

On the Confidence Intervals for Failure Probability Estimates in Kriging-Based Reliability Analysis

Zeyu Wang and Abdollah Shafieezadeh
The Ohio State University, Columbus, OH, 43202, United States

Abstract

Despite recent advancements in adaptive Kriging-based reliability analysis for complex limit states, estimation of the accuracy of extant techniques when the true failure probability is unknown remains an important challenge. The present study addresses this gap by developing analytical confidence intervals (CIs) for failure probability estimates. This is facilitated here by leveraging statistical properties of Poisson Binomial distribution for the expected number of failure points in the set of candidate design samples in adaptive Kriging as well as Lindeberg's condition for central limit theorem. Concerning computational demands involved in the computation of CIs, a simpler case where Kriging correlations are neglected is also derived. The performance of the proposed CIs is subsequently analyzed for five examples with different and varying complexities. Results indicate that the proposed CI with correlations considered offers the most accurate intervals. Additionally, whereas the CI estimated without Kriging correlation is not entirely satisfactory at early-stages of adaptive reliability analysis, it converges to accurate bounds at later stages. The proposed CIs can be used to derive efficient stopping criteria, optimal learning strategies and derive efficient solutions for high-dimensional problems.

Key words: *Reliability analysis; Risk analysis; Surrogate models; Kriging; Poisson binomial distribution; Accuracy measure; Confidence interval*

1. Introduction

Real-world phenomena either natural or engineered are accompanied with various types and extents of uncertainties. Characterization and propagation of uncertainties in the analysis of these phenomena is a challenge and a key research area in many fields of science and engineering. For example, to assist with treatment decisions for incidental intracranial aneurysm, the risk of aneurysm rupture has been quantified via computational hemodynamic simulations [1]. In the process of design and manufacturing of various industrial products such as rockets and their propulsion systems, satellites and unmanned aerial vehicles, it is necessary to consider extreme requirements of mission success under uncertain conditions [2]. Uncertainties are ubiquitous in the fields of ecological risk assessment, exposure assessment, occupational health and safety, and security and defence [3]. Analysis of these phenomena to characterize potential response modes is often facilitated by establishing quantifiable criteria, here referred to as performance (also called limit-state) functions, which describe the onset of a particular or a set of response modes [4]. Using these performance functions in conjunction with characterization and propagation of involved uncertainties, one can determine the probability of occurrence of response modes of interest. The estimation of these probabilities is facilitated by reliability analysis methods. In the analysis of reliability, the failure probability, P_f , can be calculated as:

$$P_f = \int_{g(\mathbf{x}) \leq 0} \rho(\mathbf{x}) d\mathbf{x} \quad (1)$$

where $\mathbf{x}$ is the vector of random variables, $g(\mathbf{x})$ is the performance function, $\rho(\mathbf{x})$ is the joint probability density function (PDF) of $\mathbf{x}$. In many cases, evaluation of $g(\mathbf{x})$ requires analysis of sophisticated, computationally demanding numerical models, making the estimation of P_f a challenging task. This challenge has been tackled through reliability analysis techniques such as the crude Monte-Carlo simulation (MCS) [5], [6], first- or second-order reliability analysis methods [7], [8], importance sampling (IS) [9], and subset simulation (SS) [10], [11]. Apart from techniques above, the state-of-the-art metamodel-based

approaches [12]–[15] can also provide solutions for Eq. (1) due to its computational efficiency and accuracy [4], [12], [16].

Since the early developments of Kriging-based reliability analysis methods such as EGRA (Efficient Global Reliability Analysis) proposed by Bichon et al. [13] and AK-MCS (Adaptive Kriging-based MCS) proposed by Echard et al. [12], a number of more efficient variants have been proposed. These advancements typically contribute to learning functions, sampling strategies, stopping criteria and accuracy measures. For learning functions, aside from the existing EFF and U functions, information entropy-based H function and least improvement function (LIF) have been proposed by Lv et al. [17] and Sun et al. [18], respectively. H learning function aims to seek the training point in the vicinity of the limit state, which follows the same principle as EFF but in a way of information entropy. Moreover, LIF takes advantage of probability density of each point to highlight the training points with large probability densities. Xiao et al. [19] combined two learning functions, Ψ_d and Ψ_σ to pick the best training point defined as the one that is not in the proximity of the existing training points but is close to the limit state, and has high variance. Moreover, a number of parallel training point-enriching strategies, such as k-means clustering and pseudo-Kriging, have been proposed, respectively, by Lelièvre et al. [20] and Wen et al. [21]. Kriging-based simulation techniques with IS [12], [22], [23] and SS [24]–[26] have shown promising results in the analysis of rare-event problems with small number of evaluations to performance function. Furthermore, excluding unimportant candidate design samples in the analysis of failure probabilities has been found to enhance the computational efficiency [21], [27]. For sampling strategies, methods proposed by Echard et al. [9], Balesdent et al. [22] and Dubourg et al. [23] that rely on importance sampling are shown to be computationally efficient for rare events reliability analysis. Furthermore, in *AKOIS*, Zhang et al. [28] optimized the process of searching for importance sampling center, which facilitates the fast identification of multiple Most Probable Points (MPPs). Moreover, Chen et al. [29] proposed a method that substitutes the original sample population with multiple equivalent ones, which can be leveraged to enhance the learning process with sufficiently large sample pool close to the limit state. Other techniques that integrate the subset simulation with Kriging surrogate model can be found in [24], [25] [30]. By not considering the point with small probability density, Wen et al. [21], Yang et al. [27] and Wang and Shafieezadeh [4] proposed sampling strategies that focus on samples with large contributions to probability of failure. Moreover, Kriging surrogate model can be leveraged to improve the computational efficiency of system reliability problems [27], [31], [32]. Gaspar et al. [33] assess the computation efficiency for Kriging-based structural reliability methods. As the advancement of sensors and monitoring techniques for civil infrastructure among other systems is gaining more momentum, it is shown that Kriging surrogate models can facilitate real-time reliability updating [34]. The Kriging model can also be integrated with structural reliability analysis for p-boxes considering both aleatory and epistemic uncertainty [35]. Aside from aforementioned topics, Kriging surrogate models can also be applied to reliability-based design optimization [36]–[38] and other uncertainty quantification (UQ) techniques such as Bayesian updating [39], [40].

A critical consideration in adaptive Kriging-based reliability analysis methods is the definition of the stopping criterion for active learning. Many approaches that use EFF learning functions have prescribed the stopping criterion as $\max(\text{EFF}) \leq 0.001$. Otherwise, they have used the U learning function primarily in the form of $\min(U) \geq 2$ [9], [41], [42]. Similarly, Sun et al. [23] proposed an uncertainty function and Wang et al. [34] defined cumulative confidence level to measure the expected error in wrong sign estimation according to the U learning function. Moreover, Gaspar et al. [44] proposed a stopping criterion based on the uncertainty of the estimated failure probability. Fauriat et al. [31] points out that Kriging models are often sufficiently accurate if 98% of the candidate design samples satisfy $\min(U) \geq 2$. Different from

1 aforementioned approaches, Wang and Shafieezadeh [45] proposed an efficient stopping criterion that
2 analytically derives the maximum error for the estimated failure probability. Following the same path, Jiang
3 et al. [46] proposed the real-time estimation of the maximum error for time-dependent reliability analysis.
4 Additionally, Schöbi et al. [47] proposed using the difference between the failure probabilities estimated in
5 current and last iterations in the stopping criterion. However, Wang et al. [48] pointed out that these
6 accuracy measures can be further improved and they derived the CI of the estimated failure probability via
7 probabilistic classification-based MCS. In their approach, as the probability distribution of estimated failure
8 probability was unknown, Chebyshev's inequality was applied to estimate the corresponding CI. The use
9 of Chebyshev's inequality in this approach, however, can lead to over conservative estimates of CI for
10 failure probability estimates.

11 In this paper, CIs for failure probability estimates are derived within the framework of adaptive Kriging-
12 based reliability analysis, once with Kriging correlations neglected and another time considering such
13 correlations. Statistical properties of Poisson Binomial distribution (PBD) and Central Limit Theorem (CLT)
14 are leveraged here to facilitate the derivations. The proposed CIs considering Kriging correlation are
15 expected to be tighter than those in [48], as the use of Chebyshev's inequality in that approach is based on
16 the assumption that the probabilistic distribution of the failure probability is unknown. The performance of
17 the proposed CIs are examined for five numerical examples, and results are compared to the approach based
18 on Chebyshev's inequality proposed by Wang et al. in [48]. The organization of this paper is as follows.
19 Section 2 introduces the Kriging model, the probabilistic classification-based MCS and the CIs based on
20 Chebyshev's inequality. In Section 3, the CIs with and without considering Kriging correlation are derived.
21 Section 4 explores the performance of the proposed CIs for five examples with different and varying
22 complexities. Section 5 draws the conclusions of this study.

24 **2. Review of Kriging**

25 **2.1 Kriging model**

26 Kriging, also known as the Gaussian process regression, has been widely used for deterministic computer-
27 based experiment. In this section, Kriging model with correlated outputs are briefly summarized. The model
28 for Kriging $\hat{g}(\mathbf{x})$ can be presented as:

$$29 \quad \hat{g}(\mathbf{x}) = F(\mathbf{x}, \boldsymbol{\beta}) + Z(\mathbf{x}) = \mathbf{f}^T(\mathbf{x})\boldsymbol{\beta} + Z(\mathbf{x}), \quad (2)$$

30 where $F(\mathbf{x}, \boldsymbol{\beta})$ is the regression base representing the Kriging trend, which can be a constant or a
31 polynomial. $g(\mathbf{x})$ can be seen as the realization of the random process defined in Eq.(2). $\mathbf{f}(\mathbf{x})$ is the Kriging
32 basis and $\boldsymbol{\beta}$ is the regression coefficients. $\mathbf{f}^T(\mathbf{x})\boldsymbol{\beta}$ usually have ordinary (β_0), linear ($\beta_0 + \sum_{n=1}^N \beta_n x_n$) or
33 quadratic ($\beta_0 + \sum_{n=1}^N \beta_n x_n + \sum_{n=1}^N \sum_{k=1}^N \beta_{nk} x_n x_k$) forms, whereas n is the dimension of the random input
34 vector, $\mathbf{x}$. The ordinary Kriging model is used in this paper and it is assumed that the data used in training
35 the Kriging model are not noisy. $Z(\mathbf{x})$ is the Kriging interpolation following a stationary Gaussian process
36 with zero mean and a covariance matrix between two points, $\mathbf{x}_i$ and $\mathbf{x}_j$, as defined below:

$$37 \quad \text{COV}(Z(\mathbf{x}_i), Z(\mathbf{x}_j)) = \sigma^2 R(\mathbf{x}_i, \mathbf{x}_j; \boldsymbol{\theta}) \quad (3)$$

38 where σ^2 is the process variance or the generalized mean square error from the regression part, based on
39 the best linear unbiased predictor. $R(\mathbf{x}_i, \mathbf{x}_j; \boldsymbol{\theta})$ is the correlation function or the kernel function,
40 representing the correlation function of the process with hyper-parameter $\boldsymbol{\theta}$. The correlation functions
41 implemented in Kriging include linear, exponential, Gaussian, and Matérn models, among others [49]. In
42 this paper, the Gaussian kernel function is implemented based on the assumption that the correlations
43 between samples are smooth:

$$44 \quad R(\mathbf{x}_i, \mathbf{x}_j; \boldsymbol{\theta}) = \prod_{n=1}^N \exp\left(-\theta^{\{n\}} \left(x_i^{\{n\}} - x_j^{\{n\}}\right)^2\right) \quad (4)$$

where N is the dimension of the random input vector, $\mathbf{x}_i$ or $\mathbf{x}_j$ and $x_i^{\{n\}}$ and $\theta^{\{n\}}$ are the elements of $\mathbf{x}_i$ and $\boldsymbol{\theta}$ in the n^{th} dimension, respectively. The hyper-parameter $\boldsymbol{\theta}$ can be estimated via maximum likelihood estimation (MLE) or cross-validation [49]. $\boldsymbol{\theta}$ has a significant impact on Kriging performance [21], [48], [50]. The MLE can be represented as:

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \left(|\mathbf{R}(\mathbf{x}_i, \mathbf{x}_j; \boldsymbol{\theta})|^{\frac{1}{m}} \sigma^2 \right), \quad (5)$$

where m is the number of known training points or design-of-experiment (DoE) points. Thus, for a number of DoE points, $S_{DoE} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m]$, with associated responses $\mathbf{Y} = [g(\mathbf{x}_1), g(\mathbf{x}_2), \dots, g(\mathbf{x}_m)]$, the Kriging-estimated response at candidate design points follow a multivariate normal distribution, which can be represented as [48], [51]–[53]:

$$\hat{g}(\mathbf{x}) \sim N(\boldsymbol{\mu}_{\hat{g}}, \boldsymbol{\Sigma}_{\hat{g}}), \quad \mathbf{x} \in S, \quad (6)$$

where S denotes the set of candidate design samples and N_{MCS} is the size of S . Therefore, the mean values are represented in a matrix form:

$$\boldsymbol{\mu}_{\hat{g}} = \mathbf{F}\boldsymbol{\beta} + \mathbf{r}^T \mathbf{R}^{-1}(\mathbf{Y} - \mathbf{F}\boldsymbol{\beta}), \quad (7)$$

and the covariance matrix is:

$$\boldsymbol{\Sigma}_{\hat{g}} = \sigma^2 (\mathbf{R} + \mathbf{u}_{\hat{g}}^T (\mathbf{F}^T \mathbf{R}^{-1} \mathbf{F})^{-1} \mathbf{u}_{\hat{g}} - \mathbf{r}^T \mathbf{R}^{-1} \mathbf{r}), \quad (8)$$

where

$$\begin{aligned} \mathbf{R} &= \left(\mathbf{R}(\mathbf{x}_p, \mathbf{x}_q; \boldsymbol{\theta}) \right)_{N_{MCS} \times N_{MCS}}^T, \quad \mathbf{x}_p, \mathbf{x}_q \in S, \\ \mathbf{r} &= \left(\mathbf{R}(\mathbf{x}_l, \mathbf{x}_p; \boldsymbol{\theta}) \right)_{N_{MCS} \times m}^T, \quad \mathbf{x}_l \in S_{DoE}, \\ \boldsymbol{\mu}_{\hat{g}} &= \mathbf{F}^T \mathbf{R}^{-1} \mathbf{r} - \mathbf{F}^T, \\ \mathbf{F}(\mathbf{x}) &= [f(\mathbf{x}_1), f(\mathbf{x}_2), \dots, f(\mathbf{x}_{N_{MCS}})]^T. \end{aligned} \quad (9)$$

