,}131$ kg. The stiffness between stories is modeled as $k_1 = X_1 k_0$ and $k_2 = X_2 k_0$, where X_1 and X_2 are stiffness factors and $k_0 = 29{,}700$ kN/m. The prior distributions of X_1 and X_2 are uncorrelated lognormal distributions with modes 1.3 and 0.8 and standard deviations $\sigma_{X_1} = \sigma_{X_2} = 1$.

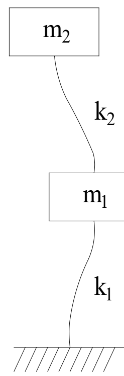


Fig. 4 Illustration of the dynamic system

The first two frequencies are used for updating the distribution, and the observations are $\hat{f}_{r1} = 3.13$ Hz and $\hat{f}_{r2} = 9.83$ Hz. The likelihood function can be expressed as:

$$\mathrm{L}(\boldsymbol{x}) \propto \exp\left(\frac{-\sum_{i=1}^{2} \lambda_i^{2} \left[\frac{f_{ri}(\boldsymbol{x})}{\hat{f}_{ri}}\right]^{2}}{2\sigma_{\varepsilon}^{2}}\right) \tag{63}$$

where $\boldsymbol{x} = [x_1, x_2]$ is a realization of $\boldsymbol{X} = [X_1, X_2]$, $f_{ri}(\cdot)$ is the prediction function for the ith frequency, λ_1 and λ_2 are the means of the prediction error for the first and second frequencies, respectively, and σ_ε is the standard deviation of the prediction error. In this example, $\lambda_1 = \lambda_2 = 1$ and $\sigma_\varepsilon = 1/16$.

The proposed method, BUAK-AIS, is compared with BUS-MCS, BUS-SuS, BUAK-MCS and BUAK-SuS. Table 4 lists the number of MCS samples (N_{MCS}), the number of IS samples (N_{IS}), the number of samples in each subset in Subset Simulation (N_{SS}), the required number of calls of the limit state function (N_{Call}), the required number of training points for the Kriging model (N_T) and the estimated means and standard deviations for the left and right clusters of X_1. Fig. 5 shows the distribution of the accepted samples for the proposed method. Since the analytical solution is not easily obtained in this example, the results from BUS-MCS can be taken as the benchmark.

Table 4 Bayesian updating results of Example Three

Method*	N_{MCS} / N_{IS} / N_{SS}	N_{Call} / N_T	$\hat{\mu}(L)$	$\hat{\sigma}(L)$	$\hat{\mu}(R)$	$\hat{\sigma}(R)$
BUS-MCS	2×10^5	2×10^5	0.502	0.038	1.817	0.141
BUS-SuS	1×10^3	3,674.52	0.505	0.044	1.824	0.137
BUAK-MCS	2×10^5	252.68	0.502	0.038	1.816	0.143
BUAK-SuS	1×10^3	252.68	0.498	0.049	1.829	0.135
BUAK-AIS	10,100	67.8	0.5025	0.0380	1.8144	0.1399

* Results for the proposed method are averaged over 20 runs. Other result can be found in [34].

As Table 4 illustrates, the estimated mean and standard division on the left side using BUS-MCS are $\hat{\mu}(L)$= 0.502 and $\hat{\sigma}(L)$= 0.038, and the ones on the right side are $\hat{\mu}(R)$= 1.817 and $\hat{\sigma}(R)$= 0.141. BUAK-MCS can achieve high accuracy, while reducing the number of calls to 252.68. The acceptance rate of MCS is only about 0.16%. Thus, 2×10^5 realizations are required by MCS to achieve reliable estimation of the posterior distribution. The mean and standard deviation estimated by BUAK-SuS are $\hat{\mu}(L)$=0.498, $\hat{\sigma}(L)$= 0.049, $\hat{\mu}(R)$= 1.829 and $\hat{\sigma}(R)$= 0.135. It is noted that the estimated standard deviation for the left cluster of X_1 by BUAK-SuS (0.049) is considerably higher than that obtained by the MCS (0.038). The proposed method, BUAK-AIS, is the most efficient approach. Only 67.8 calls of the limit state function are required on average. And the results obtained by the proposed method ($\hat{\mu}(L)$=0.5025, $\hat{\sigma}(L)$= 0.0380, $\hat{\mu}(R)$= 1.8144, $\hat{\sigma}(R)$= 0.1399) are highly accurate.

