Highlights

Parallel active learning reliability analysis: A multi-point look-ahead paradigm

Tong Zhou, Tong Guo, Chao Dang, Lei Jia, You Dong

- Overall error of active learning reliability analysis is decomposed into two separate parts.
- The probability density evolution method is deployed in an 'over-kill' manner.
- A measure of epistemic uncertainty about Kriging-based failure probability estimation is proved.
- A multi-point look-ahead learning function is analytically deduced.
- The number of new samples can be either prescribed or adaptively determined per iteration.
- Multi-point enrichment process is based on learning function itself, without additional strategies.

Parallel active learning reliability analysis: A multi-point look-ahead paradigm

Tong Zhou[a], Tong Guo[b], Chao Dang[c], Lei Jia[d,e,*], You Dong[f]

[a]*Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, China*
[b]*School of Civil Engineering, Southeast University, Nanjing 210098, China*
[c]*Chair for Reliability Engineering, TU Dortmund University, Dortmund 44227, Germany*
[d]*School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 611731, Sichuan, China*
[e]*Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China, Shenzhen 518110, Guangdong, China*
[f]*Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hong Kong, China*

Abstract

To alleviate the intensive computational burden of reliability analysis, a new parallel active learning reliability method is proposed from the multi-point look-ahead paradigm. First, in the framework of probability density evolution method, a global measure of epistemic uncertainty about Kriging-based failure probability estimation, referred to as targeted integrated mean squared error (TIMSE), is defined and well proved. Then, three key ingredients are developed in the workflow of parallel active learning reliability analysis: (i) A look-ahead learning function called k-point targeted integrated mean square error reduction (k-TIMSER) is deduced in closed form, quantifying explicitly the reduction of TIMSE induced by adding a batch of $k(\geq 1)$ new points in expectation. (ii) A hybrid convergence criterion is specified according to the actual reduction of TIMSE at each iteration. (iii) Both prescribed scheme and adaptive scheme are devised to identify the rational size of batch of new points added per iteration. The most distinctive feature of the proposed approach lies in that the multi-point enrichment process is fully guided by the learning function k-TIMSER itself, without resorting to additional batch selection strategies. Hence, it is much more theoretically elegant and easy to implement. The effectiveness of the proposed approach is testified on three examples, and comparisons are made against several existing reliability methods. The results show that the proposed method achieves fair superiority over other existing ones in terms of the accuracy of failure probability estimate and the number of iterations. Particularly, the advantage of the total computational time becomes very evident in the proposed method, when computationally-expensive reliability problems are considered.

Keywords: Parallel active learning reliability analysis, Multi-point look-ahead paradigm, Probability density evolution method, Kriging, Epistemic uncertainty

Nomenclature

AK-MCS	adaptive Kriging-Monte Carlo simulation
ALR	active learning reliability
CDF	cumulative distribution function
COV	coefficient of variation
ED	experimental design
EFF	expected feasibility function
EM	ensemble of metamodels
EPDF	evolutionary probability density function
GDEE	generalized probability density evolution equation
IS	importance sampling
MCS	Monte Carlo simulation
PABQ	parallel adaptive Bayesian quadrature
PDEM	probability density evolution method
PDF	probability density function
PEIF	PDEM-oriented expected improvement function
PIE	PDEM-oriented information entropy
ROI	regions of interest
SuS	subset simulation
TIMSE	targeted integrated mean squared error
TIMSER	targeted integrated mean square error reduction

1. Introduction

Probabilistic reliability analysis aims to quantify the failure probability of an engineering system with respect to some relevant failure criteria, accounting for various uncertainties in the physical properties, external loads and

*Corresponding author.
Email addresses: tong.ce.zhou@outlook.com (Tong Zhou), guotong@seu.edu.cn (Tong Guo), chao.dang@tu-dortmund.de (Chao Dang), jialeishenzhen@gmail.com (Lei Jia), you.dong@polyu.edu.hk (You Dong)

environmental conditions, etc. Nowadays, it has emerged as a core task in the design, performance assessment and maintenance decision-making of complex engineering structures and infrastructures [1].

The uncertainties in engineering systems are generally represented through a set of d continuous random variables $\boldsymbol{X} = \{X_1, \ldots, X_d\} \in \mathbb{X} \subset \mathbb{R}^d$, with a joint probability density function (PDF) $f_{\boldsymbol{X}}(\boldsymbol{x})$. Then, a single structural response or the extreme value of multiple structural responses is often of interest to reliability analysis and is a function of $\boldsymbol{x}$, with $\boldsymbol{x}$ one realization of $\boldsymbol{X}$. Such input-output mapping can be generalized as a computational model $y = \mathcal{M}(\boldsymbol{x}), \mathcal{M} : \mathbb{X} \in \mathbb{R}^d \mapsto \mathbb{R}^1$. Basically, a single evaluation of $\mathcal{M}(\cdot)$ is very computationally expensive, especially when a high-fidelity finite element model is involved. Conventionally, the failure domain is defined as $\mathbb{F} = \{\boldsymbol{x} \in \mathbb{X} : \mathcal{M}(\boldsymbol{x}) \geq h\}$, with h the associated failure threshold. Then, the failure probability P_f of this system can be defined as

$$P_f = \mathbb{P}(Y \geq h) = \begin{cases} \displaystyle\int_{\mathbb{X}} \mathbf{1}_{\mathbb{F}}(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}, & \text{routine ①}, \\ \displaystyle\int_h^{\infty} f_Y(y) \mathrm{d}y, & \text{routine ②}, \end{cases} \tag{1}$$

where $\mathbb{P}(\cdot)$ represents the probability measure; $\mathbf{1}_{\mathbb{F}}(\boldsymbol{x}) = \begin{cases} 1, & \boldsymbol{x} \in \mathbb{F} \\ 0, & \text{otherwise} \end{cases}$ denotes a failure indicator function; $f_Y(y)$ is the PDF of Y.

Eq. (1) indicates that the computation of P_f can be essentially categorized into two distinct routines. The *routine* ① is dedicated to tackling with a d-dimensional integral in the input space ($\boldsymbol{X}$-space). Then, typical reliability methods in this category include analytical approximation methods, e.g., first- and second-order reliability methods [2, 3], and simulation methods, e.g., Monte Carlo simulation (MCS) [4], importance sampling (IS) [5], line sampling [6], and subset simulation (SuS) [7]. The *routine* ② is concerned with dealing with a one-dimensional integral in the output space (Y-space). Then, typical reliability methods in this category include moment methods [8, 9] and probability density evolution method (PDEM) [10, 11]. The PDEM is fully non-parametric and highly flexible, yielding favorable performances in both static and dynamic reliability problems [12]. However, the computational burden of PDEM is still relatively intensive to practical engineering problems.