Let $\sigma_{\hat{g}}^2$ denote the diagonal elements of $\boldsymbol{\Sigma}_{\hat{g}}$, thus, the responses of unknown points without considering Kriging correlation can be represented as:

$$\hat{g}(\mathbf{x}) \sim N(\boldsymbol{\mu}_{\hat{g}}, \sigma_{\hat{g}}^2), \quad \mathbf{x} \in S \quad (10)$$

2.2 Probabilistic Classification-based MCS (PC-MCS)

In reliability analysis, the true failure probability, P_f , is computationally unavailable, an estimation based on probabilistic simulation techniques is preferred. Crude MCS with sufficiently large number of samples is often considered as a benchmark. In the Kriging model with MCS, there are two indicator functions: deterministic classification [12], [54] and the probabilistic classification [23], [48]. For deterministic classification, the crude MCS can be represented as:

$$\hat{P}_f^{dc} = \frac{1}{N_{MCS}} \sum_{i=1}^{N_{MCS}} I_{\mu_{\hat{g}}}^{dc}(\mathbf{x}_i), \quad \mathbf{x} \in S \quad (11)$$

where $\hat{P}_f^{dc}$ is the estimated failure probability with deterministic classification on the Kriging model. N_{MCS} is the number of sampling realizations, $\mathbf{x}_i$, from the probabilistic distribution of random variables and $I_{\mu_{\hat{g}}}^{dc}$ is the corresponding indicator for deterministic classifications:

$$I_{\mu_{\hat{g}}}^{dc}(\mathbf{x}_i) = \begin{cases} 1, & \mu_{\hat{g}}(\mathbf{x}_i) \leq 0 \\ 0, & \mu_{\hat{g}}(\mathbf{x}_i) > 0 \end{cases}, \quad \mathbf{x}_i \in S \quad (12)$$

where $\mu_{\hat{g}}(\mathbf{x}_i)$ is the mean value of $\hat{g}(\mathbf{x})$. The coefficient of variation of $\hat{P}_f^{dc}$ can be calculated as:

$$COV_{\hat{P}_f^{dc}} = \sqrt{\frac{1 - \hat{P}_f^{dc}}{N_{MCS} \hat{P}_f^{dc}}} \quad (13)$$

The failure probability based on probabilistic classification can be estimated as:

$$\hat{P}_f^{pc} = E[\tilde{P}_f^{pc}] = \frac{1}{N_{MCS}} E[\mathbb{U}] = \frac{1}{N_{MCS}} \sum_{i=1}^{N_{MCS}} \Phi\left(\frac{-\mu_{\hat{g}}(\mathbf{x}_i)}{\sigma_{\hat{g}}(\mathbf{x}_i)}\right), \quad \mathbf{x}_i \in S \quad (14)$$

$$\mathbb{U} = \sum_{i=1}^{N_{MCS}} I_{\hat{g}}^{pc}(\mathbf{x}_i), \quad \mathbf{x}_i \in S \quad (15)$$

where $\tilde{P}_f^{pc}$ is the augmented stochastic estimator and $\hat{P}_f^{pc}$ is the mean value of $\tilde{P}_f^{pc}$. $E[\cdot]$ is the expectation operator, $\sigma_{\hat{g}}(\mathbf{x}_i)$ denote the standard deviation of $\mathbf{x}_i$ according to Eq. (10) and $I_{\hat{g}}^{pc}(\mathbf{x}_i)$ denotes the probabilistic classification-based indicator:

$$I_{\hat{g}}^{pc}(\mathbf{x}_i) = \begin{cases} 1, & w.p \ \Phi\left(\frac{-\mu_{\hat{g}}(\mathbf{x}_i)}{\sigma_{\hat{g}}(\mathbf{x}_i)}\right) \\ 0, & w.p \ 1 - \Phi\left(\frac{-\mu_{\hat{g}}(\mathbf{x}_i)}{\sigma_{\hat{g}}(\mathbf{x}_i)}\right) \end{cases}, \quad \mathbf{x}_i \in S \quad (16)$$

where $\Phi(x)$ is the cumulative distribution function (CDF) of the univariate standard normal distribution and $\mu_{\hat{g}}(\mathbf{x}_i)$ and $\sigma_{\hat{g}}(\mathbf{x}_i)$ are the mean and standard deviation, respectively, of Kriging predictors. The performance of probability classification-based MCS is equivalent to the deterministic classification-based approach [48]. Note that $\mathbb{U}$ is a random variable, thus, $\tilde{P}_f^{pc}$ is the mean value of $\frac{\mathbb{U}}{N_{MCS}}$ according to Eq. (14).

The purpose of using probability classification-based MCS in this paper is to facilitate the derivation of the CIs for $\tilde{P}_f^{pc}$. The probabilistic classification estimates the failure probability considering uncertainties associated with Kriging-based classification. The coefficient of variation of failure probability estimate can be represented as [23]:

$$COV_{\tilde{P}_f^{pc}} = \frac{1}{\sqrt{N_{MCS}} \tilde{P}_f^{dc}} \sqrt{\frac{\sum_{i=1}^{N_{MCS}} \Phi^2\left(\frac{-\mu_{\hat{g}}(\mathbf{x}_i)}{\sigma_K(\mathbf{x}_i)}\right)}{N_{MCS}} - \tilde{P}_f^{dc^2}}, \quad \mathbf{x}_i \in S \quad (17)$$

2.3 CI Based on Chebyshev's inequality

Recently, Wang et al. [48] derived $\sigma_{\tilde{P}_f^{pc}}^2$ considering Kriging correlation, as shown in Eq. (6). In that study, $\hat{\sigma}_{\tilde{P}_f^{pc}}^2$ was derived as below [48]:

$$\sigma_{\tilde{P}_f^{pc}}^2 = \frac{1}{N_{MCS}^2} \left[\sum_{i=1}^{N_{MCS}} \Phi(U(\mathbf{x}_i))\Phi(-U(\mathbf{x}_i)) + \sum_{i=1, j=1, i \neq j}^{N_{MCS}} [P(\hat{g}(\mathbf{x}_i) \leq 0, \hat{g}(\mathbf{x}_j) \leq 0) - P(\hat{g}(\mathbf{x}_i) \leq 0) \cdot P(\hat{g}(\mathbf{x}_j) \leq 0)] \right] \quad (18)$$

where $\Phi(-U(\mathbf{x}_i))$ is the probability that the sign of the limit state function at $\mathbf{x}_i$ is incorrectly estimated by the Kriging model, and $P(\hat{g}(\mathbf{x}_i) \leq 0, \hat{g}(\mathbf{x}_j) \leq 0)$ is the bivariate normal distribution with the mean and covariance matrix given in Eq. (6). Because the probabilistic distribution of $\tilde{P}_f^{pc}$ was unknown, the Chebyshev's inequality was applied in [48] to find the CI of $\tilde{P}_f^{pc}$, as shown below:

$$P\left(|\tilde{P}_f^{pc} - P_f^{mcs}| < e \hat{\sigma}_{\tilde{P}_f^{pc}}\right) \geq 1 - \frac{1}{e^2}, \quad (19)$$

where $\hat{\sigma}_{\tilde{P}_f^{pc}}$ is the standard deviation of $\tilde{P}_f^{pc}$, and e is a constant. Therefore, per this inequality, the 95% confidence level CI (i.e. $\alpha = 0.05, 1 - \frac{1}{e^2} = 1 - \alpha, e = 4.472$, where α denotes the significance level) for $\tilde{P}_f^{pc}$ can be represented as:

$$\tilde{P}_f^{pc} \in \left[\mu_{\tilde{P}_f^{pc}} - e \hat{\sigma}_{\tilde{P}_f^{pc}}, \mu_{\tilde{P}_f^{pc}} + e \hat{\sigma}_{\tilde{P}_f^{pc}} \right], \quad \alpha = 0.05, e = 4.472. \quad (20)$$

The CI in Eq. (20) is conservatively estimated, because the distribution of $\tilde{P}_f^{pc}$ is treated as an unknown. However, as shown in the next section, $\mathbb{U}$ follows a Poisson binomial distribution if the Kriging correlation is not considered.

3. The proposed CIs for $\tilde{P}_f^{pc}$

3.1 CI for $\tilde{P}_f^{pc}$ without considering Kriging correlation

Let's consider S as the set of candidate design samples. N_{MCS} and N_f denote the total number of samples and the true number of failure points in S , respectively. Thus, P_f^{mcs} can be estimated as:

$$P_f^{mcs} = \frac{N_f}{N_{MCS}}, \quad (21)$$

Similarly, the probability of failure via PC-MCS can be estimated as:

$$\hat{P}_f^{pc} = \frac{\hat{N}_f^{pc}}{N_{MCS}}, \quad (22)$$

where $\hat{N}_f^{pc}$ is the expected number of failure points in S . In this approach, for each candidate design sample, $\mathbf{x}_i$, the outcome of the probabilistic classification-based indicator function, $I_{\hat{g}}^{pc}(\mathbf{x}_i)$, defined in Eq. (16), follows a Bernoulli distribution:

$$I_{\hat{g}}^{pc}(\mathbf{x}_i) \sim B(\mu_b(\mathbf{x}_i)), \mathbf{x}_i \in S, \quad (23)$$

where B is the Bernoulli distribution, $\mu_b(\mathbf{x}_i)$ is the Bernoulli mean with $\mu_b(\mathbf{x}_i) = \Phi\left(\frac{-\mu_{\hat{g}}(\mathbf{x}_i)}{\sigma_{\hat{g}}(\mathbf{x}_i)}\right)$ and σ_b^2 is the variance of the Bernoulli distribution, $\sigma_b^2 = \mu_b(\mathbf{x}_i)(1 - \mu_b(\mathbf{x}_i))$. It can be inferred that $\hat{N}_f^{pc}$ is the expected value of $\mathbb{U}$ in Eq. (14), where each $I_{\hat{g}}^{pc}(\mathbf{x}_i)$ follows a Bernoulli distribution. In probability theory and statistics, a random variable, PB , is considered to follow Poisson Binomial distribution if it can be represented as [55]:

$$PB = \sum_{i=1}^{N_B} B_i \quad (24)$$

where B_i s denote random variables with mutually independent Bernoulli distribution. Note that B_i s are not identically distributed. Based on this definition, it follows that $\mathbb{U}$ has a Poisson binomial distribution, if the Kriging correlations are not considered. Subsequently, $\mathbb{U}$ can be defined as:

$$\mathbb{U} \sim PB(\mu_{\mathbb{U}}, \sigma_{\mathbb{U}}^2), \mathbf{x}_i \in S, \quad (25)$$

where $\mu_{\mathbb{U}}$ and $\sigma_{\mathbb{U}}^2$ are the mean value and variance of $\mathbb{U}$. According to the probabilistic properties of PBD, $\mu_{\mathbb{U}} = \sum_{i=1}^{N_{MCS}} \mu_{P_f}(\mathbf{x}_i)$ and $\sigma_{\mathbb{U}}^2 = \sum_{i=1}^{N_{MCS}} \mu_b(\mathbf{x}_i)(1 - \mu_b(\mathbf{x}_i))$. Note that $\hat{N}_f^{pc} = \mu_{\mathbb{U}}$. Consequently, the CI for $\mathbb{U}$ with confidence level α can be calculated as:

$$\mathbb{U} \in \left(\boldsymbol{\theta}_{\mathbb{U}}^{-1}\left(\frac{\alpha}{2}\right), \boldsymbol{\theta}_{\mathbb{U}}^{-1}\left(1 - \frac{\alpha}{2}\right) \right), \quad (26)$$

where $\boldsymbol{\theta}_{\mathbb{U}}^{-1}(\cdot)$ is the inverse CDF of PBD with mean $\mu_{\mathbb{U}}$ and variance $\sigma_{\mathbb{U}}^2$. α is the confidence level (e.g. $\alpha = 0.05$). Analytical solutions for the above CIs are typically not available. Therefore, numerical approaches or approximate analytical methods can be pursued. Sampling techniques can be used to numerically determine the inverse CDF of PBDs. In most cases, N_{MCS} is sufficiently large so that the CI of $\mathbb{U}$ can be approximately determined using the Central Limit Theorem (CLT) [45]. As Lindeberg's condition for the central limit theorem for the sum of independent, not identically distributed random variables [56] is satisfied for sufficiently large N_{MCS} :

$$\lim_{N_{MCS} \rightarrow \infty} \left(\max_{i=1, \dots, N_{MCS}} \frac{\text{Var}[I_{\hat{g}}^{pc}(\mathbf{x}_i)]}{\text{Var}[\mathbb{U}]} \right) = 0, \quad \mathbf{x}_i \in S \quad (27)$$

it can be shown that $\mathbb{U}$ in distribution converges to a normal distribution:

$$\mathbb{U} \sim N(\mu_{\mathbb{U}}, \sigma_{\mathbb{U}}^2), \quad \mathbf{x}_i \in S. \quad (28)$$

The CI of $\mathbb{U}$ can then be obtained as:

$$\mathbb{U} \in [\mu_{\mathbb{U}} - \gamma_{ci}\sigma_{\mathbb{U}}, \mu_{\mathbb{U}} + \gamma_{ci}\sigma_{\mathbb{U}}], \quad \mathbf{x} \in S, \quad (29)$$

where $\gamma_{ci} = 1.96$ for the confidence level $\alpha = 0.05$. As N_{MCS} is large in Kriging-based reliability analysis problems, the above confidence bounds for $\mathbb{U}$ are accurate. However, the CLT cannot be appropriately applied in cases where N_{MCS} is small. According to Le cam's theorem [45], [55], [57], the CI of $\mathbb{U}$ without considering Kriging correlation can be approximately obtained using Poisson distribution:

$$\sum_{k=0}^{\infty} \left| Pr(\mathbb{U} = k) - \frac{\mu_{\mathbb{U}}^k e^{-\mu_{\mathbb{U}}}}{k!} \right| < 2 \sum_{i=1}^{N_{MCS}} \left(\Phi \left(\frac{-\mu_{\hat{g}}(\mathbf{x}_i)}{\sigma_{\hat{g}}(\mathbf{x}_i)} \right) \right)^2, \quad \mathbf{x}_i \in S \quad (30)$$

which indicates that the distribution of $\mathbb{U}$ can be approximately represented as a Poisson distribution:

$$Pr(\mathbb{U} = k) \approx \frac{\mu_{\mathbb{U}}^k e^{-\mu_{\mathbb{U}}}}{k!}, \quad k = 0, 1, \dots, N_{MCS} \quad (31)$$