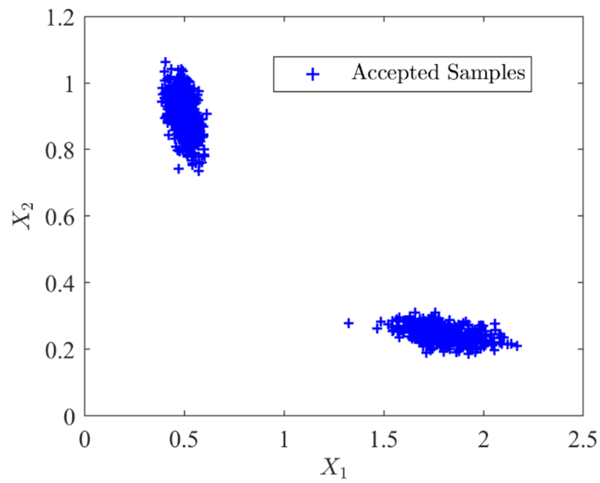


Fig. 5 The accepted samples in BUAK-AIS

4.4 Example Four: Model Updating of a Cable-stayed Bridge during Construction

Cable-stayed bridges have received much attention and been applied widely around world in the last few decades in part due to the advantages they offer in terms of mechanistic performance, construction, maintenance and aesthetic characteristics [52-54]. The cantilever method is the most common approach for the construction of cable-stayed bridges [55]. After the construction of the tower and the first segment of the girder, the general erection procedure of the cantilever method consists of three steps. First, the next segment of girder is installed symmetrically in each side of the tower. Second, the stay cables are installed, and then the cranes move to the end of the constructed girder for the installation of the next segment. These three steps continue until the closure of the girder in the side span or in the main span. Fig. 6 illustrates basic procedures of the cantilever method.

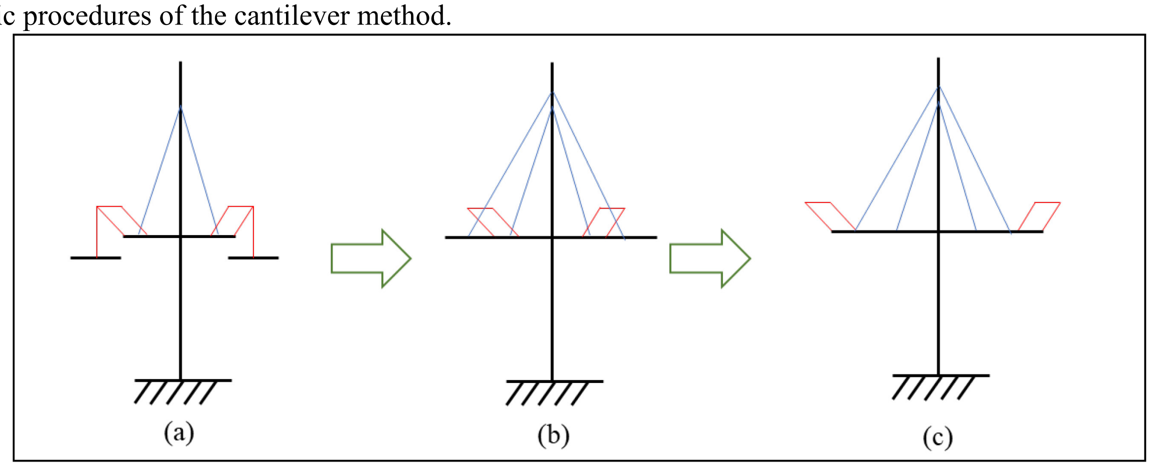


Fig. 6 The illustration of the cantilever construction method: (a) installation of a girder segment, (b) installation of the stay cables, (c) movement of the crane

As a highly redundant structure, the deformations and the internal force distribution of a cable-stayed bridge depends on the cable pretension forces. Many studies have focused on the deterministic optimization of the pretension forces for the stay cables [52-54, 56-59]. However, uncertainties, especially those stemming from the weight of the girder and applied cable forces, may pose challenge to sequential construction as segments have to be aligned very accurately. For instance, when the weight of the girder is underestimated, the desired configuration cannot be achieved with the predefined optimal cable forces. Moreover, the deviation of the applied cable forces can also significantly influence the configuration, yielding a considerably large deflection. Collecting information during the construction process therefore becomes essential for model updating and therefore optimizing the later construction procedures. For example, forces in the cables that have been constructed can be modified in order to adjust for deviations from the design plans. This paper investigates the performance of the proposed method for determining the posterior distribution of key parameters of a 1210-m long-span cable-stayed bridge based on the observations in the construction process.