In the past two decades, active learning reliability (ALR) methods [13] have constituted an active field of research, owing to its favorable computational efficiency over those traditional reliability methods above. Its core lies in replacing the actual computational model with a cheap-to-evaluate surrogate model, which is used in conjunction with a reliability estimation algorithm to provide an estimate of failure probability. Then, guided by a learning function, the experimental design (ED) for the surrogate model is sequentially enriched with new training samples and their model responses to improve the accuracy of failure probability estimate progressively. This sequential enrichment process is ended when a relevant convergence criterion is met. Hence, the ALR framework is comprised of four main ingredients, namely surrogate model, reliability estimation algorithm, learning function, and convergence criterion [14]. Commonly-used surrogate models include polynomial chaos expansion [15, 16], Kriging [17, 18], support vector regression [19, 20], radial basis function [21], and ensemble of metamodels (EM) [22], to name just a few. Kriging is the most popular method for its exact interpolation property and the ability to quantify epistemic uncertainty of prediction. Efficient global reliability analysis [23] and adaptive Kriging-Monte Carlo simulation (AK-MCS) [24] are two pioneering contributions in the framework of Kriging-based active learning reliability analysis. Then, the active learning methods combining Kriging with those traditional reliability estimation methods, e.g., MCS [25], IS [14], SuS [26] or PDEM [27, 28], have been well developed in the literature. The interested readers are referred to [13, 14] for a comprehensive review of this topic.

Sequential experimental design process is the most prominent feature of the ALR methods. It is essentially a problem of sequential decision making under uncertainty [29]. This process is often achieved by defining a learning function, which assigns a score to each candidate point reflecting its preference over others for the next evaluation of computational model. In the context of active learning-based simulation methods, common Kriging-based learning functions include the U function [24], expected feasibility function (EFF) [23], stepwise uncertainty reduction [30, 31], expected uncertainty reduction [25], expected margin volume reduction [32], stepwise margin reduction [33], expected integrated error reduction [34], and integrated probability of misclassification [35]. Then, in the context of active learning-based PDEM, common Kriging-based learning functions include the PDEM-oriented information entropy (PIE) [27], PDEM-oriented expected improvement function (PEIF) [28], and stepwise truncated variance reduction [36]. Note that single-point enrichment process is generally used with these aforementioned learning functions, due to their inherent point-to-point nature. This bottleneck makes the overall computational time of active learning reliability methods still very intensive, especially when considering the large-scale engineering problems.

Recently, parallel computing has drawn increasing attention due to its potential of speeding up the overall computation and maximizing the available computing resources [37]. Hence, parallel active learning reliability

methods that allow adding multiple new points per iteration have been explored by researchers, so as to reduce the total computational time dramatically. This goal is often achieved by integrating a single-point learning function with some additional multi-point selection strategies. Those strategies can be basically categorized into the following four groups. (i) *The clustering strategy*. The new samples are selected as the centroids of K clusters of candidate points weighted by the learning function values. Popular clustering techniques include the K-means clustering strategy [38, 39, 40, 41], the K-medoids clustering strategy [42, 43], or the combination of them [44]. However, it is often challenging to specify a reasonable value of K in advance. (ii) *Kriging believer or constant liar strategy* [45]. The first new point is selected by the learning function; then, the Kriging is retrained by adding a new point and the Kriging mean at this point [38, 46, 47] or a constant value fixed by the user [48]; next, the learning function can be updated to add the second new point. This process iterates until K new points are selected. Clearly, there are a total of K runs of retraining Kriging per iteration. (iii) *EM-based strategy* [45]. The EM is a weighted average metamodel that combines K distinct types of metamodel. Then, a learning function is used with each metamodel within the EM to select a total of K new points, at most, per iteration [22]. Obviously, the number of new samples is restricted by the amount of metamodels considered in the EM. (iv) *Pseudo learning function strategy* [49]. The first new point is selected by the learning function; then, the learning function is updated by multiplying itself by an influence function of the first new point, and the second new sample can be thus selected. This process iterates until K new samples are selected [50, 51]. However, the influence function is often expressed as a kernel function, which may be too empirical to reflect the actual impact of adding a new point on the learning function.

In this study, a new parallel active learning reliability method is proposed from the multi-point look-ahead perspective. Its distinctive feature lies in that the multi-point enrichment process is directly conducted based on the learning function itself, eliminating the need for additional parallel selection procedures. The major contributions of this study can be summarized as follows.

- Based on the overall error analysis of active learning reliability framework, a metric called targeted integrated mean squared error (TIMSE) is proved, for the first time, as the upper bound of Kriging-induced error in the PDEM. Hence, the TIMSE acts as a measure of epistemic uncertainty about Kriging-based failure probability estimation and guides the deployment of those key ingredients of active learning paradigm.

- With the aim of minimizing the TIMSE, a goal-oriented learning function called k-point targeted integrated mean square error reduction (k-TIMSER) is derived in closed form. It quantifies the expected impact of adding a batch of $k(\geq 1)$ new points on the reduction of TIMSE. Notably, the k-TIMSER is the first multi-point learning function in the framework of active learning-based PDEM.

- In the multi-point enrichment process, the number of new points added per iteration can be specified either by a prescribed scheme or by an adaptive scheme, thanks to the good ability of k-TIMSER to quantify the expected gain of adding each new point in the batch. Notably, the proposed adaptive scheme is also the first one in the parallel active learning reliability framework.

The rest of this paper is arranged as follows. Section 2 reviews some basic concepts. Section 3 is dedicated to the multi-point learning function k-TIMSER. Section 4 details the parallel active learning reliability analysis based on k-TIMSER. The efficacy of the proposed approach is illustrated in Section 5 through three examples. Finally, some concluding remarks are given in Section 6.

2. Preliminaries

Section 2.1 revisits structural reliability analysis based on the PDEM. Section 2.2 reviews the regions of interest (ROI) in the PDEM. Section 2.3 details the overall error analysis of active learning reliability framework. Section 2.4 states the motivation of this study.

2.1. Probability density evolution method (PDEM)

Following the routine ② in Eq. (1), the core of PDEM lies in deriving the PDF $f_Y(y)$ of Y. To achieve this goal, a virtual stochastic process $V(\tau)$ associated with Y needs to be constructed, typically expressed as [12]

$$V(\tau) = Y \sin\left(\frac{5\pi}{2}\tau\right), \tau \in [0,1], \tag{2}$$

where τ is a virtual time parameter. One may easily find that $V(\tau)|_{\tau=0} = 0$, and $V(\tau)|_{\tau=1} = Y$. Hence, $f_Y(y)$ equals the evolutionary probability density function (EPDF) of $V(\tau)$, denoted as $f_V(v,\tau)$, at the ending instant $\tau = 1$, i.e.,

$$f_Y(y) = f_V(v,\tau)|_{v=y,\tau=1}. \tag{3}$$

Eq. (2) indicates that the augmented system $[V(\tau), \boldsymbol{X}]$ is probability-preserved, owing to all its randomness coming from $\boldsymbol{X}$. After a series of mathematical manipulations, the so-called generalized probability density evolution equation (GDEE) reads [10]

$$\begin{cases} \dfrac{\partial f_{V\boldsymbol{X}}(v,\boldsymbol{x},\tau)}{\partial \tau} + \dot{V}(\boldsymbol{X},\tau)\dfrac{\partial f_{V\boldsymbol{X}}(v,\boldsymbol{x},\tau)}{\partial v} = 0, \\ f_{V\boldsymbol{X}}(v,\boldsymbol{x},\tau)|_{\tau=0} = \delta(v) f_{\boldsymbol{X}}(\boldsymbol{x}), \end{cases} \tag{4}$$

where $f_{V\boldsymbol{X}}(v,\boldsymbol{x},\tau)$ is the joint PDF of $[V(\boldsymbol{\tau}), \boldsymbol{X}]$; $\dot{V}(\cdot)$ is the derivative of $V(\tau)$ with respect to τ; $\delta(\cdot)$ is the Dirac's delta function.