Therefore, the CI of $\mathbb{U}$ can be determined as:

$$\mathbb{U} \in \left[\Gamma_{\mathbb{U}}^{-1} \left(\frac{\alpha}{2} \right), \Gamma_{\mathbb{U}}^{-1} \left(1 - \frac{\alpha}{2} \right) \right] \quad (32)$$

where $\Gamma_{\mathbb{U}}^{-1}(\cdot)$ is the inverse CDF of the Poisson distribution with both mean and variance equal to $\mu_{\mathbb{U}}$ defined in Eq. (25), and α is the confidence level. According to Eq. (22), the CI of $\tilde{P}_f^{pc}$ can be obtained by:

$$\tilde{P}_f^{pc} \in \left[\frac{\mu_{\mathbb{U}} - \gamma_{ci}\sigma_{\mathbb{U}}}{N_{MCS}}, \frac{\mu_{\mathbb{U}} + \gamma_{ci}\sigma_{\mathbb{U}}}{N_{MCS}} \right], \quad \alpha = 0.05, \gamma_{ci} = 1.96. \quad (33)$$

For Kriging-based reliability analysis methods, there is a prior assumption that all the responses for training and testing points are mutually and normally (correlated) distributed. There are two groups of uncertainty in the stochastic estimator. The first group can be defined as the one that the training samples are not sufficient but the Gaussian Process is satisfied and the second group stems from the fact the prior assumption for Gaussian process is not satisfied. For the first case, the difference between g and $\hat{g}$ already exists in the stochastic estimator of $I_{\hat{g}}^{pc}(\mathbf{x}_i)$. The stochastic indicator takes the value of 1 or 0 according to the probability that the sample is classified as failure or safe. The Confidence interval are derived in our manuscript based on the first case. For the second group of uncertainty that is related the GP uncertainty, it is not investigated in this manuscript. This GP uncertainty commonly exists in many Kriging-based reliability analysis algorithm.

3.2 CI for $\tilde{P}_f^{pc}$ considering Kriging correlation

The CI derived above is based on the assumption that responses from the Kriging model follow independent normal distributions. However, these responses follow a multivariate normal distribution, as stated in Section 2.1. Let $\mathbb{U}^c$ denote $\sum_{i=1}^{N_{MCS}} I_{\hat{g}}^{pc}(\mathbf{x}_i)$, $\mathbf{x}_i \in S$, where $I_{\hat{g}}^{pc}(\mathbf{x}_i)$ are mutually correlated indicators in this subsection. The covariance matrix in Eq. (6) for the responses of the Kriging model can be represented as:

$$\mathbf{\Sigma}_{\hat{g}} = \begin{bmatrix} \mathbf{\Sigma}_{1,1} = \sigma_{\hat{g}}^2(\mathbf{x}_1) & \mathbf{\Sigma}_{1,2} & \dots & \mathbf{\Sigma}_{1,N_{MCS}} \\ \mathbf{\Sigma}_{2,1} & \mathbf{\Sigma}_{2,2} = \sigma_{\hat{g}}^2(\mathbf{x}_2) & \dots & \mathbf{\Sigma}_{2,N_{MCS}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Sigma}_{N_{MCS},1} & \mathbf{\Sigma}_{N_{MCS},2} & \dots & \mathbf{\Sigma}_{N_{MCS},N_{MCS}} = \sigma_{\hat{g}}^2(\mathbf{x}_{N_{MCS}}) \end{bmatrix}, \quad (34)$$

where $\mathbf{x}_i$ is a candidate sample from $S = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{N_{MCS}}]$ whose true value of the performance function is not known, $\sigma_{\hat{g}}^2(\mathbf{x}_i)$ is the variance defined in Eq. (10), and $\rho_{i,j}$ is the correlation coefficient between any two Kriging responses, $\hat{g}(\mathbf{x}_i)$ and $\hat{g}(\mathbf{x}_j)$. According to the definition of Pearson's correlation coefficient and Eq. (34), the correlation $\rho_{i,j}$ between $\hat{g}(\mathbf{x}_i)$ and $\hat{g}(\mathbf{x}_j)$ can be calculated as follows:

$$\rho_{i,j} = \frac{\mathbf{\Sigma}_{i,j}}{\sigma_{\hat{g}}(\mathbf{x}_i)\sigma_{\hat{g}}(\mathbf{x}_j)}. \quad (35)$$

It should be noted that the true correlation form between $\hat{g}(\mathbf{x}_i)$ and $\hat{g}(\mathbf{x}_j)$ can have a nonlinear form. Here, the use of Pearson's correlation model have an underlying assumption that this form is linear. While as later presented, this approach provides satisfactory results, further investigations are needed to examine this assumption. Moreover, let $P_{i,j}^B$ denote the probability that two correlated Bernoulli random variables $I_{\hat{g}}^{pc}(\mathbf{x}_i)$ and $I_{\hat{g}}^{pc}(\mathbf{x}_j)$ are both estimated as failure (i.e., $I_{\hat{g}}^{pc}(\mathbf{x}_i) = 1$ and $I_{\hat{g}}^{pc}(\mathbf{x}_j) = 1$). Subsequently, $P_{i,j}^B$ can be calculated as:

$$P_{i,j}^B = \int_{-\infty}^0 \int_{-\infty}^0 \psi_{i,j}([x, y]; \boldsymbol{\mu}_{\hat{g}}, \mathbf{\Sigma}_{\hat{g}}) dx dy, \quad (36)$$

where $\psi_{i,j}(x, y)$ is the PDF of the multivariate normal distribution with mean $\boldsymbol{\mu}_{\hat{g}} = [\mu_{\hat{g}}(\mathbf{x}_i) \ \mu_{\hat{g}}(\mathbf{x}_j)]$ and the covariance matrix $\mathbf{\Sigma}_{\hat{g}} = \begin{bmatrix} \sigma_{\hat{g}}^2(\mathbf{x}_i) & \rho_{i,j}\sigma_{\hat{g}}(\mathbf{x}_i)\sigma_{\hat{g}}(\mathbf{x}_j) \\ \rho_{i,j}\sigma_{\hat{g}}(\mathbf{x}_i)\sigma_{\hat{g}}(\mathbf{x}_j) & \sigma_{\hat{g}}^2(\mathbf{x}_j) \end{bmatrix}$ in Eq. (6). The covariance between two correlated Bernoulli random variables can be represented as:

$$COV(I_{\hat{g}}^{pc}(\mathbf{x}_i), I_{\hat{g}}^{pc}(\mathbf{x}_j)) = \rho_{i,j}^b \cdot \sigma_b(\mathbf{x}_i) \cdot \sigma_b(\mathbf{x}_j), \quad (37)$$

where $\rho_{i,j}^b \in [-1, 1]$ is the unknown correlation coefficient of two correlated Bernoulli distributions, $I_g(\mathbf{x}_i)$ and $I_g(\mathbf{x}_j)$. $\sigma_b(\mathbf{x}_i)$ and $\sigma_b(\mathbf{x}_j)$ are the standard deviations ($\sigma_b(\mathbf{x}_i) = \sqrt{\mu_b(\mathbf{x}_i)(1 - \mu_b(\mathbf{x}_i))}$). When $\rho_{i,j}^b = 0$, it means that the failure probabilities of two points, $\mathbf{x}_i$ and $\mathbf{x}_j$, are uncorrelated, which can further infer that $\rho_{i,j} = 0$ (i.e. $P_{i,j}^B = P(\hat{g}(\mathbf{x}_i) \leq 0, \hat{g}(\mathbf{x}_j) \leq 0) = P(\hat{g}(\mathbf{x}_i) \leq 0) \cdot P(\hat{g}(\mathbf{x}_j) \leq 0)$). When $\rho_{i,j}^b = 1$ or -1 , it means that the two responses are positively or negatively, linearly correlated. It also means that $\rho_{i,j} = 1$ or -1 . Because the correlated Bernoulli distribution in Eq. (36) is actually a linear integration form of bivariate normal distribution, the covariance matrix for the points with Bernoulli distributions can be expressed as:

$$\mathbf{\Sigma}_{pb} = \begin{bmatrix} \sigma_b^2(\mathbf{x}_1) & \rho_{1,2}\sigma_b(\mathbf{x}_1)\sigma_b(\mathbf{x}_2) & \dots & \rho_{1,N_{MCS}}\sigma_b(\mathbf{x}_1)\sigma_b(\mathbf{x}_{N_{MCS}}) \\ \rho_{2,1}\sigma_b(\mathbf{x}_2)\sigma_b(\mathbf{x}_1) & \sigma_b^2(\mathbf{x}_2) & \dots & \rho_{2,N_{MCS}}\sigma_b(\mathbf{x}_2)\sigma_b(\mathbf{x}_{N_{MCS}}) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{N_{MCS},1}\sigma_b(\mathbf{x}_{N_{MCS}})\sigma_b(\mathbf{x}_1) & \rho_{N_{MCS},2}\sigma_b(\mathbf{x}_{N_{MCS}})\sigma_b(\mathbf{x}_2) & \dots & \sigma_b^2(\mathbf{x}_{N_{MCS}}) \end{bmatrix}, \quad (38)$$

where $\sigma_b^2(\mathbf{x}_i)$ is the Bernoulli variance, and can be calculated as:

$$\sigma_b^2(\mathbf{x}_i) = \Phi\left(\frac{-\mu_{\hat{g}}(\mathbf{x}_i)}{\sigma_{\hat{g}}(\mathbf{x}_i)}\right) \cdot \Phi\left(\frac{\mu_{\hat{g}}(\mathbf{x}_i)}{\sigma_{\hat{g}}(\mathbf{x}_i)}\right), \mathbf{x}_i \in S. \quad (39)$$

Construction of the correlation coefficient for every pair of points in S is computationally highly demanding. Therefore, the strategy in [48] for points satisfying $U(\mathbf{x}_i) \leq 2$ can be applied here. Thus, the covariance corresponding to $U(\mathbf{x}_i) \leq 2$ can be represented as:

$$\mathbf{\Sigma}_{pb}^{u \leq 2} = \begin{bmatrix} \sigma_b^2(\mathbf{x}_1^{u \leq 2}) & \rho_{1,2}\sigma_b(\mathbf{x}_1^{u \leq 2})\sigma_b(\mathbf{x}_2^{u \leq 2}) & \dots & \rho_{1,N_{u \leq 2}}\sigma_b(\mathbf{x}_1^{u \leq 2})\sigma_b(\mathbf{x}_{N_{u \leq 2}}^{u \leq 2}) \\ \rho_{2,1}\sigma_b(\mathbf{x}_2^{u \leq 2})\sigma_b(\mathbf{x}_1^{u \leq 2}) & \sigma_b^2(\mathbf{x}_2^{u \leq 2}) & \dots & \rho_{2,N_{u \leq 2}}\sigma_b(\mathbf{x}_2^{u \leq 2})\sigma_b(\mathbf{x}_{N_{u \leq 2}}^{u \leq 2}) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{N_{u \leq 2},1}\sigma_b(\mathbf{x}_{N_{u \leq 2}}^{u \leq 2})\sigma_b(\mathbf{x}_1^{u \leq 2}) & \rho_{N_{u \leq 2},2}\sigma_b(\mathbf{x}_{N_{u \leq 2}}^{u \leq 2})\sigma_b(\mathbf{x}_2^{u \leq 2}) & \dots & \sigma_b^2(\mathbf{x}_{N_{u \leq 2}}^{u \leq 2}) \end{bmatrix}, \quad (40)$$

where $\mathbf{x}_i^{u \leq 2}$ is a point in $S_{U \leq 2}$ with size $N_{u \leq 2}$, which is defined as:

$$S_{U \leq 2} = [\text{all } \mathbf{x}_i \text{ satisfying } U(\mathbf{x}_i^u) \leq 2 \mid \mathbf{x}_i \in S]. \quad (41)$$

Earlier, it was shown that the summation of uncorrelated Bernoulli random variables follow a PBD. However, as the number of candidate design samples is very large, the Central Limit theorem for dependent random variables is applicable here [58]. Thus, $\mathbb{U}^c$ can be approximated as a normal distribution when $N_{MCS} \rightarrow \infty$. Thus, the distribution of $\mathbb{U}^c$ can be represented as:

$$\mathbb{U}^c \sim N(\mu_{\mathbb{U}^c}, \sigma_{\mathbb{U}^c}^2), \quad (42)$$

$$\mu_{\mathbb{U}^c} = \sum_{i=1}^{N_{MCS}} \Phi\left(\frac{-\mu_{\hat{g}}(\mathbf{x}_i)}{\sigma_{\hat{g}}(\mathbf{x}_i)}\right), \mathbf{x}_i \in S, \quad (43)$$

$$\sigma_{\mathbb{U}^c}^2 = \sum_{i=1}^{N_{MCS}} \sum_{j=1}^{N_{MCS}} \mathbf{\Sigma}_{pb_{i,j}} = \sum_{i=1}^{N_{u \leq 2}} \sum_{j=1}^{N_{u \leq 2}} \mathbf{\Sigma}_{pb_{i,j}}^{u \leq 2} + \sum_{i=1}^{N_{u > 2}} \sum_{j=1}^{N_{u > 2}} \mathbf{\Sigma}_{pb_{i,j}}^{u > 2} \approx \sum_{i=1}^{N_{u \leq 2}} \sum_{j=1}^{N_{u \leq 2}} \mathbf{\Sigma}_{pb_{i,j}}^{u \leq 2}, \quad (44)$$

where $\mathbf{\Sigma}_{pb_{i,j}}$ are the elements of the covariance matrix, $\mathbf{\Sigma}_{pb}$, in Eq. (38), $N_{u \leq 2}$ and $N_{u > 2}$ are sizes of $S_{U \leq 2}$ and $S_{U > 2}$ and $\mathbf{\Sigma}_{pb_{i,j}}^{u \leq 2}$ are the elements of the reduced covariance matrix, $\mathbf{\Sigma}_{pb}^{u \leq 2}$, in Eq. (40). Based on the properties of the normal distribution, the CI of $\mathbb{U}^c$ can be calculated as:

$$\mathbb{U}^c \in [\mu_{\mathbb{U}^c} - \gamma_{ci}\sigma_{\mathbb{U}^c}, \mu_{\mathbb{U}^c} + \gamma_{ci}\sigma_{\mathbb{U}^c}], \quad (45)$$

where $\gamma_{ci} = 1.96$ in this paper, which corresponds to $\alpha = 0.05$. However, for small values of N_{MCS} CLT may not guarantee appropriate bounds. To address this issue, the CI of $\mathbb{U}^c$ can be estimated through simulations. First, we generate a group of samples (e.g. 10^4) as follows:

$$T = \begin{pmatrix} B_1^1 & \dots & B_{n_t}^1 \\ \vdots & \ddots & \vdots \\ B_1^{m_t} & \dots & B_{n_t}^{m_t} \end{pmatrix} \quad (46)$$

where n_t is the n_{th} Bernoulli random variable and m_t is the m_{th} group of simulated samples. Summing up all the rows, one can get:

$$T_{sum} = \begin{bmatrix} B_1^1 + B_2^1 \dots + B_{n_t}^1 \\ \vdots \\ B_1^{m_t} + B_2^{m_t} \dots + B_{n_t}^{m_t} \end{bmatrix} \quad (47)$$