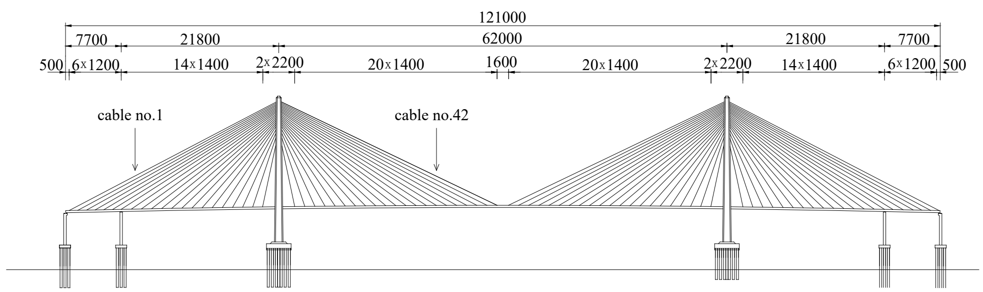


Fig. 7 Span arrangement of the cable-stayed bridge (cm)

The span arrangement of the cable-stayed bridge is shown in Fig. 7. There are 168 cables with a semi-fan design. The main span and side span are 620 m and 320 m long, respectively. The total height of the tower is 203.5 m, and the deck is placed 45 m above the foundation. The construction process is modeled in ANSYS. Note that as the stiffness of the girder is relatively small considering the long span of the bridge, the applied cable forces and the weight of the girder become crucial to the configuration of the bridge during the construction process. Thus, this example only considers the density of the girder and the applied cable force as random variables to investigate the performance of the proposed method. The manufacturing error in the girder welding, the wind-induced vibration and the difference between the temperature inside and outside the girder are not considered here. Table 5 lists parts of the section properties, material properties and the construction load of the bridge. More details of the bridge and the optimization of cable forces can be found in [54].

Table 5 Parameters of the cable-stayed bridge

Variable	Distribution	Mean	Standard deviation
The mass density of girder	Normal	7850 (kg/m^3)	785
The cable pretension forces	Normal	1859.5 (kN)	185.95
The moment of inertia of girder*	Deterministic	3.468 (m^3)	-
The section area of girder	Deterministic	1.1811 (m^2)	-
The moment of inertia of tower (top)*	Deterministic	131.830 (m^3)	-

The section area of tower (top)	Deterministic	21.821 (m^2)	-
The moment of inertia of tower (bottom)*	Deterministic	769.899 (m^3)	-
The section area of tower (bottom)	Deterministic	44.535 (m^2)	-
Elasticity modulus of the steel	Deterministic	2.05×10^{11} (Pa)	-

* The moment of inertia to the lateral direction of the bridge

During the construction of the first segment using the cantilever construction method, the relative deflection of the end of the installed girder can be measured. As the installation of a segment consists of three steps as shown in Fig. 6, there are totally 3 observed deflections. Hypothetical observations of these defections are: o_1=-3.52 cm, o_2=8.24 cm and o_3=-2.46 cm. The errors of the observations are assumed to follow a normal distribution with zero mean and σ_ε=1 mm standard deviation. Thus, the likelihood function can be formulated as:

$$\mathrm{L}(\boldsymbol{x}) \propto \prod_{i=1}^{3} \exp\left(\frac{[o_i - O_i(\boldsymbol{x})]^2}{2{\sigma_\varepsilon}^2}\right) \tag{64}$$

where $\boldsymbol{x} = [x_1, x_2]$ is a realization of $\boldsymbol{X} = [X_1, X_2]$, X_1 represents the mass density of girder, X_2 represents the pretension force for the stay cables of the first segment, o_i ($i = [1,2,3]$) is the observed deflection, and $O_i(\cdot)$ is the predicted deflection from the finite element analysis in ANSYS.