Once Eq. (4) is solved, $f_V(v,\tau)$ is obtained as

$$f_V(v,\tau) = \int_{\mathbb{X}} f_{V\boldsymbol{X}}(v,\boldsymbol{x},\tau)\mathrm{d}\boldsymbol{x}. \tag{5}$$

Then, $f_Y(y)$ is gained from Eq. (3), and P_f is finally evaluated by Eq. (1).

Clearly, the crux of PDEM lies in solving the GDEE (Eq. (4)), and the numerical workflow consists of four main steps as follows.

(1) **Partition of probability space**.
According to the GF discrepancy-based strategy [52], generate a set of n_{rp} representative points, denoted as $\mathcal{X}_{\mathrm{rp}} = \{\boldsymbol{x}^{(i)}, i = 1,\ldots,n_{\mathrm{rp}}\}$. Meanwhile, their assigned probabilities are denoted as $\mathcal{P}_{\mathrm{rp}} = \{p^{(i)}, i = 1,\ldots,n_{\mathrm{rp}}\}$, with the $p^{(i)}$ computed as

$$p^{(i)} = \int_{\mathbb{X}^{(i)}} f_{\boldsymbol{X}}(\boldsymbol{x})\mathrm{d}\boldsymbol{x},\ i = 1,\ldots,n_{\mathrm{rp}}, \tag{6}$$

where $\mathbb{X}^{(i)}$ represents the Voronoi cell of $\boldsymbol{x}^{(i)}$. For illustration, a set of well-selected representative points and their associated assigned probabilities are displayed in the top panel of Fig. 1(a).

(2) **Computational model evaluations**.
Evaluate the computational model $\mathcal{M}(\cdot)$ on each $\boldsymbol{x}^{(i)} \in \mathcal{X}_{\mathrm{rp}}$ to obtain the concerned quantity $y^{(i)} = \mathcal{M}(\boldsymbol{x}^{(i)})$. Then, they are collected as $\mathcal{Y}_{\mathrm{rp}} = \{y^{(i)}, i = 1,\ldots,n_{\mathrm{rp}}\}$.

(3) **Discretization of GDEE**.
For $\boldsymbol{x}^{(i)}, i = 1,\ldots,n_{\mathrm{rp}}$, the GDEE in Eq. (4) can be discretized as [28]

$$\begin{cases} \dfrac{\partial f_V^{(i)}(v,\tau)}{\partial \tau} + \dot{V}(\boldsymbol{x}^{(i)},\tau)\dfrac{\partial f_V^{(i)}(v,\tau)}{\partial v} = 0, \\ f_V^{(i)}(v,\tau)|_{\tau=0} = \delta(v)p^{(i)}, \end{cases} \tag{7}$$

where $f_V^{(i)}(v,\tau)$ represents the partial EPDF of $V(\tau)$ conditional on $\boldsymbol{x}^{(i)}$, expressed as

$$f_V^{(i)}(v,\tau) = \int_{\mathbb{X}^{(i)}} f_{V\boldsymbol{X}}(v,\boldsymbol{x},\tau)\mathrm{d}\boldsymbol{x}. \tag{8}$$

Then, Eq. (7) can be solved by the finite difference method, e.g., the total variation diminishing scheme [1], resulting in $\{f_V^{(i)}(v,\tau), i = 1,\ldots,n_{\mathrm{rp}}\}$.

(4) **Computation of failure probability**.
Based on the results of Step (3), the EPDF $f_V(v,\tau)$ in Eq. (5) is assembled as

$$f_V(v,\tau) \approx \sum_{i=1}^{n_{\mathrm{rp}}} f_V^{(i)}(v,\tau). \tag{9}$$

Then, substituting Eq. (9) into Eq. (3), $f_Y(y)$ is given as

$$f_Y(y) \approx \left(\sum_{i=1}^{n_{\mathrm{rp}}} f_V^{(i)}(v,\tau)\right)\Bigg|_{v=y,\tau=1} = \sum_{i=1}^{n_{\mathrm{rp}}}\left(f_V^{(i)}(v,\tau)\Big|_{v=y,\tau=1}\right) = \sum_{i=1}^{n_{\mathrm{rp}}} f_Y^{(i)}(y), \tag{10}$$

where $f_Y^{(i)}(y) = f_V^{(i)}(v,\tau)|_{v=y,\tau=1}$ denotes the partial PDF of Y conditional on $\boldsymbol{x}^{(i)}$. Further, substituting Eq. (10) into Eq. (1) yields

$$P_f \approx \widehat{P}_f = \int_h^{+\infty}\left(\sum_{i=1}^{n_{\mathrm{rp}}} f_Y^{(i)}(y)\right)\mathrm{d}y = \sum_{i=1}^{n_{\mathrm{rp}}}\int_h^{+\infty} f_Y^{(i)}(y)\mathrm{d}y = \sum_{i=1}^{n_{\mathrm{rp}}} \widehat{P}_f^{(i)}, \tag{11}$$

where $\widehat{P}_f$ denotes the failure probability computed by the PDEM; $\widehat{P}_f^{(i)} = \int_h^{+\infty} f_Y^{(i)}(y)\mathrm{d}y$ is the partial failure probability brought by $\boldsymbol{x}^{(i)}$.