T_{sum} is the stochastic realization of $\mathbb{U}^c$, which can be used to estimate the CI of $\mathbb{U}^c$ numerically. The CI of $\tilde{P}_f^{pc}$ considering Kriging correlation can therefore be estimated as:

$$\tilde{P}_f^{pc} \in \left[\frac{\mu_{\mathbb{U}^c} - \gamma_{ci}\sigma_{\mathbb{U}^c}}{N_{MCS}}, \frac{\mu_{\mathbb{U}^c} + \gamma_{ci}\sigma_{\mathbb{U}^c}}{N_{MCS}} \right], \quad \alpha = 0.05, \gamma_{ci} = 1.96. \quad (48)$$

Compared to the CI of $\tilde{P}_f^{pc}$ in Eq. (20), which is derived from Chebyshev's inequality, the CI in Eq. (48) is more precise. In the next section, the performance of the proposed CIs for $\tilde{P}_f^{pc}$ is investigated.

4. Numerical Investigations

In this section, the performance of proposed CIs are investigated for five numerical examples. In this paper, the proposed CIs are compared with the P_f^{mcs} and not the true failure probability P_f . This is because the true failure probability P_f is not available prior to implementing any reliability analysis method. This approach is acceptable since the coefficient of variation of the failure probability estimate based on Monte Carlo simulation is set to be small; therefore, it is expected that P_f^{mcs} will be close to P_f . Note that the level of probability of failure determines the number of candidate design samples according to Eq.(13) and (17). For example, N_{MCS} needs to be as large as 4×10^6 for $\hat{P}_f^{dc} = 10^{-4}$ and $COV_{\hat{P}_f^{dc}} = 0.05$. Therefore, the size of associated covariance matrix defined in Eq.(38) is as large as 4×10^6 by 4×10^6 , which may cause computation burden for computers. Considering our purpose is to compare the Confidence Intervals with and without considering Kriging correlation, Chebyshev's inequality by Wang et al [48], cases with probability of failure below 10^{-4} are not explored due to aforementioned computational limitations. To ensure that the application of the CLT is appropriate, the number of candidate design samples, N_{MCS} , should be sufficiently large and a reasonable threshold for N_{MCS} is necessary. According to our numerical investigations, $N_{MCS} \geq 10^4$ is sufficient for the application of CLT in the derivation of CI for $\tilde{P}_f^{pc}$. The number of initial training points is set as three times of the number of random variables for all examples. Moreover, Latin Hypercube Sampling(LHS) strategy based on the X-space without normalization is applied in these examples. For consistency, the set of candidate design samples, S , remains exactly the same throughout the process of reliability analysis for all CIs for a given example. The EFF learning function is applied for selection of best training points. This function is defined below [13], [59]:

$$\begin{aligned}
EFF(\mathbf{x}) = & (\mu_{\hat{g}}(\mathbf{x}) - a) \left[2\Phi\left(\frac{a - \mu_{\hat{g}}(\mathbf{x})}{\sigma_{\hat{g}}(\mathbf{x})}\right) - \Phi\left(\frac{a^- - \mu_{\hat{g}}(\mathbf{x})}{\sigma_{\hat{g}}(\mathbf{x})}\right) - \Phi\left(\frac{a^+ - \mu_{\hat{g}}(\mathbf{x})}{\sigma_{\hat{g}}(\mathbf{x})}\right) \right] \\
& - \sigma_{\hat{g}}(\mathbf{x}) \left[2\phi\left(\frac{a - \mu_{\hat{g}}(\mathbf{x})}{\sigma_{\hat{g}}(\mathbf{x})}\right) - \phi\left(\frac{a^- - \mu_{\hat{g}}(\mathbf{x})}{\sigma_{\hat{g}}(\mathbf{x})}\right) - \phi\left(\frac{a^+ - \mu_{\hat{g}}(\mathbf{x})}{\sigma_{\hat{g}}(\mathbf{x})}\right) \right] \\
& + 2\sigma_{\hat{g}}(\mathbf{x}) \left[\Phi\left(\frac{a^+ - \mu_{\hat{g}}(\mathbf{x})}{\sigma_{\hat{g}}(\mathbf{x})}\right) - \Phi\left(\frac{a^- - \mu_{\hat{g}}(\mathbf{x})}{\sigma_{\hat{g}}(\mathbf{x})}\right) \right],
\end{aligned} \tag{49}$$

where $\phi(\cdot)$ is the PDF and $\Phi(\cdot)$ is the CDF of the standard normal distribution. In this paper, $a = 0$, $a^+ = 2\sigma_{\hat{g}}(\mathbf{x})$ and $a^- = -2\sigma_{\hat{g}}(\mathbf{x})$. The point that maximizes the EFF response is chosen as the next-best training point. The number of calls to a performance function is denoted as N_{call} . Because this paper only investigates the CI of $\hat{P}_f^{pc}$, it is reasonable that the stopping criterion for Kriging refinement is chosen as $N_{call} \geq N_{thr}$. Thus, training point enrichment stops when it reaches a threshold number of evaluations of the performance function. The steps of numerical investigation are elaborated in Algorithm 1.

Algorithm 1. Steps for numerical investigations

1. Generating initial candidate design samples S with Latin Hypercube Sampling (LHS)
 2. Randomly select initial training samples $\mathbf{x}_{tr}$ from S and evaluate their responses $g(\mathbf{x}_{tr})$
 3. Construct the Kriging model $\hat{g}(\mathbf{x})$ based on $\mathbf{x}_{tr}$ and $g(\mathbf{x}_{tr})$
 4. Estimate the mean $\sigma_{\hat{g}}(\mathbf{x})$, standard deviation $\sigma_{\hat{g}}(\mathbf{x})$, $\hat{P}_f^{pc}$ and its confidence interval for S with $\hat{g}(\mathbf{x})$
 5. Search for the next best training points $\mathbf{x}_{tr}^*$ using learning function and update the training samples $\mathbf{x}_{tr}$
 6. Check if the stopping criterion $N_{call} \leq N_{thr}$ is satisfied or not:
 - (a). Satisfied. Go to step 7.
 - (b). Unsatisfied. Estimate the response $g(\mathbf{x}_{tr}^*)$ for $\mathbf{x}_{tr}^*$ and go back to Step 3.
 7. Report $\hat{P}_f^{MCS}$ and its confidence interval.
-

4.1 Two dimensional problems

This non-linear four-branch series system problem has been investigated in many studies [12], [24], [48]. In this problem, random variables all follow mutually independent standard normal distributions as described in Table 1. The performance function, $g(\mathbf{x})$, is given as:

$$g(x_1, x_2) = \min \begin{cases} 3 + 0.1(x_1 - x_2)^2 - \frac{(x_1 + x_2)}{\sqrt{2}} \\ 3 + 0.1(x_1 - x_2)^2 + \frac{(x_1 + x_2)}{\sqrt{2}} \\ (x_1 - x_2) + \frac{6}{\sqrt{2}} \\ -(x_1 - x_2) + \frac{6}{\sqrt{2}} \end{cases} \tag{50}$$

Table 1. Random variables in example 4.1.

Random variable	Distribution	Mean	Standard deviation
x_1, x_2	Normal	0	1

The number of candidate design samples for Kriging Monte Carlo simulations is chosen as $N_{MCS} = 1 \times 10^5$, in order to satisfy $COV_{P_f} < 0.05$ as suggested in [12]. In this example, the threshold number for the stopping criterion is set to $N_{thr} = 70$, such that $\hat{P}_f^{pc} \cong P_f^{mcs}$ when N_{call} is close to N_{thr} . The initial training samples are illustrated in Fig. 1(a), where x_{in} means the initial training samples. The 95% CIs of $\hat{P}_f^{pc}$ without considering Kriging correlation is shown in Fig. 1(b), whereas the comparison between the 95% CI using Chebyshev's inequality proposed by Wang et al. [48] and the CI considering Kriging correlation is shown in Fig. 1(c). Additionally, to clearly explore the performance of the proposed approaches, the true absolute difference between $\hat{P}_f^{pc}$ and P_f^{mcs} is investigated:

$$\epsilon = |P_f^{mcs} - \hat{P}_f^{pc}|. \quad (51)$$

Different from the work in [48], which directly uses $P_f^{mcs} = 4.5 \times 10^{-3}$ from the literature, P_f^{MCS} here is computed based on the current set S , because P_f^{mcs} can be different for different simulations. In this example, $P_f^{mcs} = 4.60 \times 10^{-3}$ for the generated set, S , in this simulation. Note that S for both P_f^{mcs} and $\hat{P}_f^{pc}$ are exactly the same. The estimated bounds of error, $\hat{\epsilon}$, versus N_{call} , for the three different approaches, are illustrated in Fig. 1(d). Fig. 1(e) records the coefficient of variant of $\hat{P}_f^{pc}$ for both cases with and without considering Kriging correlation:

$$C.O.V_{\hat{P}_f^{pc}} = \frac{\sigma_{\mathbb{U}}/N_{MCS}}{\mu_{\mathbb{U}}/N_{MCS}} = \frac{\sigma_{\mathbb{U}}}{\mu_{\mathbb{U}}} \text{ or } \frac{\sigma_{\mathbb{U}^c}}{\mu_{\mathbb{U}^c}} \quad (52)$$

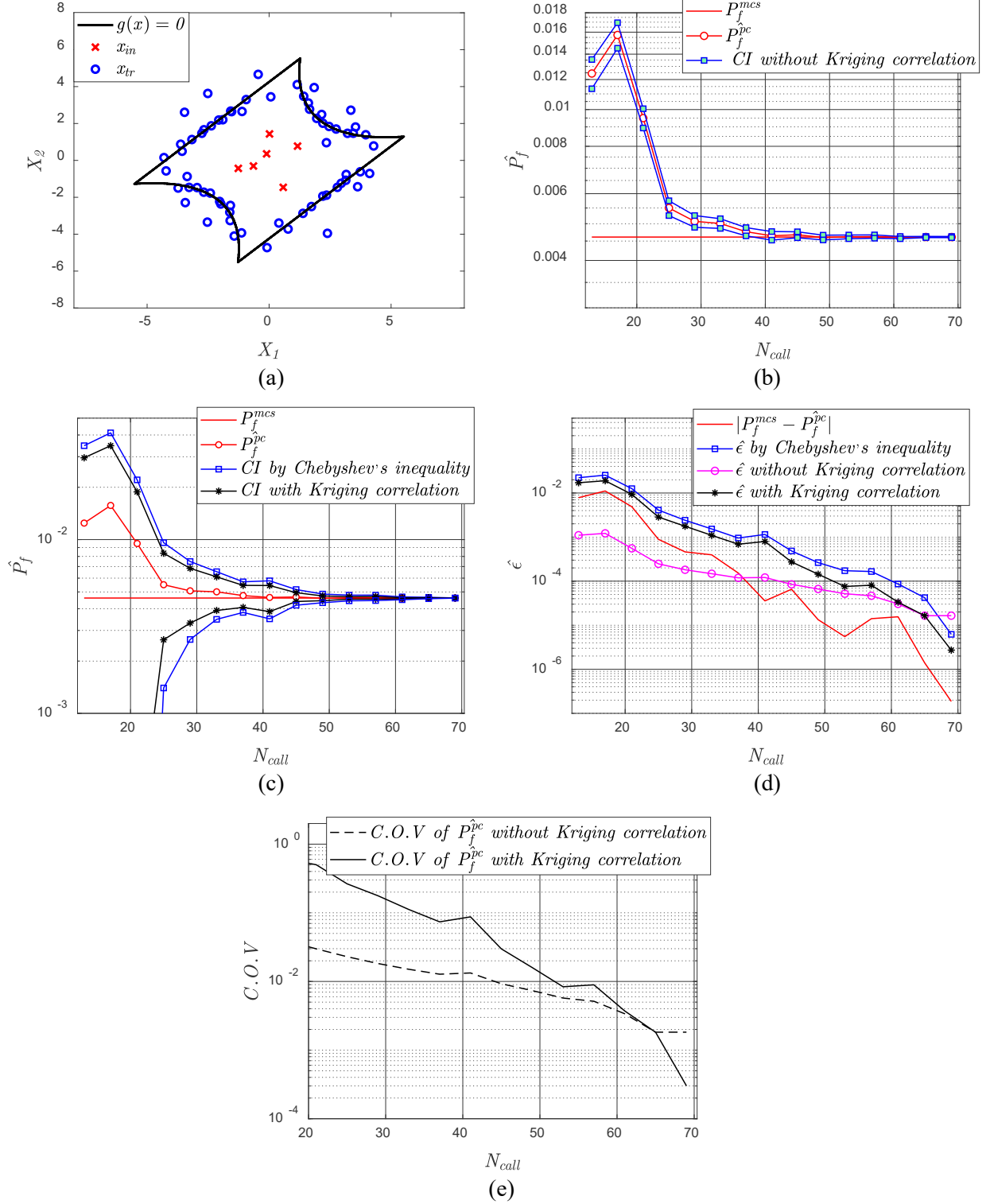


Fig. 1. Reliability analysis results for Example 1 including (a) Illustration of initial training samples (b) $\hat{P}_f$ and the CI without Kriging correlation vs N_{call} , (c) $\hat{P}_f$ and the CI with Kriging correlation vs N_{call} , and (d) $\hat{\epsilon}$ vs N_{call} (e) Curves of $C.O.V_{\hat{P}_f^{pc}}$