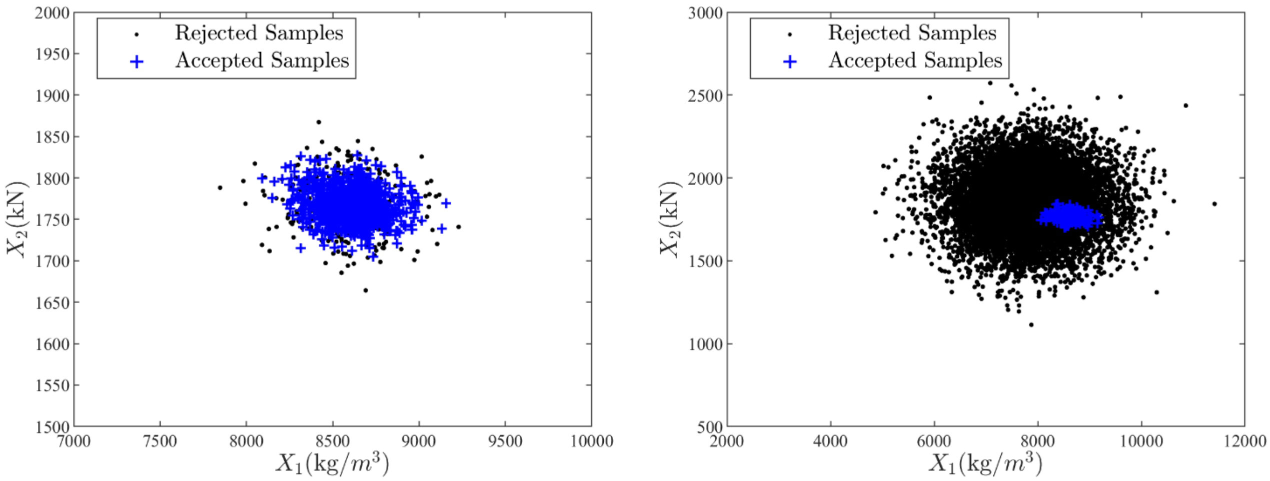
The proposed method, BUAK-AIS, is compared with the existing method BUAK-MCS for illustrating the performance. As the evaluation of the limit state function is time-consuming, BUS with the original limit state function is not applied here. Table 6 lists the number of MCS samples (N_{MCS}), the number of IS samples (N_{IS}), the required number of training points for the Kriging model (N_T), and the estimated mean and standard deviation for the posterior distribution.

Table 6 Bayesian updating results of Example Four

Method*	N_{MCS} / N_{IS}	N_T	$\hat{\mu}(X_1)$	$\hat{\sigma}(X_1)$	$\hat{\mu}(X_2)$	$\hat{\sigma}(X_2)$
BUAK-MCS	1×10^7	35.1	8570.63	219.61	1769.34	25.90
BUAK-AIS	1×10^3	11.5	8571.80	214.08	1769.96	25.24

* Results are averaged over 20 runs.

As Table 6 illustrates, the estimated mean and standard division for the posterior distribution of X_1 with BUAK-MCS are 8570.63 and 219.61, respectively. In addition, the estimated mean and standard division for the posterior distribution of X_2 with BUAK-MCS are 1769.34 and 25.90, respectively. The proposed method BUAK-AIS can achieve very close results, which demonstrates its high accuracy. Moreover, the proposed method can reduce the number of calls from 35.1 to 11.5 on average, showing around 67% improvement in efficiency. Fig. 8 illustrates the distribution of the realizations in BUAK-AIS and BUAK-MCS. The acceptance rate of the proposed method is about 77%, while that for BUAK-MCS is only 2.1%. The sampling efficiency of the proposed method in this practical application can thus be illustrated.