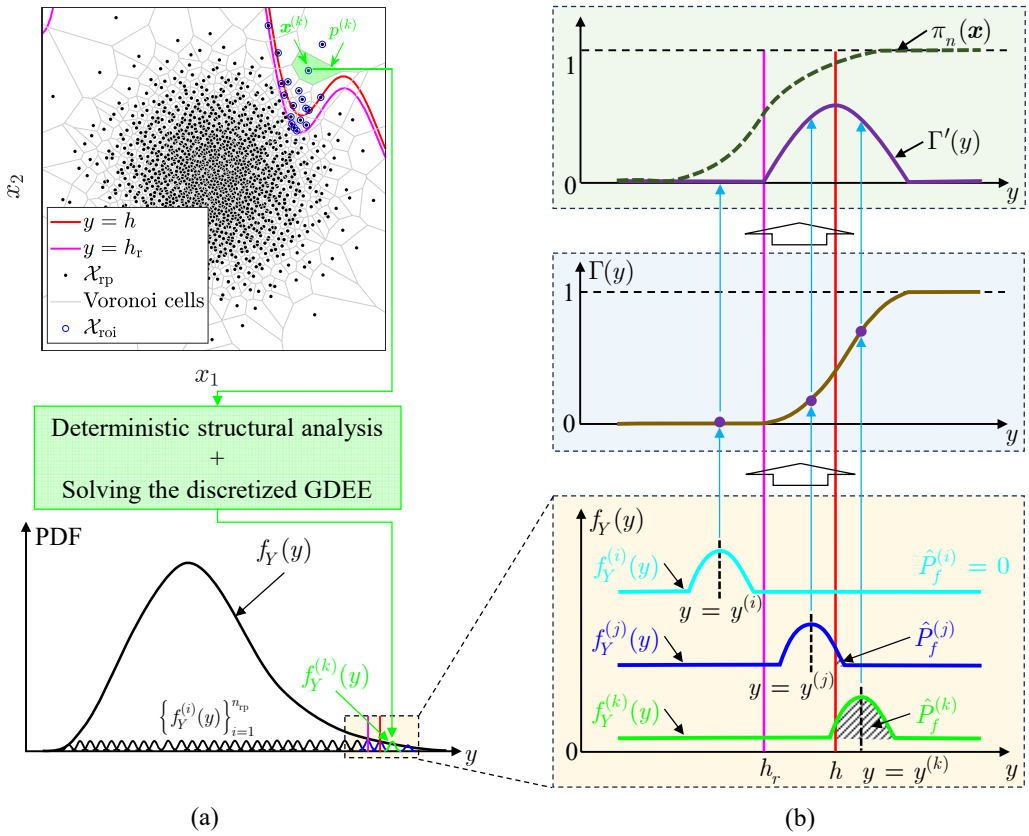


Figure 1: Schematic of the computation of $\widehat{P}_f$ in PDEM

2.2. Regions of interest (ROI)

The summation in Eq. (10) is illustrated in the bottom panel of Fig. 1(a). Generally, $f_Y^{(i)}(y)$ centers at the vertical coordinate $y = y^{(i)}$ and decays quickly with the distance to this coordinate. Hence, $f_Y^{(i)}(y)$ is only significant in the vicinity of $y = y^{(i)}$. In this way, only when $y^{(i)}$ is close to or within $[h, +\infty]$, the $\widehat{P}_f^{(i)}$ will gain significant value, as exemplified by three representative points in the bottom panel of Fig. 1(b).

Besides, one may find that

$$\int_{-\infty}^{+\infty} f_Y^{(i)}(y)\mathrm{d}y = \int_{-\infty}^{+\infty} f_V^{(i)}(v,\tau)|_{\tau=1}\mathrm{d}v = \int_{-\infty}^{+\infty}\int_{\mathbb{X}^{(i)}} f_{V\boldsymbol{X}}(v,\boldsymbol{x},\tau)|_{\tau=1}\mathrm{d}\boldsymbol{x}\mathrm{d}v = p^{(i)}, \tag{12}$$

which implies that $\widehat{P}_f^{(i)}$ is only a portion of $p^{(i)}$, with the proportional coefficient depending on $y^{(i)}$.

In this way, $\widehat{P}_f^{(i)}$ can be reformulated as

$$\widehat{P}_f^{(i)} = \Gamma\big(y^{(i)}\big)p^{(i)}, \tag{13}$$

where the proportional coefficient $\Gamma(\cdot) \in [0,1]$ is a monotonically non-decreasing function with respect to y; see the middle panel of Fig. 1(b).

Clearly, only the representative point $\boldsymbol{x}$ featuring $\Gamma\big(y\big) > 0$ contributes significantly to $\widehat{P}_f$. Then, the local regions covered by those critical points are called the regions of interest (ROI), defined as [28]

$$\mathbb{X}_{\mathrm{R}} = \{\boldsymbol{x} \in \mathbb{X} : \mathcal{M}(\boldsymbol{x}) \geq h_{\mathrm{r}}\}, \tag{14}$$

where $y = h_{\mathrm{r}}$ is the boundary of ROI, as plotted as the magenta line in Fig. 1.

According to Eqs. (11) and (14), h_{r} has to satisfy the following expression

$$\frac{\widehat{P}_f - \sum_{i=1}^{n_{\mathrm{rp}}} \widehat{P}_f^{(i)}\mathbb{1}\big(\mathcal{M}\left(\boldsymbol{x}^{(i)}\right) \geq h_{\mathrm{r}}\big)}{\widehat{P}_f} \leq \varepsilon_{\mathrm{r}}, \tag{15}$$

where $\mathbb{1}(\cdot)$ is an indicator function that equals 1 if the bracketed event is true, and 0 otherwise; ε_{r} is a minor tolerance, say 10^{-5}. Eq. (15) indicates that the ratio of failure probability induced by those representative

points in the ROI to $\widehat{P}_f$ is at least $(1-\varepsilon_{\mathrm{r}})\times 100\% = 99.999\%$. Hence, the impact of those representative points outside the ROI is almost negligible.

Moreover, it can be found that once $\mathcal{Y}_{\mathrm{rp}} = \{y^{(i)}, i = 1, \ldots, n_{\mathrm{rp}}\}$ is available, the h_{r} can be identified by trial and error. The associated workflow is given in Appendix A.

2.3. Overall error analysis of active learning reliability framework

To improve the computational efficiency of PDEM, the active learning methods that combine Kriging and PDEM have been developed by the first author and his co-workers [27, 28]. For the sake of brevity, the basics of Kriging $\widehat{\mathcal{M}}_n(\boldsymbol{x})$ are outlined in Appendix B. Then, two existing Kriging-based learning functions, i.e., PIE [27] and PEIF [28], are outlined in Appendix C.

To show the motivation of this study, Fig. 2 illustrates the overall error analysis of a general ALR framework, where surrogate model (e.g., Kriging) can be combined either with PDEM or with simulation methods (e.g., MCS, IS, or SuS). According to the sources of uncertainty, the overall error of failure probability estimation can be viewed as a compound of two categories of error as follows.

- The first category of error arises from the reliability estimation algorithm being used. Due to the aleatoric uncertainty in $\boldsymbol{X}$, those reliability estimation algorithms have to numerically compute Eq. (1), giving rise to approximation error. For example, the PDEM-caused error is mainly brought by the partition of probability space in Eq. (6) and further the discretization of GDEE in Eq. (7). Similarly, the error in those simulation methods, e.g., MCS and SuS, is mainly due to the random sampling.