According to Fig. 1(b), the CI without considering Kriging correlation tends to be accurate when $N_{call} >$

40. However, it performs poorly at the early stage when $N_{call} \leq 40$. On the other hand, in Fig. 1(c), both the CIs by Chebyshev's inequality [48] and the proposed CI considering Kriging correlation perform well. This trend is also reflected clearly in Fig. 1(d). Nevertheless, the one estimated by considering Kriging correlation offers a narrower band than the one estimated by Chebyshev's inequality. Because the CI estimated by Wang et al. [48] is based on the Chebyshev's inequality, as stated in Eq. (20), it assumes that the distribution of $\mathbb{U}$ is unknown. However, the distribution of $\mathbb{U}$ or $\mathbb{U}^c$ as proved in this paper is a PBD. This finding was subsequently used in the derivation of the proposed CIs with and without considering Kriging correlation. Moreover, according to Fig. 1(d), one can infer that error $\hat{\epsilon}$, estimated without considering Kriging correlation, is smaller than the true error, ϵ , at the early stage (e.g. $N_{call} < 40$). However, it is close to the $\hat{\epsilon}$ estimated considering Kriging correlation in the later stage (e.g. $N_{call} \geq 40$). This trend is because of the contribution of the off-diagonal elements of Σ_{pb} in Eq. (38). At the early stages, this contribution is significant resulting in a significantly smaller $\sigma_{\mathbb{U}^c}^2$ for the uncorrelated PBD compared to the case considering Kriging correlation. Therefore, the CI without considering Kriging correlation tends to be considerably narrower than the CI considering Kriging correlation. As more training points are added into the construction of the Kriging model, $\hat{P}_f^{pc}$ gradually converges to P_f^{mcs} , and the off-diagonal elements in Eq. (38) become negligible. This is reflected in the phenomenon where the CI without considering Kriging correlation performs close to CI considering Kriging correlation for larger number of training points. Hence, the CI considering Kriging correlation is recommended, because it offers accurate confidence intervals over the entire range of training points for Kriging-based reliability analysis with probabilistic classification. However, the estimates of CI considering Kriging correlation and Chebyshev's inequality are computationally demanding, owing to the large scale of the covariance matrix in Eq. (38). Although the CI without considering Kriging correlation only performs well at post stage, the coefficient of variation of $\hat{P}_f^{pc}$ by this approach finally converges to the performance via the approach considering Kriging correlation according to Fig. 1(e). This finding can draw the conclusion that the coefficients of variation of $\hat{P}_f^{pc}$ with and without considering Kriging correlation asymptotically converge to the same level of value as the training points increase.

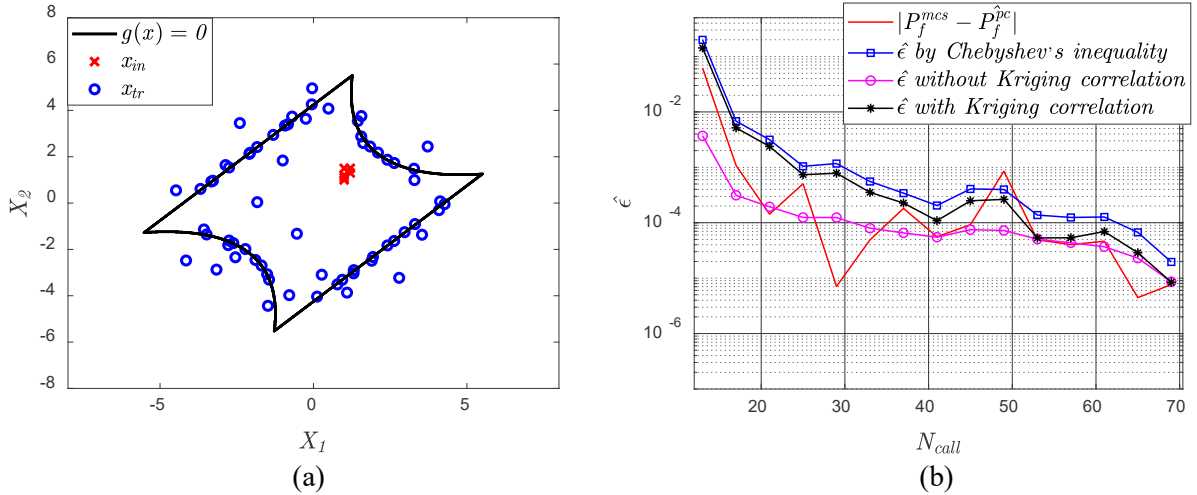


Fig. 2. Analysis of the robustness of the proposed CIs to the arrangement of initial training samples: (a) Illustration of ill-conditioned initial training samples and (b) $\hat{\epsilon}$ vs N_{call}

To investigate the robustness of the proposed CIs to the arrangement of initial training samples, we define a set of ill-conditioned initial training samples as shown in Fig. 2(a). It can be observed that these samples are clustered close to each other. The error estimates for confidence intervals are presented in Fig. 2(b). Compared to the results in Fig. 1(d), which represents the case where initial training samples are generated

following design of experiment methods, it is seen that artificially clustering initial training samples affects the performance of the proposed CIs. Although this represents an extremely unlikely scenario for generation of training samples, the proposed CI with Kriging correlations considered present a satisfactory performance except for when $N_{call} = 48$. Moreover, effects of active learning strategies on the performance of the CIs is investigated in Fig. 3. For this purpose, as shown in Fig. 3(a), training samples are randomly selected in the area $[-4, 4] \times [-4, 4]$ as compared to a strategic selection based on EFF learning function. According to Fig. 3(b), the proposed CI that considers Kriging correlations exhibits a satisfactory performance in the sense that it provides an upper bound for the true error. However, all CIs fail to show convergence compared to the case that adopts the active learning function. This can be attributed to the fact that random selection leads to training points that are not close to the limit state. As a result, it cannot reduce the uncertainty of the Kriging model for failure probability estimation.

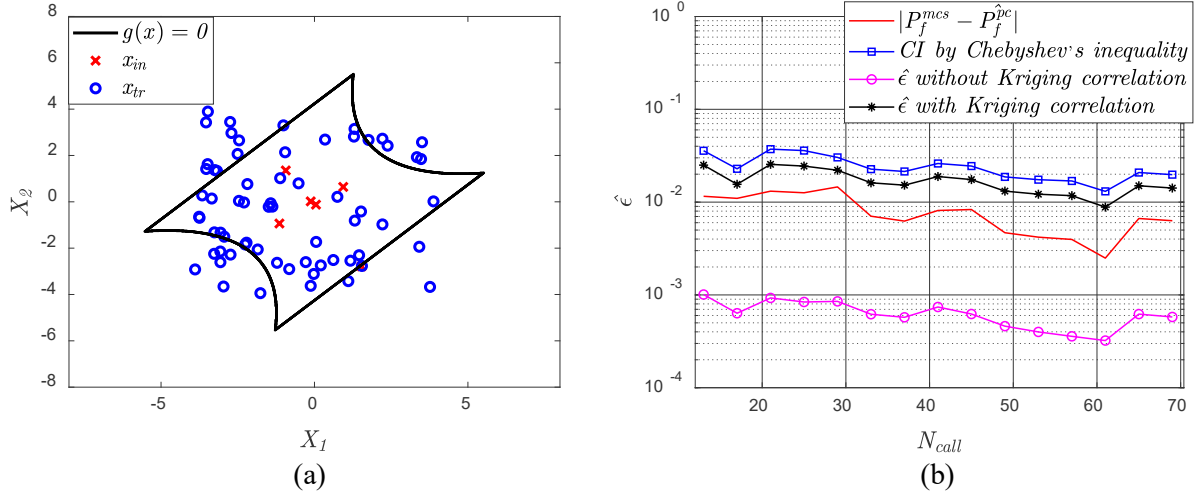


Fig. 3. Reliability analysis results for example 1 when training points are randomly selected: (a) Illustration of initial training samples and (b) $\hat{\epsilon}$ vs N_{call}

4.2 Truss problems

4.2.1 Structural Roof Truss

The second example is a structural truss problem shown in Fig. 4 with six normal random variables [60], [61]. In this example, the maximum vertical displacement at the centre of the structure is selected as the critical response to consider in the limit state function as follows:

$$G(q, l, A_c, E_c, A_s, E_s) = 0.03 - \frac{ql^2}{2} \left(\frac{3.81}{A_c E_c} + \frac{1.13}{A_s E_s} \right), \quad (53)$$

where q is uniformly distributed loading, l is the horizontal length of the roof truss, A_c and A_s are the cross-section areas, and E_c and E_s are the Young's modulus of the steel and concrete beams, respectively. The probabilistic information of the six random variables are shown in Table 2. The number of candidate design samples for Kriging-based MCS is 5×10^4 , and the P_f^{mcs} is estimated as 9.6×10^{-3} , which means that the coefficient of variation satisfies $COV_{P_f} < 0.05$. The threshold number for the stopping criterion for refining the Kriging model is set as $N_{thr} = 90$. Three CIs determined using the three approaches discussed earlier are illustrated in Fig. 5.

Table 2. Random variables of example 4.2.1.

Random variable	Unit	Mean	C.O.V (%)
q	(N/m)	2×10^4	7
l	(m)	12	1
A_c	(m ²)	4×10^{-2}	12
E_c	(N/m ²)	2×10^{10}	6
A_s	(m ²)	9.82×10^{-4}	6
E_s	(N/m ²)	1×10^{11}	6

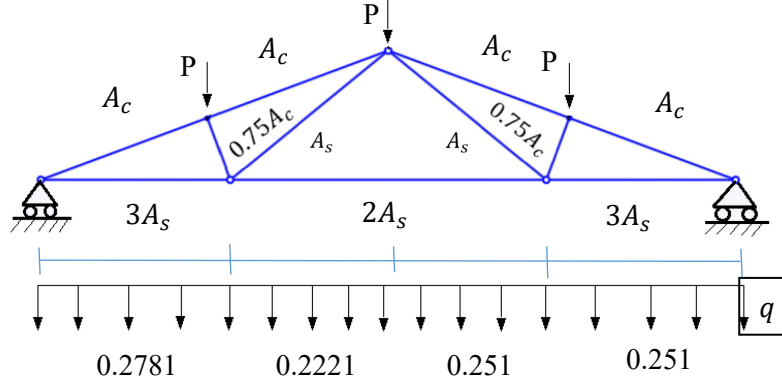


Fig. 4. Example 4.2.1, The roof truss in [23].

According to Fig. 5(a), the CI without considering Kriging correlation fails to cover the true failure probability, P_f^{mcs} , at the early stage (e.g. $N_{call} \leq 70$). However, it works well at the later stage (e.g. $N_{call} \geq 70$). The reason behind this phenomenon has been explained in example 4.1. In Figs. 5(b) and 5(c), both CIs estimated with Kriging correlation and by Chebyshev's inequality [48] perform well. However, the correlated PBD-based CI offers a narrower bound and closer to the true error, ϵ , compared to the approach via Chebyshev's inequality [48]. As shown in Fig. 5(d), the coefficients of variation of $\bar{P}_f^{pc}$ estimated through the proposed two approaches gradually converge to 10^{-3} . However, $C.O.V_{\bar{P}_f^{pc}}$ considering the Kriging correlation keeps decreasing while $C.O.V_{\bar{P}_f^{pc}}$ without considering Kriging correlation does not change too much after 10^{-3} .

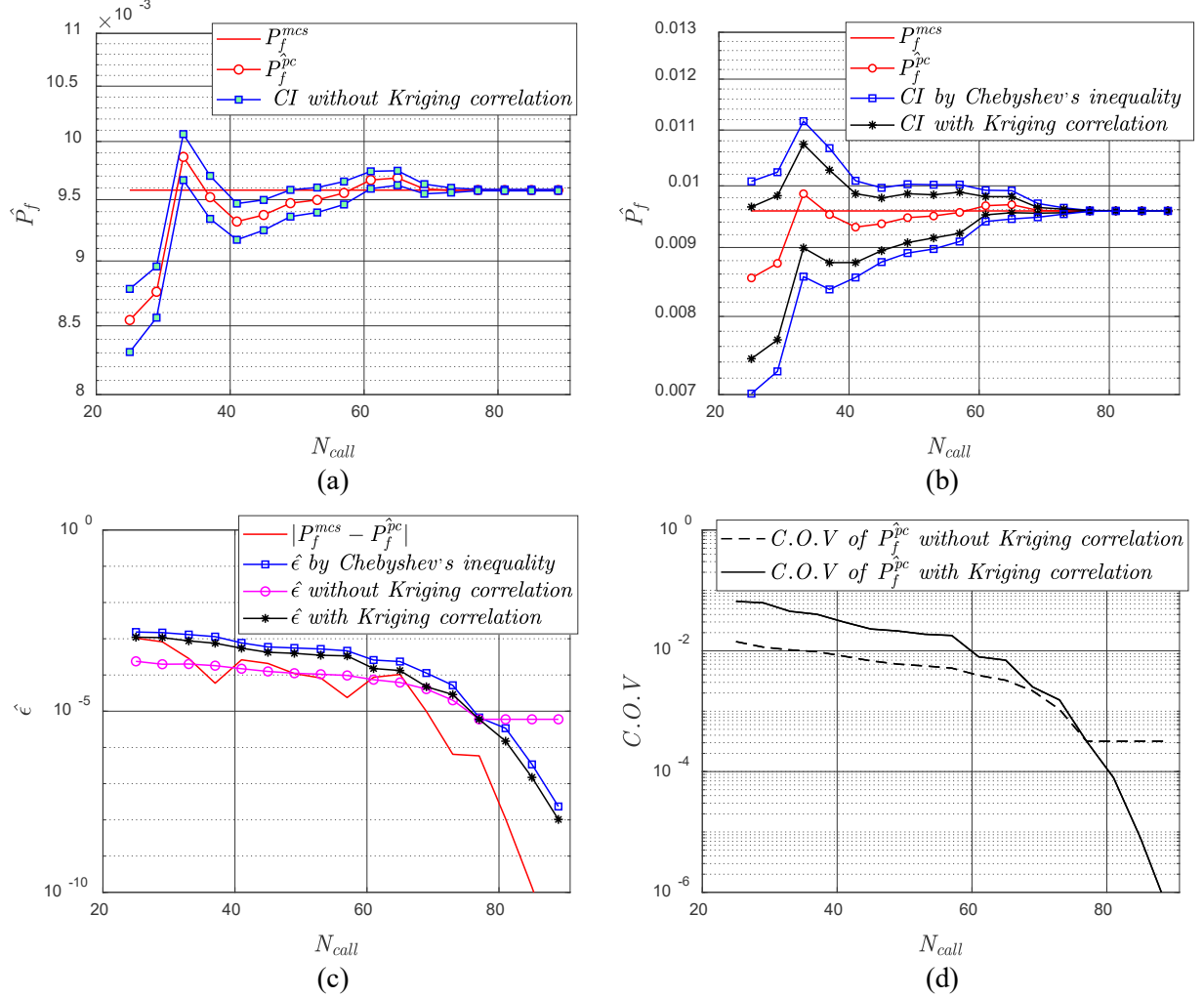


Fig. 5. Reliability analysis results for Example 4.2.1 including: (a) $\hat{P}_f$ vs N_{call} without considering Kriging correlation (b) $\hat{P}_f$ vs N_{call} considering Kriging correlation (c) ϵ vs N_{call} (d) Curves of $C.O.V_{\hat{P}_f^{pc}}$