(a) The proposed method: BUAK-AIS (b) The existing method: BUAK-MCS

Fig. 8 Illustration of the realizations in Example four

5. Conclusion

Reformulating the rejection sampling into reliability analysis, Bayesian Updating with Structural reliability methods (BUS) has shown high potential to improve computational efficiency of Bayesian updating. However, the transformed reliability problem may face the challenge of characterizing the probability of a rare event. To address this issue, this paper proposes an efficient Bayesian Updating method with Active learning Kriging-based Adaptive Importance Sampling, called BUAK-AIS. In the proposed framework, the vicinity of the limit state function is adaptively explored by the U learning function. A Gaussian mixture distribution is utilized as the quasi-optimal importance sampling distribution, and the parameters of the Gaussian mixture distribution are optimized in each iteration based on the current Kriging model. The estimate for the first moment of the posterior distribution is discussed, and a stopping criterion is proposed accordingly for a robust estimate of the posterior distribution using importance sampling. A new stopping criterion for the active learning process of Kriging is also developed by quantifying the error in the estimation of the posterior distribution. Three numerical examples and an application regrading model updating of a cable-stayed bridge during the construction process are investigated to examine the performance of the proposed methods. Results indicate that the proposed method, BUAK-AIS, can provide highly accurate estimates of the posterior distribution, while reducing the required calls of expensive-to-evaluate likelihood functions compared with existing approaches. A path for future research is to address limitations of Kriging for high-dimensional problems, which subsequently prevents the application of the proposed method to such cases. As the performance of the proposed method also depends on the quality of the proposal IS distribution, investigating and improving the performance of the cross-entropy-based IS can also be studied in the future. Some or all data, models, or code generated or used during the study are available from the corresponding author upon reasonable request.

CRediT authorship contribution statement

Chaolin Song: Conceptualization, Methodology, Formal Analysis, Writing - original draft. **Zeyu Wang**: Methodology, Validation, Writing - review & editing. **Abdollah Shafieezadeh**: Methodology, Validation, Writing - review & editing. **Rucheng Xiao**: Conceptualization, Writing - review & editing.

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

This research was partly funded by the U.S. National Science Foundation (NSF) through award CMMI-2000156; the Lichtenstein endowment at The Ohio State University; and the China Scholarship Council. Opinions and findings presented are those of the authors and do not necessarily reflect the views of the sponsors.

References

[1] Zamanian S, Hur J, Shafieezadeh A. Significant variables for leakage and collapse of buried concrete sewer pipes: A global sensitivity analysis via Bayesian additive regression trees and Sobol'indices. Structure and infrastructure engineering. 2020:1-13.

[2] Zamanian S, Hur J, Shafieezadeh A. A high-fidelity computational investigation of buried concrete sewer pipes exposed to truckloads and corrosion deterioration. Eng Struct. 2020;221:111043.

[3] Zhang C, Shafieezadeh A. A quantile-based sequential approach to reliability-based design optimization via error-controlled adaptive Kriging with independent constraint boundary sampling. Structural and Multidisciplinary Optimization. 2021;63:2231-52.

[4] Straub D. Value of information analysis with structural reliability methods. Structural Safety. 2014;49:75-85.

[5] Zhang C, Wang Z, Shafieezadeh A. Value of Information Analysis via Active Learning and Knowledge Sharing in Error-Controlled Adaptive Kriging. IEEE Access. 2020;8:51021-34.

[6] Pozzi M, Der Kiureghian A. Assessing the value of information for long-term structural health monitoring. Health monitoring of structural and biological systems 2011: International Society for Optics and Photonics; 2011. p. 79842W.

[7] Song C, Zhang C, Shafieezadeh A, Xiao R. Value of information analysis in non-stationary stochastic decision environments: A reliability-assisted POMDP approach. Reliability Engineering & System Safety. 2022;217:108034.

[8] DiazDelaO F, Garbuno-Inigo A, Au S, Yoshida I. Bayesian updating and model class selection with subset simulation. Computer Methods in Applied Mechanics and Engineering. 2017;317:1102-21.

[9] Feng X, Jimenez R, Zeng P, Senent S. Prediction of time-dependent tunnel convergences using a Bayesian updating approach. Tunnelling and Underground Space Technology. 2019;94:103118.

[10] Walter G, Flapper SD. Condition-based maintenance for complex systems based on current component status and Bayesian updating of component reliability. Reliability Engineering & System Safety. 2017;168:227-39.

[11] Zheng D, Huang J, Li D-Q, Kelly R, Sloan SW. Embankment prediction using testing data and monitored behaviour: A Bayesian updating approach. Computers and Geotechnics. 2018;93:150-62.