- The second category of error is brought by introducing surrogate model, e.g., Kriging. In essence, this error is the epistemic uncertainty of failure probability estimation, due to only n evaluations of $\mathcal{M}(\cdot)$ (very limited) in the ED $\mathcal{D}_n$, rather than the exact knowledge of $\mathcal{M}(\cdot)$. For example, the epistemic uncertainty of Kriging $\widehat{\mathcal{M}}_n(\boldsymbol{x})$ is encapsulated by its posterior variance $\sigma_n^2(\boldsymbol{x})$, which will further propagate to the epistemic uncertainty of Kriging-based failure probability estimation.

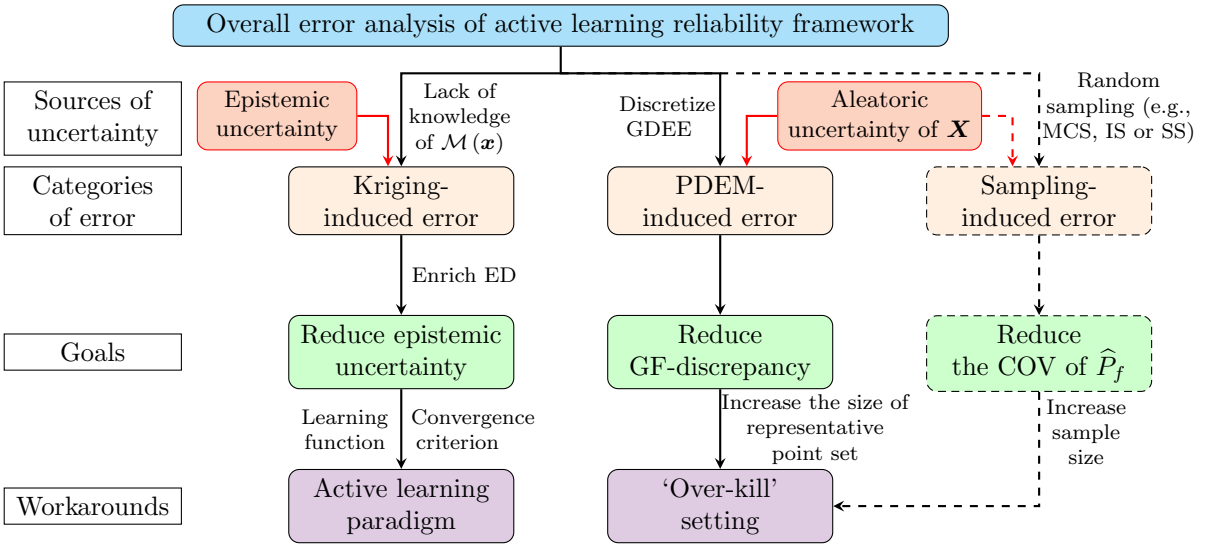


Figure 2: Overall error analysis of active learning reliability framework

To reduce the overall error of failure probability estimation in the ALR framework, great efforts should be made to alleviate those two categories of error jointly, i.e.,

- To reduce the approximation error, the so-called 'over-kill' setting [14] is advocated for the reliability estimation algorithm in the ALR framework. Specifically, the parameters of the reliability estimation algorithm are tuned to dramatically reduce the variation of the failure probability estimate. For example, the number n_{rp} of representative points in PDEM shall be significantly increased to reduce the corresponding GF discrepancy. Hence, the n_{rp} is often set as $\mathcal{O}(10^3)$, instead of the traditional $\mathcal{O}(10^2)$ [28]. Similarly, to reduce the coefficient of variation (COV) of failure probability estimate in SuS, the number n_{s} of samples at each subset and the conditional probability p_0 should be increased significantly. Hence, they are generally set as $\mathcal{O}(10^5)$ and $0.2-0.3$, respectively, rather than 10^3 and 0.1 in the usual setting [14]. Although the number of generated samples will increase to some extent, this overhead can be cheaply covered by the relatively negligible expense of surrogate model to provide predictions.

- To reduce surrogate-induced error, the epistemic uncertainty about failure probability estimation can be substantially reduced by adding more and more new samples. This can be readily achieved by the active learning workflow, where a learning function selects a sequence of new samples that reduce a specific form of epistemic uncertainty measure at most.

Obviously, the first aspect can be readily realized with the off-the-shelf procedures. Then, the key of the second part lies in how to build the main ingredients of the ALR framework, e.g., learning function and convergence criterion, so as to reduce the epistemic uncertainty of failure probability with high efficiency.

2.4. Motivation of this study

From the perspective of Section 2.3, those previous studies are flawed in the following three aspects.

- *Lack of an explicit link with the global epistemic uncertainty of* $\widehat{P}_f$. Existing learning functions, e.g., PIE (Eq. (C.3)) and PEIF (Eq. (C.4)), are directly defined from some *local* uncertainty metrics about Kriging prediction at a single point $\boldsymbol{x}$, rather than from the *global* residual uncertainty of $\widehat{P}_f$ (a joint effect of all points in the probability space). Besides, the convergence criterion is sometimes defined according to the maximum/minimum value of learning function, which only reflects the local uncertainty measure of $\widehat{P}_f$, rather than the global effect, and is too conservative [27, 39]. Hence, existing active learning workflow is not goal-oriented somewhat, and is not as parsimonious as possible.

- *Inability of quantifying the gain of adding a new sample.* Due to the first drawback, those existing learning functions are unable to measure the impact of adding a new sample on reducing the epistemic uncertainty of $\widehat{P}_f$. This drawback is also reflected by their expressions, which are functions of the predictive mean $\mu_n(\cdot)$ and variance $\sigma_n^2(\cdot)$ of Kriging at a candidate point $\boldsymbol{x}$, without involving the covariance $c_n(\boldsymbol{x}, \boldsymbol{x}')$ of Kriging. In essence, they ignore the effect of adding a new point $\boldsymbol{x}$ on any other point around it. Note that this limitation is also observed in the learning function U (Eq. (C.1)) and EFF (Eq. (C.2)) in the AK-MCS. In fact, those learning functions are collectively called as pointwise learning functions in [53, 32].

- *Need of additional multi-point selection procedures.* Due to their point-to-point nature, those existing learning functions cannot work directly with a batch of $k(>1)$ new samples. Then, the parallel enrichment process has to be conducted by combining a pointwise learning function with an additional multi-point procedure, e.g., the K-means clustering strategy. Given the second drawback, it is impossible for such practice to quantify the total impact of adding a batch of $k(>1)$ new samples on reducing the epistemic uncertainty of $\widehat{P}_f$.

To summarize, although existing (parallel) active learning reliability methods may provide favorable empirical results, they are not theoretically sound and consistent to some extent. Hence, there is still great room for improvement.