4.2.2 Modified Truss Structure

In the fourth example, a 23-bar truss with 10 input random variables is investigated [15], [48]. In this truss, as shown in Fig. 6, 11 bars are horizontal and 12 are diagonal. The performance function is defined as:

$$g(x) = 0.14 - |dis(x)|, \quad (54)$$

where $dis(x)$ is the vertical displacement of the truss at point E . The truss is subject to six vertical loadings, P_1 to P_6 , which follow Gumbel distributions. A_1 and A_2 are the cross-section areas and E_1 and E_2 are the Young's modulus of the horizontal and diagonal bars, respectively. The 10 mutually independent random variables are described in Table 3. For this high dimensional example, $N_{MCS} = 5 \times 10^4$, and $P_f^{mcs} = 9.10 \times 10^{-3}$, such that $COV_{P_f} < 0.05$. Moreover, $N_{thr} = 100$. The simulation results are presented in Fig.

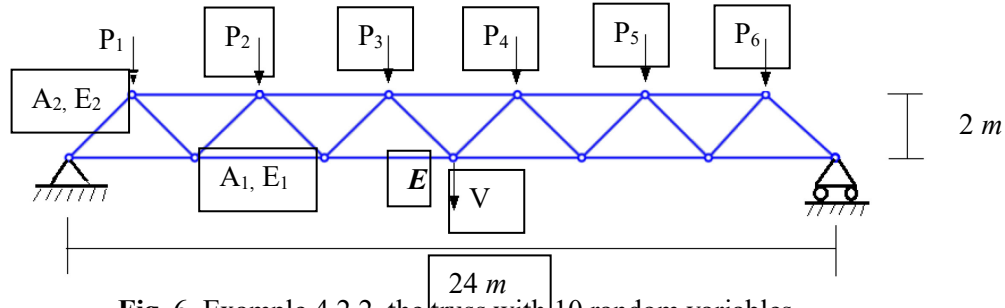


Fig. 6. Example 4.2.2, the truss with 10 random variables

Table 3. Random variables in example 4.2.2.

Random variable	Distribution	Mean	Standard deviation
$P_1 - P_6$	Gumbel	6.5×10^4	6.5×10^3
A_1	Lognormal	2×10^{-3}	2×10^{-4}
A_2	Lognormal	1×10^{-3}	1×10^{-4}
E_1	Lognormal	2.1×10^{11}	2.1×10^{10}
E_2	Lognormal	2.1×10^{11}	2.1×10^{10}

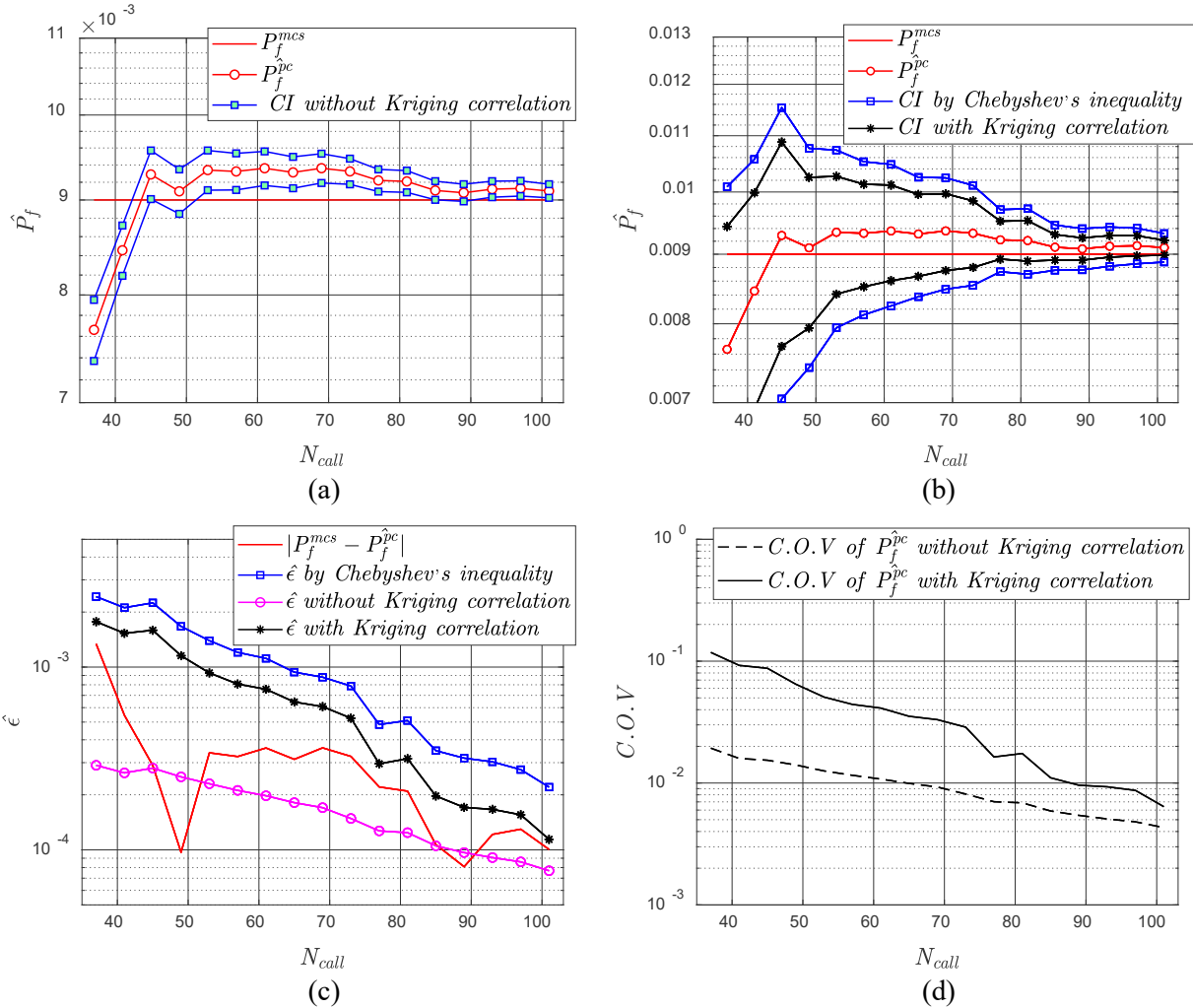


Fig. 7. Reliability analysis results for Example 4.2.2 including: (a) $\hat{P}_f$ vs N_{call} without considering Kriging correlation (b) $\hat{P}_f$ vs N_{call} considering Kriging correlation (c) ϵ vs N_{call} (d) Curves of C.O.V of P_f^{pc}

In Fig. 7, the CI of $\tilde{P}_f^{pc}$ without considering Kriging correlation exhibits poor performance, while the CI considering Kriging correlation works well for all stages of the adaptive Kriging-based reliability analysis. Moreover, the CI considering Kriging correlation is tighter than the CI by Chebyshev's inequality [48], as shown in Figs. 7(b) and 7(c). According to Fig. 7(d), the $C.O.V_{\tilde{P}_f^{pc}}$ estimated via proposed two approaches show the same convergence direction.

4.3 Non-linear oscillator

4.3.1 One-degree-of-freedom oscillator

This example investigated in this section is a mechanical non-linear oscillator with six random variables, as illustrated in Fig. 8 [12], [48]. For this non-linear problem, the performance function is defined as:

$$g(k_1, k_2, m, r, t_1, F_1) = 3r - \left| \frac{2F_1}{m\omega_0^2} \sin\left(\frac{\omega_0 t_1}{2}\right) \right|, \quad (55)$$

where $\omega_0 = \sqrt{\frac{k_1+k_2}{m}}$ is the system frequency. The probabilistic information of the six input random variables are summarized in Table 4. In this example, $N_{MCS} = 7 \times 10^4$ and P_f^{mcs} is estimated as 2.89×10^{-2} with $COV_{P_f} < 0.05$. Moreover, $N_{thr} = 76$. The CIs estimated with/without considering Kriging correlation and via Chebyshev's inequality [48], are all presented in Fig. 9.

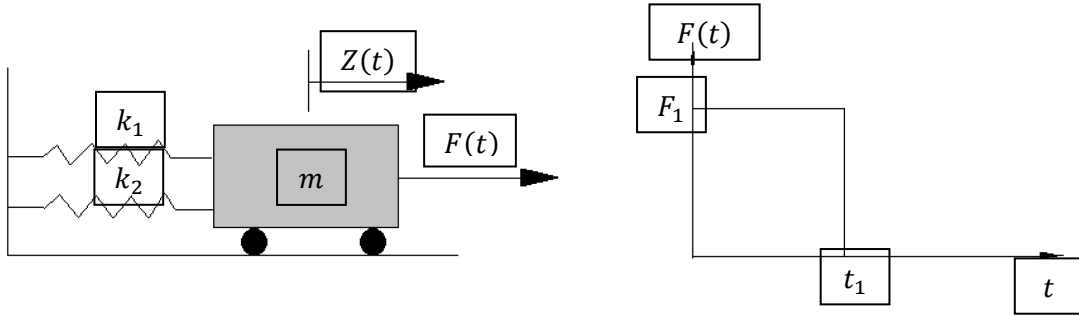


Fig. 8 Example 4.3.1, A two-degree-of-freedom

As illustrated in Figs. 9(a) and 9(c), the CI estimated without considering Kriging correlation works well when $N_{call} > 65$. Moreover, while both the CIs through considering Kriging correlation and Chebyshev's inequality are accurate, the former approach yields a tighter bound. This trend is also observed in Fig. 9(c). Generally, CI estimated without considering Kriging correlation is efficient for implementation, but is inaccurate at early stages. However, the CI estimated with Kriging correlation is accurate in all stages, but remains computationally demanding. Moreover, the $C.O.V_{\tilde{P}_f^{pc}}$ estimated via proposed two approaches converges to the same value as shown in Fig. 9(d).

Table 4. Random variables in example 4.3.1.

Random variable	Distribution type	Mean	Standard Deviation
m	Normal	1	0.05
k_1	Normal	1	0.1
k_2	Normal	0.1	0.01
r	Normal	0.5	0.05
F_1	Normal	1	0.2
t_1	Normal	1	0.2

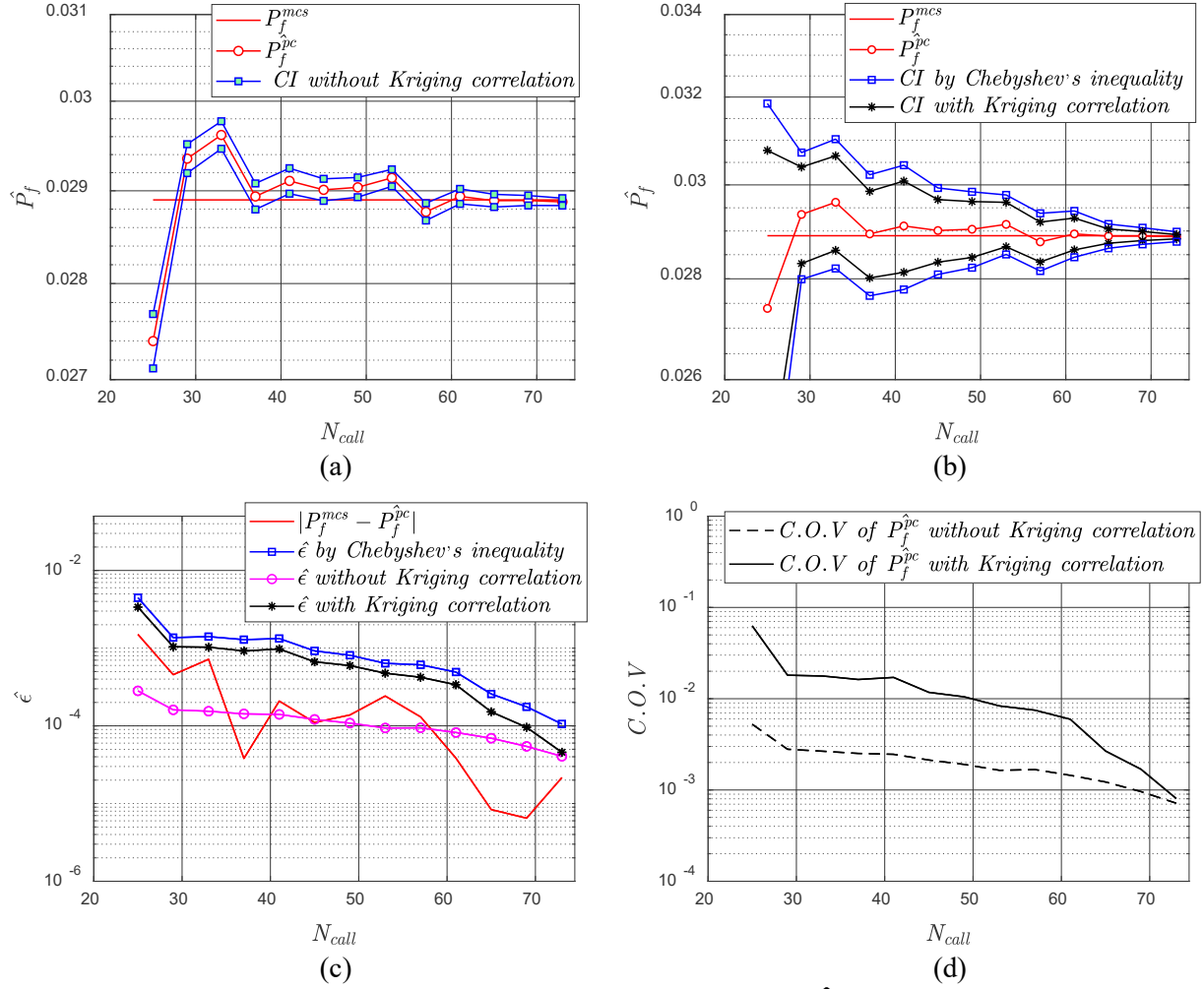


Fig. 9. Reliability analysis results for Example 4.3.1 including: (a) $\hat{P}_f$ vs N_{call} without considering Kriging correlation (b) $\hat{P}_f$ vs N_{call} considering Kriging correlation (c) $\hat{\epsilon}$ vs N_{call} (d) Curves of $C.O.V.$ of $\hat{P}_f^{pc}$