[12] Cordeiro SGF, Leonel ED, Beaurepaire P. Quantification of cohesive fracture parameters based on the coupling of Bayesian updating and the boundary element method. Engineering Analysis with Boundary Elements. 2017;74:49-60.

[13] Huang Y, Beck JL, Li H. Bayesian system identification based on hierarchical sparse Bayesian learning and Gibbs sampling with application to structural damage assessment. Computer Methods in Applied Mechanics and Engineering. 2017;318:382-411.

[14] Savvas D, Papaioannou I, Stefanou G. Bayesian identification and model comparison for random property fields derived from material microstructure. Computer Methods in Applied Mechanics and Engineering. 2020;365:113026.

[15] Ni P, Li J, Hao H, Han Q, Du X. Probabilistic model updating via variational Bayesian inference and adaptive Gaussian process modeling. Computer Methods in Applied Mechanics and Engineering. 2021;383:113915.

[16] Kalman RE. A new approach to linear filtering and prediction problems. 1960.

[17] Chatzi EN, Smyth AW, Masri SF. Experimental application of on-line parametric identification for nonlinear hysteretic systems with model uncertainty. Structural Safety. 2010;32:326-37.

[18] Corigliano A, Mariani S. Parameter identification in explicit structural dynamics: performance of the extended Kalman filter. Computer methods in applied mechanics and engineering. 2004;193:3807-35.

[19] Ebrahimian H, Astroza R, Conte JP. Extended Kalman filter for material parameter estimation in nonlinear structural finite element models using direct differentiation method. Earthquake Engineering & Structural Dynamics. 2015;44:1495-522.

[20] Hoshiya M, Saito E. Structural identification by extended Kalman filter. Journal of engineering mechanics. 1984;110:1757-70.

[21] Julier SJ, Uhlmann JK. New extension of the Kalman filter to nonlinear systems. Signal processing, sensor fusion, and target recognition VI: International Society for Optics and Photonics; 1997. p. 182-93.

[22] Saha N, Roy D. Extended Kalman filters using explicit and derivative-free local linearizations. Applied Mathematical Modelling. 2009;33:2545-63.

[23] Thrun S. Probabilistic robotics. Communications of the ACM. 2002;45:52-7.