In this study, a new parallel active learning reliability method is developed, fully complying with the theoretical philosophy of Fig. 2. First, a measure of epistemic uncertainty of Kriging-based failure probability estimation in the framework of PDEM is defined. With the aim of minimizing such epistemic uncertainty, three key ingredients, i.e., learning function, convergence criterion, and the number of new samples added per iteration, are built in a goal-oriented way. Thanks to the core role of such epistemic uncertainty measure, those main ingredients are consistently assembled, as illustrated in Fig. 3. To the best of the authors' knowledge, such theoretical philosophy has not been explored in the field of parallel active learning-based PDEM, and is totally different from those state-of-the-art approaches. More details will be provided in Sections 3 and 4.

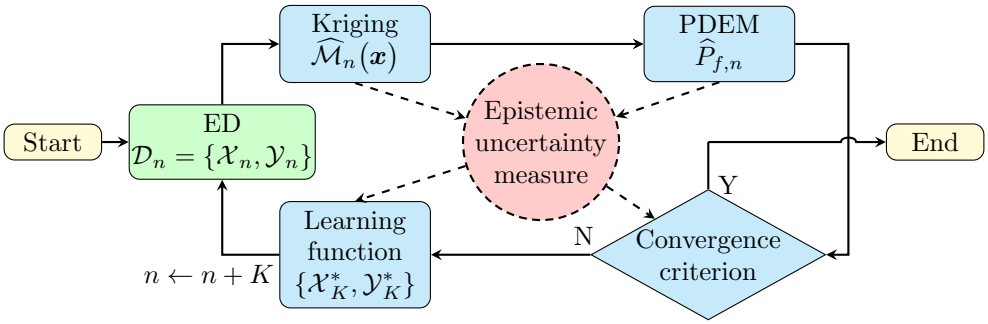


Figure 3: Theoretical philosophy of the proposed multi-point look-ahead paradigm

3. The proposed multi-point learning function

The proposed approach starts by treating the response of computational model $\mathcal{M}(\boldsymbol{x})$ as one possible sample path of Kriging $\widehat{\mathcal{M}}_n(\boldsymbol{x})$. Then, a measure of epistemic uncertainty about Kriging-based failure probability estimation $\widehat{P}_{f,n}$ in the PDEM is defined in Section 3.1. In this way, a new learning function is defined in Section 3.2 by quantifying the expected reduction of this epistemic uncertainty measure induced by adding a batch of k new samples. Finally, numerical implementation of the resulting multi-point enrichment process is detailed in Section 3.3.

3.1. An epistemic uncertainty measure of failure probability estimation

Given an ED $\mathcal{D}_n = \{\mathcal{X}_n, \mathcal{Y}_n\}$ of size n, a Kriging $\widehat{\mathcal{M}}_n(\boldsymbol{x})$ can be readily trained, with its mean $\mu_n(\boldsymbol{x})$, variance $\sigma_n^2(\boldsymbol{x})$ and covariance $c_n(\boldsymbol{x}, \boldsymbol{x}')$ given by Eqs. (B.7), (B.8) and (B.9), respectively. Then, by performing PDEM on the Kriging predictions $\{\widehat{\mathcal{M}}_n(\boldsymbol{x}^{(i)}), i = 1, \ldots, n_{\mathrm{rp}}\}$ at the $\mathcal{X}_{\mathrm{rp}}$, the failure probability estimate can be obtained from Eqs. (11) and (13) such that

$$\widehat{P}_{f,n} = \sum_{i=1}^{n_{\mathrm{rp}}} \widehat{P}_{f,n}^{(i)} = \sum_{i=1}^{n_{\mathrm{rp}}} \Gamma\big(\widehat{\mathcal{M}}_n(\boldsymbol{x}^{(i)})\big) p^{(i)}, \tag{16}$$

where $\widehat{P}_{f,n}^{(i)}, i = 1, \ldots, n_{\mathrm{rp}}$, denotes the partial failure probability estimate conditional on $\widehat{\mathcal{M}}_n(\boldsymbol{x}^{(i)})$.

Since the Kriging $\widehat{\mathcal{M}}_n(\boldsymbol{x})$ is a Gaussian process, $\widehat{P}_{f,n}$ is actually a random variable. Then, if replacing $\widehat{\mathcal{M}}_n(\boldsymbol{x})$ by its posterior mean $\mu_n(\boldsymbol{x})$ in Eq. (16), there exists

$$\tilde{P}_{f,n} = \sum_{i=1}^{n_{\mathrm{rp}}} \tilde{P}_{f,n}^{(i)} = \sum_{i=1}^{n_{\mathrm{rp}}} \Gamma\big(\mu_n(\boldsymbol{x}^{(i)})\big) p^{(i)}, \tag{17}$$

where $\tilde{P}_{f,n}^{(i)}, i = 1, \ldots, n_{\mathrm{rp}}$, denotes the partial failure probability estimate conditional on $\mu_n(\boldsymbol{x}^{(i)})$. Obviously, the $\tilde{P}_{f,n}$ is a deterministic quantity and is a natural estimate of $\widehat{P}_{f,n}$.

Although $\tilde{P}_{f,n}$ (Eq. (17)) is not the mean of $\widehat{P}_{f,n}$ (Eq. (16)), it satisfies the following proposition.

Proposition 1. *Denote the targeted integrated mean squared error (TIMSE) as*

$$H_n = n_{\mathrm{rp}} \sum_{i=1}^{n_{\mathrm{rp}}} (p^{(i)})^2 \sigma_n^2(\boldsymbol{x}^{(i)}) \pi_n(\boldsymbol{x}^{(i)}), \tag{18}$$

where $\pi_n(\boldsymbol{x})$ represents the probability of $\boldsymbol{x}$ being located in the ROI, expressed as

$$\pi_n(\boldsymbol{x}) = \mathbb{P}\Big(\widehat{\mathcal{M}}_n(\boldsymbol{x}) \geq h_{\mathrm{r}}\Big) = \Phi\left(\frac{\mu_n(\boldsymbol{x}) - h_{\mathrm{r}}}{\sigma_n(\boldsymbol{x})}\right), \tag{19}$$

where $\Phi(\cdot)$ is the cumulative distribution function (CDF) of a standard Gaussian random variable. Then, there holds

$$\mathbb{E}_n\left[\left(\widehat{P}_{f,n} - \tilde{P}_{f,n}\right)^2\right] \leq H_n, \tag{20}$$

where $\mathbb{E}_n[\cdot]$ denotes the expectation with respect to $\widehat{\mathcal{M}}_n(\boldsymbol{x})$.

For the sake of brevity, the proof of Proposition 1 is provided in Appendix D. Eq. (20) implies that when the TIMSE $H_n \to 0$, $\tilde{P}_{f,n}$ converges to the failure probability $\widehat{P}_f$ (Eq. (11)) in expectation. Hence, the H_n is a measure of epistemic uncertainty of $\tilde{P}_{f,n}$, arising from only n evaluations of computation model within the ED $\mathcal{D}_n$. To improve the accuracy of $\tilde{P}_{f,n}$, the H_n shall be dramatically reduced by adding more and more new samples. Following this, a learning function that explicitly quantifies the expected reduction of H_n brought by adding a batch of new points will be developed in Section 3.2.