4.3.2 Two-degree-of-freedom oscillator

To explore the performance of the proposed CIs for a non-linear case, we investigate a two-degree-of-freedom primary-secondary damped oscillator as shown in Fig. 10. This example was originally used in the report by Der Kiureghian and De Stefano [8], and was further explored by Bourinet et al [62], Dubourg et al [23], [63], and Hu and Mahadevan [64]. Let m_p and m_s , k_p and k_s , $\omega_p = \sqrt{k_p/m_p}$ and $\omega_s = \sqrt{k_s/m_s}$, and ξ_p and ξ_s denote the primary and secondary masses, spring stiffness, natural frequencies and damping ratios of the oscillator system, respectively. Here, p and s denote the primary and secondary oscillator, respectively. Thus, the mean-square relative displacement of the secondary spring under a white noise base acceleration S can be computed as:

$$E_s[x_s^2] = \pi \frac{S_0}{4\xi_s\omega_s^3} \frac{\xi_a\xi_s}{\xi_p\xi_s(4\xi_a^2 + \theta^2) + \gamma\xi_a^2} \frac{(\xi_p\omega_p^3 + \xi_s\omega_s^3)\omega_p}{4\xi_a\omega_a^4} \quad (56)$$

where S_0 is the intensity of the white noise, $\gamma = m_s/m_p$, $\omega_a = (\omega_p + \omega_s)/2$, $\xi_a = (\xi_p + \xi_s)/2$ and $\theta =$

$(\omega_p - \omega_s)/\omega_a$ is the tuning parameter. For this example, the limit state function is defined as:

$$g(\mathbf{X}) = F_s - pe \cdot k_s \sqrt{E_s[x_s^2]}, \quad (57)$$

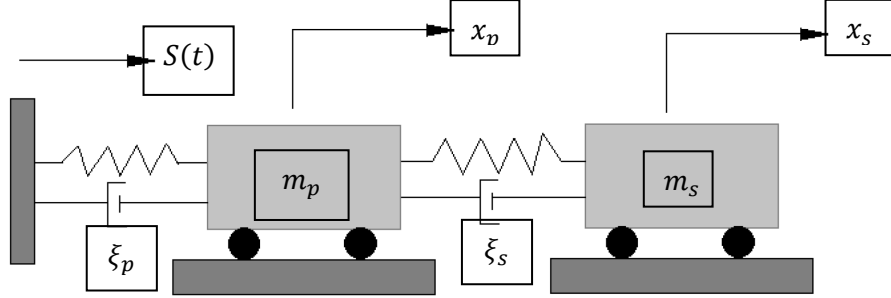


Fig. 10. Example 4.3.2, a two-degree-of-freedom oscillator under a white-noise base acceleration

Table 5. Random variables in example 4.3.2.

Random variable	Distribution type	Mean	Standard Deviation
m_p	Lognormal	1.5	0.15
m_s	Lognormal	0.01	0.001
k_p	Lognormal	1	0.2
k_s	Lognormal	0.01	0.002
ξ_p	Lognormal	0.05	0.02
ξ_s	Lognormal	0.02	0.01
F_s	Lognormal	15	1.5
S_0	Lognormal	100	10

where pe denotes the peak factor set and is equal to 3 in this example. The eight mutually independent random variables are summarized in Table 5. In this example, $N_{MCS} = 1 \times 10^5$ and P_f^{mcs} is estimated as 4.80×10^{-3} with $COV_{\hat{P}_f^{pc}} < 0.05$. Moreover, $N_{thr} = 1000$. The proposed CIs estimated with and without considering Kriging correlation and that via Chebyshev's inequality [48] are all presented in Fig. 11.

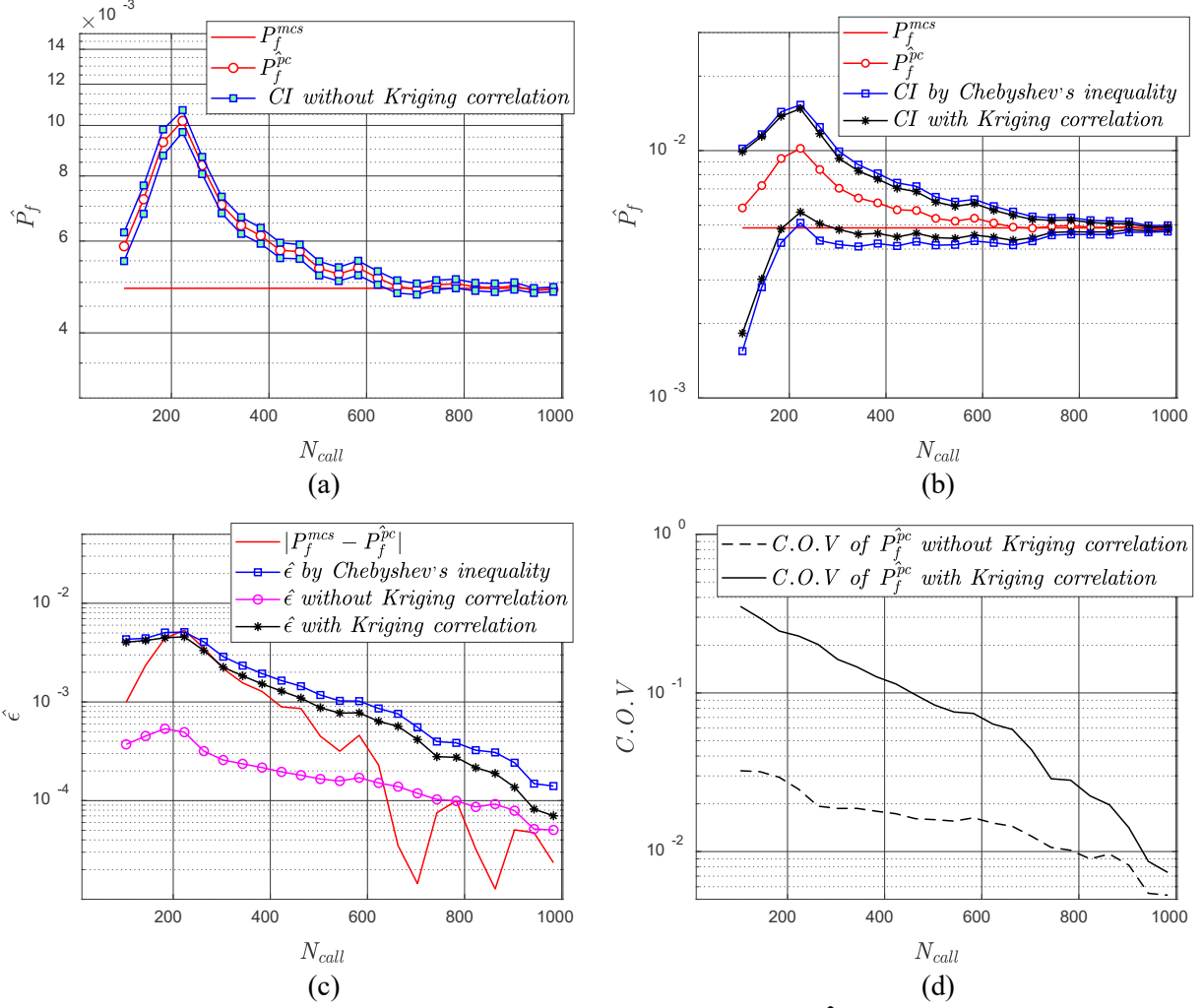


Fig. 11. Reliability analysis results for Example 4.3.2 including: (a) $\hat{P}_f$ and the CI without Kriging correlation vs N_{call} , (b) $\hat{P}_f$ and the CI with Kriging correlation vs N_{call} , (c) ϵ vs N_{call} , and (d) $C.O.V_{\hat{P}_f^{pc}}$ vs N_{call}

According to the plots in Figs. 11(a) and (c), the CI estimated without considering Kriging correlation works well when $N_{call} > 630$. On the other hand, the CIs that consider Kriging correlation and Chebyshev's inequality are accurate for N_{call} larger than 240. Fig. 11(d) illustrates a similar convergence pattern of the coefficient of variation of $\hat{P}_f^{pc}$ for different correlation considerations.

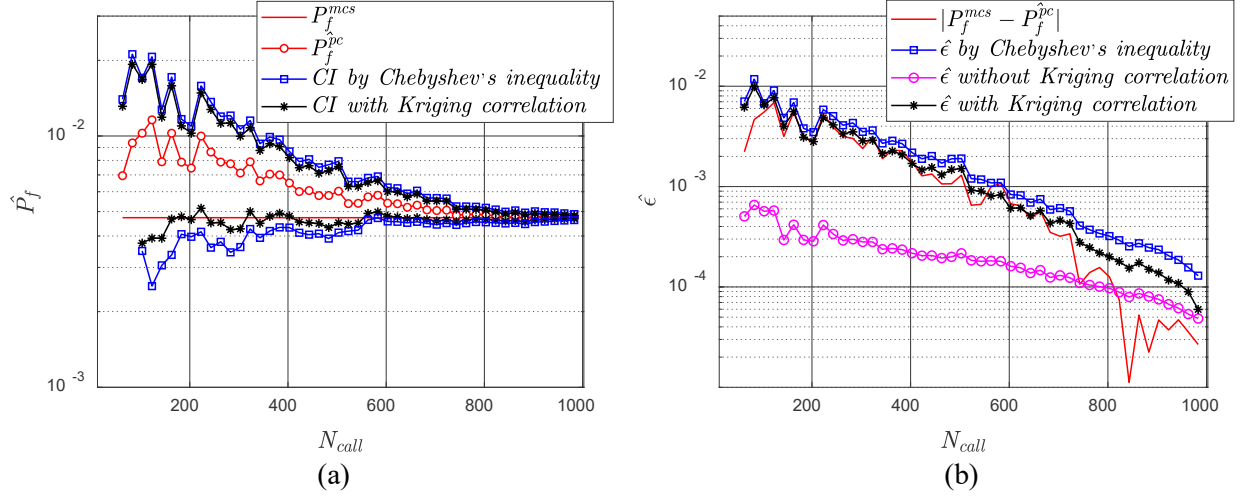


Fig. 12. Analysis of the robustness of the proposed CIs to the arrangement of initial training samples for example 4.3.2: (a) $\hat{P}_f$ and the CI with Kriging correlation vs N_{call} , (b) $\hat{\epsilon}$ vs N_{call}

To investigate the robustness of the proposed CIs to the arrangement of initial training samples, we define a set of ill-conditioned initial training samples. Results of corresponding analyses are summarized in Fig. 12. Compared to the results in Fig. 11(c), the error estimate considering Kriging correlation presented in Fig. 12(b) tends to be accurate for $N_{call} \geq 630$. It can be inferred that the initial design of training samples affects the performance of the proposed confidence intervals. However, they become more accurate as the number of training samples increases. Therefore, a strategy for the initial design of experiments is necessary and should be appropriately defined.

5. Conclusion

This paper proposes a novel approach to derive confidence intervals (CIs) for failure probability estimates in adaptive Kriging-based reliability analysis methods. The approach builds on the fact that the summation of probabilistic classification-based indicator follows a Bernoulli distribution, which is subsequently leveraged to derive the CI for failure probability estimates. Two variants of CIs are developed in this paper, where one approach disregards Kriging correlation for candidate design samples, and the other approach takes into account such correlations. Both the two proposed CIs are investigated for four groups of numerical example. Results show that the CI considering Kriging correlation offers more accurate bounds for failure probability estimates compared to the uncorrelated CI model. It is also demonstrated that the proposed CI considering Kriging correlation outperforms the existing approach that relies on Chebyshev's inequality. While the CI without considering Kriging correlation is primarily accurate at later stages of adaptive reliability analysis, it is computationally efficient compared to other techniques. The proposed CIs can be leveraged to derive efficient stopping criteria, optimal learning strategies and efficient solutions for high-dimensional problems.

It should be noted that the performance of the proposed CIs can be partially affected by the initial design of experiments. Specifically, if the initial training samples are properly generated e.g. via Latin Hypercube Sampling (LHS), the estimated CI will be appropriate. However, clustered initial samples may lead to inaccurate CIs. Such ill-conditioned initial samples fail to capture the global responses of the performance function and can lead to fast convergence in local regions. Therefore, the appropriate generation of initial training samples is necessary for estimating proper confidence intervals. Moreover, the proposed CIs considering Kriging correlation are computationally demanding. There is no concern for computational intractability for the uncertainty indicator without Kriging correlation, since it relies on computationally simple operations. However, the uncertainty indicator with Kriging correlation needs to compute a very large covariance matrix. The size of this matrix is associated with the number of candidate design samples.

As N_{call} increases, it becomes increasingly demanding to generate the covariance matrix. Due to this computational challenge, numerical examples investigated in this paper have probabilities of failure larger than 10^{-4} . Although investigating techniques for the computation of large covariance matrices is out of the scope of this paper, it is an important topic to pursue in future studies.

Acknowledgments

This research has been partly funded by the U.S. National Science Foundation (NSF) through awards CMMI-1333943, 1462183, and 1635569. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation. We also thank three anonymous reviewers for informative and valuable suggestions on paper revision.