[24] Gilks WR, Wild P. Adaptive rejection sampling for Gibbs sampling. Journal of the Royal Statistical Society: Series C (Applied Statistics). 1992;41:337-48.
[25] Beck JL, Au S-K. Bayesian updating of structural models and reliability using Markov chain Monte Carlo simulation. Journal of engineering mechanics. 2002;128:380-91.
[26] Ching J, Chen Y-C. Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging. Journal of engineering mechanics. 2007;133:816-32.
[27] Basheer IA, Hajmeer M. Artificial neural networks: fundamentals, computing, design, and application. Journal of microbiological methods. 2000;43:3-31.
[28] Straub D, Papaioannou I. Bayesian updating with structural reliability methods. Journal of Engineering Mechanics. 2015;141:04014134.
[29] Crestaux T, Le Maıtre O, Martinez J-M. Polynomial chaos expansion for sensitivity analysis. Reliability Engineering & System Safety. 2009;94:1161-72.
[30] Sudret B. Global sensitivity analysis using polynomial chaos expansions. Reliability engineering & system safety. 2008;93:964-79.
[31] Bourinet J-M. Rare-event probability estimation with adaptive support vector regression surrogates. Reliability Engineering & System Safety. 2016;150:210-21.
[32] Dai H, Zhang H, Wang W, Xue G. Structural reliability assessment by local approximation of limit state functions using adaptive Markov chain simulation and support vector regression. Computer-Aided Civil and Infrastructure Engineering. 2012;27:676-86.
[33] Cressie N. The origins of kriging. Mathematical geology. 1990;22:239-52.
[34] Wang Z, Shafieezadeh A. Highly efficient Bayesian updating using metamodels: An adaptive Kriging-based approach. Structural Safety. 2020;84:101915.
[35] Lophaven SN, Nielsen HB, Søndergaard J. DACE: a Matlab kriging toolbox: Citeseer; 2002.
[36] Betz W, Papaioannou I, Straub D. Adaptive variant of the BUS approach to Bayesian updating. EURODYN 20142014.
[37] Betz W, Papaioannou I, Beck JL, Straub D. Bayesian inference with subset simulation: strategies and improvements. Computer Methods in Applied Mechanics and Engineering. 2018;331:72-93.
[38] Kaymaz I. Application of kriging method to structural reliability problems. Structural Safety. 2005;27:133-51.
[39] Bichon BJ, Eldred MS, Swiler LP, Mahadevan S, McFarland JM. Efficient global reliability analysis for nonlinear implicit performance functions. AIAA journal. 2008;46:2459-68.
[40] Echard B, Gayton N, Lemaire M. AK-MCS: an active learning reliability method combining Kriging and Monte Carlo simulation. Structural Safety. 2011;33:145-54.
[41] Lv Z, Lu Z, Wang P. A new learning function for Kriging and its applications to solve reliability problems in engineering. Computers & Mathematics with Applications. 2015;70:1182-97.
[42] Sun Z, Wang J, Li R, Tong C. LIF: A new Kriging based learning function and its application to structural reliability analysis. Reliability Engineering & System Safety. 2017;157:152-65.
[43] Zhang X, Wang L, Sørensen JD. REIF: A novel active-learning function toward adaptive Kriging surrogate models for structural reliability analysis. Reliability Engineering & System Safety. 2019;185:440-54.
[44] Wang Z, Shafieezadeh A. REAK: Reliability analysis through Error rate-based Adaptive Kriging. Reliability Engineering & System Safety. 2019;182:33-45.
[45] Wang Z, Shafieezadeh A. ESC: an efficient error-based stopping criterion for kriging-based reliability analysis methods. Structural and Multidisciplinary Optimization. 2019;59:1621-37.
[46] Wang Z, Shafieezadeh A. On confidence intervals for failure probability estimates in Kriging-based reliability analysis. Reliability Engineering & System Safety. 2020;196:106758.
[47] Zhang C, Wang Z, Shafieezadeh A. Error quantification and control for adaptive kriging-based reliability updating with equality information. Reliability Engineering & System Safety. 2021;207:107323.
[48] Marelli S, Sudret B. UQLab: A framework for uncertainty quantification in Matlab. Vulnerability, uncertainty, and risk: quantification, mitigation, and management2014. p. 2554-63.
[49] Dubourg V, Sudret B, Deheeger F. Metamodel-based importance sampling for structural reliability analysis. Probabilistic Engineering Mechanics. 2013;33:47-57.
[50] Kurtz N, Song J. Cross-entropy-based adaptive importance sampling using Gaussian mixture. Structural Safety. 2013;42:35-44.
[51] Giovanis DG, Papaioannou I, Straub D, Papadopoulos V. Bayesian updating with subset simulation using artificial neural networks. Computer Methods in Applied Mechanics and Engineering. 2017;319:124-45.
[52] Martins AMB, Simãµes LSMC, Negrã£O JOHJO. Optimum design of concrete cable-stayed bridges. International Journal for Computational Methods in Engineering Science & Mechanics. 2016;48:772-91.

[53] Martins AMB, Simões LMC, Negrão JHJO. Optimization of cable forces on concrete cable-stayed bridges including geometrical nonlinearities. Comput Struct. 2015;155:18-27.
[54] Song C, Xiao R, Sun B. Optimization of cable pre-tension forces in long-span cable-stayed bridges considering the counterweight. Eng Struct. 2018;172:919-28.
[55] Wang P-H, Tang T-Y, Zheng H-N. Analysis of cable-stayed bridges during construction by cantilever methods. Comput Struct. 2004;82:329-46.
[56] Hassan MM, Nassef AO, Damatty AAE. Determination of optimum post-tensioning cable forces of cable-stayed bridges. Eng Struct. 2012;44:248-59.
[57] Lute V, Upadhyay A, Singh KK. Computationally efficient analysis of cable-stayed bridge for GA-based optimization. Engineering Applications of Artificial Intelligence. 2009;22:750-8.
[58] Martins AMB, Simões LMC, Negrão JHJO. Cable stretching force optimization of concrete cable-stayed bridges including construction stages and time-dependent effects. Structural & Multidisciplinary Optimization. 2015;51:757-72.
[59] Negrão JHO, Simões LMC. Optimization of cable-stayed bridges with three-dimensional modelling. Comput Struct. 1997;64:741-58.