3.2. The learning function k-TIMSER

First, the H_n can be readily computed based on the current ED $\mathcal{D}_n$; see Eq. (18). Then, assume that the $\mathcal{D}_n$ is enriched with a batch of k new points $\mathcal{X}_k^+ = \{\boldsymbol{x}_+^{(1)}, \ldots, \boldsymbol{x}_+^{(k)}\}^\top$ and their associated responses $\mathcal{Y}_k^+ = \{y_+^{(1)}, \ldots, y_+^{(k)}\}^\top$, the posterior of Kriging can be updated according to the multi-point Kriging update formulas (Appendix E), with the look-ahead mean $\mu_{n+k}(\boldsymbol{x})$, variance $\sigma_{n+k}^2(\boldsymbol{x})$ and covariance $c_{n+k}(\boldsymbol{x}, \boldsymbol{x}')$ given by Eqs. (E.1), (E.2) and (E.3), respectively.

In this way, the future TIMSE, arising from the addition of $\{\mathcal{X}_k^+, \mathcal{Y}_k^+\}$, can be expressed as

$$H_{n+k}(\mathcal{X}_k^+, \mathcal{Y}_k^+) = n_{\mathrm{rp}} \sum_{i=1}^{n_{\mathrm{rp}}} (p^{(i)})^2 \sigma_{n+k}^2(\boldsymbol{x}^{(i)}) \pi_{n+k}(\boldsymbol{x}^{(i)}; \mathcal{X}_k^+, \mathcal{Y}_k^+), \tag{21}$$

which is a function of both $\mathcal{X}_k^+$ and $\mathcal{Y}_k^+$; $\pi_{n+k}(\boldsymbol{x}; \mathcal{X}_k^+, \mathcal{Y}_k^+)$ is given by

$$\begin{aligned} \pi_{n+k}(\boldsymbol{x}; \mathcal{X}_k^+, \mathcal{Y}_k^+) &= \Phi\left(\frac{\mu_{n+k}(\boldsymbol{x}) - h_{\mathrm{r}}}{\sigma_{n+k}(\boldsymbol{x})}\right), \\ &= \Phi\left(\frac{\mu_n(\boldsymbol{x}) + c_n(\boldsymbol{x}, \mathcal{X}_k^+)^\top (\mathcal{C}_k^+)^{-1} (\mathcal{Y}_k^+ - \mu_n(\mathcal{X}_k^+)) - h_{\mathrm{r}}}{\sigma_{n+k}(\boldsymbol{x})}\right), \\ &= \Phi\left(\frac{\mu_n(\boldsymbol{x}) - h_{\mathrm{r}}}{\sigma_{n+k}(\boldsymbol{x})} + \frac{c_n(\boldsymbol{x}, \mathcal{X}_k^+)^\top (\mathcal{C}_k^+)^{-1}}{\sigma_{n+k}(\boldsymbol{x})} \left(\mathcal{Y}_k^+ - \mu_n(\mathcal{X}_k^+)\right)\right), \\ &= \Phi\big(a(\boldsymbol{x}) + \boldsymbol{b}(\boldsymbol{x})^\top \mathcal{U}_k^+\big), \end{aligned} \tag{22}$$

where $a(\boldsymbol{x}) = \dfrac{\mu_n(\boldsymbol{x}) - h_{\mathrm{r}}}{\sigma_{n+k}(\boldsymbol{x})}$ and $\boldsymbol{b}(\boldsymbol{x}) = \dfrac{(\mathcal{C}_k^+)^{-1} c_n(\boldsymbol{x}, \mathcal{X}_k^+)}{\sigma_{n+k}(\boldsymbol{x})}$ are a scalar quantity and a $k \times 1$ vector independent of $\mathcal{Y}_k^+$, respectively; $\mathcal{U}_k^+ = \mathcal{Y}_k^+ - \mu_n(\mathcal{X}_k^+)$. Since $\mathcal{Y}_k^+$ will not be exactly known without evaluating the actual computational model $\mathcal{M}(\cdot)$ on $\mathcal{X}_k^+$, the $\pi_{n+k}(\boldsymbol{x}; \mathcal{X}_k^+, \mathcal{Y}_k^+)$ in Eq. (22) and the $H_{n+k}(\mathcal{X}_k^+, \mathcal{Y}_k^+)$ in Eq. (21) are unknown as well.

In view of this, the $\mathcal{Y}_k^+$ is replaced by the Kriging predictions on $\mathcal{X}_k^+$, i.e., $\boldsymbol{Y}_k^+ = \widehat{\mathcal{M}}_n(\mathcal{X}_k^+) \sim \mathcal{N}_k\big(\mu_n(\mathcal{X}_k^+), \mathcal{C}_k^+\big)$. Then, Eq. (21) is reformulated as

$$\mathcal{H}_{n+k}(\mathcal{X}_k^+) = n_{\mathrm{rp}} \sum_{i=1}^{n_{\mathrm{rp}}} (p^{(i)})^2 \sigma_{n+k}^2(\boldsymbol{x}^{(i)}) \Pi_{n+k}(\boldsymbol{x}^{(i)}; \mathcal{X}_k^+), \tag{23}$$

which becomes a function of $\mathcal{X}_k^+$ solely; $\Pi_{n+k}\left(\boldsymbol{x}; \mathcal{X}_k^+\right)$ takes the plain expression of $\pi_{n+k}\left(\boldsymbol{x}; \mathcal{X}_k^+, \mathcal{Y}_k^+\right)$ in Eq. (22), but the related term $\mathcal{U}_k^+$ is replaced by $\boldsymbol{U}_k^+$, that is,

$$\Pi_{n+k}\left(\boldsymbol{x}; \mathcal{X}_k^+\right) = \Phi\big(a(\boldsymbol{x}) + \boldsymbol{b}(\boldsymbol{x})^\top \boldsymbol{U}_k^+\big), \tag{24}$$

where $\boldsymbol{U}_k^+ = \boldsymbol{Y}_k^+ - \mu_n(\mathcal{X}_k^+) \sim \mathcal{N}_k(\boldsymbol{0}, \mathcal{C}_k^+)$.

The potential impact of adding $\mathcal{X}_k^+$ on the TIMSE can be expressed as

$$\Delta\mathcal{H}_{n+k}(\mathcal{X}_k^+) = H_n - \mathcal{H}_{n+k}(\mathcal{X}_k^+), \tag{25}$$

where the 'mathcal' font acts as a reminder that it is not the actual reduction of TIMSE and is essentially a random quantity through $\boldsymbol{U}_k^+$. Hence, Eq. (25) fails to be a deterministic criterion.