References

- [1] Y. Qian, H. Takao, M. Umezu, and Y. Murayama, "Risk Analysis of Unruptured Aneurysms Using Computational Fluid Dynamics Technology: Preliminary Results," *Am. J. Neuroradiol.*, vol. 32, no. 10, pp. 1948–1955, Nov. 2011.
- [2] J. H. Choi, N.-H. Kim, and N. Kogiso, "Reliability for aerospace systems: Methods and applications," *Adv. Mech. Eng.*, vol. 8, no. 11, p. 1687814016677092, Nov. 2016.
- [3] T. Aven, "Risk assessment and risk management: Review of recent advances on their foundation," *Eur. J. Oper. Res.*, vol. 253, no. 1, pp. 1–13, Aug. 2016.
- [4] Z. Wang and A. Shafieezadeh, "REAK: Reliability analysis through Error rate-based Adaptive Kriging," *Reliab. Eng. Syst. Saf.*, vol. 182, pp. 33–45, Feb. 2019.
- [5] R. Y. Rubinstein and D. P. Kroese, *Simulation and the Monte Carlo method*. John Wiley & Sons, 2016.
- [6] G. Fishman, *Monte Carlo: concepts, algorithms, and applications*. Springer Science & Business Media, 2013.
- [7] R. Rackwitz and B. Flessler, "Structural reliability under combined random load sequences," *Comput. Struct.*, vol. 9, no. 5, pp. 489–494, Nov. 1978.
- [8] Kiureghian Armen Der and Stefano Mario De, "Efficient Algorithm for Second-Order Reliability Analysis," *J. Eng. Mech.*, vol. 117, no. 12, pp. 2904–2923, Dec. 1991.
- [9] B. Echard, N. Gayton, M. Lemaire, and N. Relun, "A combined importance sampling and kriging reliability method for small failure probabilities with time-demanding numerical models," *Reliab. Eng. Syst. Saf.*, vol. 111, pp. 232–240, 2013.
- [10] S. K. Au and J. L. Beck, "A new adaptive importance sampling scheme for reliability calculations," *Struct. Saf.*, vol. 21, no. 2, pp. 135–158, 1999.
- [11] S. K. Au and J. L. Beck, "Subset simulation and its application to seismic risk based on dynamic analysis," *J. Eng. Mech.*, vol. 129, no. 8, pp. 901–917, 2003.
- [12] B. Echard, N. Gayton, and M. Lemaire, "AK-MCS: an active learning reliability method combining Kriging and Monte Carlo simulation," *Struct. Saf.*, vol. 33, no. 2, pp. 145–154, 2011.
- [13] B. J. Bichon, M. S. Eldred, L. P. Swiler, S. Mahadevan, and J. M. McFarland, "Efficient global reliability analysis for nonlinear implicit performance functions," *ALAA J.*, vol. 46, no. 10, pp. 2459–2468, 2008.
- [14] U. Alibrandi, A. M. Alani, and G. Ricciardi, "A new sampling strategy for SVM-based response surface for structural reliability analysis," *Probabilistic Eng. Mech.*, vol. 41, pp. 1–12, Jul. 2015.
- [15] N. Roussouly, F. Petitjean, and M. Salaun, "A new adaptive response surface method for reliability analysis," *Probabilistic Eng. Mech.*, vol. 32, pp. 103–115, Apr. 2013.
- [16] X. Shi, A. P. Teixeira, J. Zhang, and C. G. Soares, "Kriging response surface reliability analysis of a ship-stiffened plate with initial imperfections," *Struct. Infrastruct. Eng.*, vol. 11, no. 11, pp. 1450–1465, Nov. 2015.
- [17] Z. Lv, Z. Lu, and P. Wang, "A new learning function for Kriging and its applications to solve reliability problems in engineering," *Comput. Math. Appl.*, vol. 70, no. 5, pp. 1182–1197, Sep. 2015.
- [18] Z. Sun, J. Wang, R. Li, and C. Tong, "LIF: A new Kriging based learning function and its application to structural reliability analysis," *Reliab. Eng. Syst. Saf.*, vol. 157, pp. 152–165, 2017.
- [19] N.-C. Xiao, M. J. Zuo, and C. Zhou, "A new adaptive sequential sampling method to construct surrogate models for efficient reliability analysis," *Reliab. Eng. Syst. Saf.*, vol. 169, pp. 330–338, Jan. 2018.
- [20] N. Lelièvre, P. Beaurepaire, C. Mattrand, and N. Gayton, "AK-MCSi: A Kriging-based method to deal with small failure probabilities and time-consuming models," *Struct. Saf.*, vol. 73, pp. 1–11, Jul. 2018.

- [21] Z. Wen, H. Pei, H. Liu, and Z. Yue, "A Sequential Kriging reliability analysis method with characteristics of adaptive sampling regions and parallelizability," *Reliab. Eng. Syst. Saf.*, vol. 153, pp. 170–179, 2016.
- [22] M. Balesdent, J. Morio, and J. Marzat, "Kriging-based adaptive Importance Sampling algorithms for rare event estimation," *Struct. Saf.*, vol. 44, pp. 1–10, Sep. 2013.
- [23] V. Dubourg, B. Sudret, and F. Deheeger, "Metamodel-based importance sampling for structural reliability analysis," *Probabilistic Eng. Mech.*, vol. 33, pp. 47–57, Jul. 2013.
- [24] X. Huang, J. Chen, and H. Zhu, "Assessing small failure probabilities by AK-SS: an active learning method combining Kriging and subset simulation," *Struct. Saf.*, vol. 59, pp. 86–95, 2016.
- [25] V. Dubourg, B. Sudret, and J.-M. Bourinet, "Reliability-based design optimization using kriging surrogates and subset simulation," *Struct. Multidiscip. Optim.*, vol. 44, no. 5, pp. 673–690, Nov. 2011.
- [26] J. Bect, L. Li, and E. Vazquez, "Bayesian Subset Simulation," *SIAMASA J. Uncertain. Quantif.*, vol. 5, no. 1, pp. 762–786, Jan. 2017.
- [27] X. Yang, Y. Liu, C. Mi, and C. Tang, "System reliability analysis through active learning Kriging model with truncated candidate region," *Reliab. Eng. Syst. Saf.*, vol. 169, pp. 235–241, 2018.
- [28] X. Zhang, L. Wang, and J. D. Sørensen, "AKOIS: An adaptive Kriging oriented importance sampling method for structural system reliability analysis," *Struct. Saf.*, vol. 82, p. 101876, Jan. 2020.
- [29] W. Chen, C. Xu, Y. Shi, J. Ma, and S. Lu, "A hybrid Kriging-based reliability method for small failure probabilities," *Reliab. Eng. Syst. Saf.*, vol. 189, pp. 31–41, Sep. 2019.
- [30] N. Pedroni and E. Zio, "An Adaptive Metamodel-Based Subset Importance Sampling approach for the assessment of the functional failure probability of a thermal-hydraulic passive system," *Appl. Math. Model.*, vol. 48, pp. 269–288, Aug. 2017.
- [31] W. Fauriat and N. Gayton, "AK-SYS: An adaptation of the AK-MCS method for system reliability," *Reliab. Eng. Syst. Saf.*, vol. 123, pp. 137–144, 2014.
- [32] X. Yang, C. Mi, D. Deng, and Y. Liu, "A system reliability analysis method combining active learning Kriging model with adaptive size of candidate points," *Struct. Multidiscip. Optim.*, vol. 60, no. 1, pp. 137–150, Jul. 2019.
- [33] B. Gaspar, A. P. Teixeira, and C. G. Soares, "Assessment of the efficiency of Kriging surrogate models for structural reliability analysis," *Probabilistic Eng. Mech.*, vol. 37, pp. 24–34, Jul. 2014.
- [34] Z. Wang and A. Shafieezadeh, "Real-time high-fidelity reliability updating with equality information using adaptive Kriging," *Reliab. Eng. Syst. Saf.*, vol. 195, p. 106735, Mar. 2020.
- [35] R. Schöbi and B. Sudret, "Structural reliability analysis for p-boxes using multi-level meta-models," *Probabilistic Eng. Mech.*, vol. 48, pp. 27–38, Apr. 2017.
- [36] Z. Meng, Z. Zhang, D. Zhang, and D. Yang, "An active learning method combining Kriging and accelerated chaotic single loop approach (AK-ACSLA) for reliability-based design optimization," *Comput. Methods Appl. Mech. Eng.*, vol. 357, p. 112570, Dec. 2019.
- [37] J. Li and J. Chen, "Solving time-variant reliability-based design optimization by PSO-t-IRS: A methodology incorporating a particle swarm optimization algorithm and an enhanced instantaneous response surface," *Reliab. Eng. Syst. Saf.*, vol. 191, p. 106580, Nov. 2019.
- [38] T. Sauder, S. Marelli, and A. J. Sørensen, "Probabilistic robust design of control systems for high-fidelity cyber-physical testing," *Automatica*, vol. 101, pp. 111–119, Mar. 2019.
- [39] M. Rahimi, A. Shafieezadeh, D. Wood, E. J. Kubatko, and N. C. Dormady, "Bayesian calibration of multi-response systems via multivariate Kriging: Methodology and geological and geotechnical case studies," *Eng. Geol.*, vol. 260, p. 105248, Oct. 2019.
- [40] Z. Wang and A. Shafieezadeh, "Reliability-based Bayesian Updating using Machine Learning," ICASP, Seoul, South Korea, 2019.
- [41] Z. Hu and X. Du, "Mixed Efficient Global Optimization for Time-Dependent Reliability Analysis," *J. Mech. Des.*, vol. 137, no. 5, pp. 051401–051401–9, May 2015.
- [42] Z. hu and S. Mahadevan, "A Single-Loop Kriging (SILK) Surrogate Modeling for Time-Dependent Reliability Analysis," *J. Mech. Des.*, vol. 138, Apr. 2016.
- [43] Z. Wang and P. Wang, "A Maximum Confidence Enhancement Based Sequential Sampling Scheme for Simulation-Based Design," *J. Mech. Des.*, vol. 136, p. 021006, Feb. 2014.
- [44] B. Gaspar, A. P. Teixeira, and C. Guedes Soares, "Adaptive surrogate model with active refinement combining Kriging and a trust region method," *Reliab. Eng. Syst. Saf.*, vol. 165, pp. 277–291, Sep. 2017.
- [45] Z. Wang and A. Shafieezadeh, "ESC: an efficient error-based stopping criterion for kriging-based reliability analysis methods," *Struct. Multidiscip. Optim.*, vol. 59, no. 5, pp. 1621–1637, May 2019.
- [46] C. Jiang, H. Qiu, L. Gao, D. Wang, Z. Yang, and L. Chen, "Real-time estimation error-guided active learning Kriging method for time-dependent reliability analysis," *Appl. Math. Model.*, vol. 77, pp. 82–98, 2020.

- [47] Schöbi R., Sudret B., and Marelli S., "Rare Event Estimation Using Polynomial-Chaos Kriging," *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part Civ. Eng.*, vol. 3, no. 2, p. D4016002, Jun. 2017.
- [48] J. Wang, Z. Sun, Q. Yang, and R. Li, "Two accuracy measures of the Kriging model for structural reliability analysis," *Reliab. Eng. Syst. Saf.*, vol. 167, pp. 494–505, Nov. 2017.
- [49] "UQLab Kriging (Gaussian process modelling) manual," *UQLab, the Framework for Uncertainty Quantification*. [Online]. Available: <http://www.uqlab.com/userguidekriging>. [Accessed: 13-May-2017].
- [50] I. Kaymaz, "Application of kriging method to structural reliability problems," *Struct. Saf.*, vol. 27, no. 2, pp. 133–151, 2005.
- [51] A. Marrel, B. Iooss, B. Laurent, and O. Roustant, "Calculations of Sobol indices for the Gaussian process metamodel," *Reliab. Eng. Syst. Saf.*, vol. 94, no. 3, pp. 742–751, Mar. 2009.
- [52] Forrester Alexander I.J, Sóbester András, and Keane Andy J, "Multi-fidelity optimization via surrogate modelling," *Proc. R. Soc. Math. Phys. Eng. Sci.*, vol. 463, no. 2088, pp. 3251–3269, Dec. 2007.
- [53] M. C. Kennedy and A. O'Hagan, "Bayesian calibration of computer models," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 63, no. 3, pp. 425–464, 2001.
- [54] B. Bichon, M. Eldred, L. Swiler, S. Mahadevan, and J. McFarland, "Multimodal reliability assessment for complex engineering applications using efficient global optimization," in *48th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, 2007, p. 1946.
- [55] L. L. Cam, "An approximation theorem for the Poisson binomial distribution," *Pac. J. Math.*, vol. 10, no. 4, pp. 1181–1197, 1960.
- [56] O. T. Johnson, *Information Theory and the Central Limit Theorem*. World Scientific, 2004.
- [57] L. LeCam, "On the Distribution of Sums of Independent Random Variables," in *Bernoulli 1713 Bayes 1763 Laplace 1813*, Springer, Berlin, Heidelberg, 1965, pp. 179–202.
- [58] W. Hoeffding and H. Robbins, "The central limit theorem for dependent random variables," *Duke Math. J.*, vol. 15, no. 3, pp. 773–780, 194809.
- [59] P. Ranjan, D. Bingham, and G. Michailidis, "Sequential experiment design for contour estimation from complex computer codes," *Technometrics*, vol. 50, no. 4, pp. 527–541, 2008.
- [60] S. Song, Z. Lu, and H. Qiao, "Subset simulation for structural reliability sensitivity analysis," *Reliab. Eng. Syst. Saf.*, vol. 94, no. 2, pp. 658–665, Feb. 2009.
- [61] V. Dubourg and B. Sudret, "Meta-model-based importance sampling for reliability sensitivity analysis," *Struct. Saf.*, vol. 49, pp. 27–36, Jul. 2014.
- [62] J.-M. Bourinet, F. Deheeger, and M. Lemaire, "Assessing small failure probabilities by combined subset simulation and Support Vector Machines," *Struct. Saf.*, vol. 33, no. 6, pp. 343–353, Sep. 2011.
- [63] D. Vincent, "Adaptive surrogate models for reliability analysis and reliability-based design optimization." 2011.
- [64] Z. Hu and S. Mahadevan, "Global sensitivity analysis-enhanced surrogate (GSAS) modeling for reliability analysis," *Struct. Multidiscip. Optim.*, vol. 53, no. 3, pp. 501–521, Mar. 2016.