In this regard, the learning function is defined by taking the expectation of $\Delta\mathcal{H}_{n+k}(\mathcal{X}_k^+)$ with respect to $\boldsymbol{U}_k^+$, that is,

$$\mathrm{TIMSER}_{n+k}(\mathcal{X}_k^+) = \mathbb{E}_{\boldsymbol{U}_k^+}\left[\Delta\mathcal{H}_{n+k}(\mathcal{X}_k^+)\right] = \mathbb{E}_{\boldsymbol{U}_k^+}\left[H_n - \mathcal{H}_{n+k}\left(\mathcal{X}_k^+\right)\right] = H_n - \mathbb{E}_{\boldsymbol{U}_k^+}\left[\mathcal{H}_{n+k}(\mathcal{X}_k^+)\right]. \tag{26}$$

Then, a batch of k best next points, $\mathcal{X}_k^* = \{\boldsymbol{x}^{(n+1)}, \ldots, \boldsymbol{x}^{(n+k)}\}$, will be exactly the candidate batch maximizing the expected reduction of TIMSE, i.e.,

$$\mathcal{X}_k^* = \underset{\mathcal{X}_k^+ \in \mathcal{X}_{\mathrm{cp}}}{\arg\max}\, \mathrm{TIMSER}_{n+k}(\mathcal{X}_k^+), \tag{27}$$

where $\mathcal{X}_{\mathrm{cp}} = \{\boldsymbol{x}^{(i)}, i = 1, \ldots, n_{\mathrm{cp}}\}$ denotes a candidate pool of size n_{cp} and will be updated per iteration; see Section 4.3.

Clearly, as more and more batches of k new samples are sequentially added by Eq. (27), the TIMSE H_n is expected to be gradually reduced step by step. Hence, the learning function in Eq. (26) is called k-point targeted integrated mean square error reduction (k-TIMSER) here.

Proposition 2. *The closed-form expression of k-TIMSER in Eq. (26) is given as*

$$\mathrm{TIMSER}_{n+k}(\mathcal{X}_k^+) = n_{\mathrm{rp}} \sum_{i=1}^{n_{\mathrm{rp}}} (p^{(i)})^2 \pi_n(\boldsymbol{x}^{(i)}) \big[\sigma_n^2(\boldsymbol{x}^{(i)}) - \sigma_{n+k}^2(\boldsymbol{x}^{(i)})\big], \tag{28}$$

where the impact of $\mathcal{X}_k^+$ is implicitly encoded by $\sigma_{n+k}^2(\boldsymbol{x})$; see Eq. (E.2).

For brevity, the proof of Proposition 2 is given in Appendix F. Further, Eq. (E.2) indicates that Eq. (28) involves internally the following matrix manipulation

$$\left\{\sigma_n^2(\boldsymbol{x}^{(i)})-\sigma_{n+k}^2(\boldsymbol{x}^{(i)})\right\}_{i=1}^{n_{\rm rp}}=\left\{c_n(\boldsymbol{x}^{(i)},\mathcal{X}_k^+)^\top(\mathcal{C}_k^+)^{-1}c_n(\boldsymbol{x}^{(i)},\mathcal{X}_k^+)\right\}_{i=1}^{n_{\rm rp}}=\mathrm{diag}\left(c_n(\mathcal{X}_{\rm rp},\mathcal{X}_k^+)^\top(\mathcal{C}_k^+)^{-1}c_n(\mathcal{X}_{\rm rp},\mathcal{X}_k^+)\right), \tag{29}$$

where $\mathrm{diag}(\cdot)$ returns the diagonal entries of a matrix, and the matrix size in $\mathrm{diag}(\cdot)$ is $n_{\rm rp}\times n_{\rm rp}$. Generally, $n_{\rm rp}$ is $\mathcal{O}(10^{3-4})$, and the resulting matrix size will be relatively significant.

To alleviate this computer memory issue, the following workaround is considered. For notational brevity, Eq. (29) is simplified as '$\mathrm{diag}\left(A^\top B^{-1}A\right)$', where A and B are two matrices of size $k\times n_{\rm rp}$ and $k\times k$, respectively. Then, this matrix operation is performed with the syntax '$\mathrm{sum}\left(A.*\left(B\setminus A\right),1\right)$' in MATLAB, where $.*$ denotes element-wise multiplication, and $\mathrm{sum}(\cdot,1)$ denotes the sum of each column in a matrix. In this way, the maximum matrix size reduces from $n_{\rm rp}\times n_{\rm rp}$ to $n_{\rm rp}\times k$, which is much easier to handle.

Remark 1. *In comparison with those existing learning functions, e.g., PIE and PEIF, the proposed k-TIMSER is expressed as a function of the mean $\mu_n(\boldsymbol{x})$, variance $\sigma_n^2(\boldsymbol{x})$, and covariance $c_n(\boldsymbol{x},\boldsymbol{x}')$ of Kriging; see Eqs. (28) and (29). This reflects a critical fact that the k-TIMSER enables fully accounting for the influences of adding $\mathcal{X}_k^+$ on any other point around them and, further, the global epistemic uncertainty of $\tilde{P}_f$ through $c_n(\cdot,\cdot)$. More importantly, the k-TIMSER can work directly with a batch of k new samples. Obviously, capitalizing on more useful information of Kriging and working directly with multiple new samples are two critical theoretical advantages of k-TIMSER over other existing learning functions.*

Remark 2. *The relatively complex expression of k-TIMSER exactly comes from the two aspects outlined in Remark 1. Despite relatively complex, the k-TIMSER only involves some simple matrix operations on Kriging predictions. Since Kriging is much cheaper to train and predict when compared with the computational model itself, the k-TIMSER is very cheap to evaluate. Further, the benefit of k-TIMSER in terms of improving the overall computational efficiency of reliability analysis will be manifested in Section 5.*

3.3. Multi-point enrichment process based on k-TIMSER

It is computationally-intensive to directly carry out the maximization of dimension $k\times d$ in Eq. (27). Hence, a cost-effective heuristic approach is devised here, of which the basic idea is to sequentially select a batch of new points one by one, instead of selecting them at once.

Specifically, when $k=1$, the 1-point TIMSER is expressed as

$$\mathrm{TIMSER}_{n+1}(\boldsymbol{x}_+)=n_{\rm rp}\sum_{i=1}^{n_{\rm rp}}(p^{(i)})^2\pi_n(\boldsymbol{x}^{(i)})\left[\sigma_n^2(\boldsymbol{x}^{(i)})-\sigma_{n+1}^2(\boldsymbol{x}^{(i)};\boldsymbol{x}_+)\right], \tag{30}$$

where $\sigma_{n+1}^2(\cdot;\boldsymbol{x}_+)$ acts as a reminder that $\sigma_{n+1}^2(\cdot)$ depends on $\boldsymbol{x}_+$ solely. Then, the 1-st best next point $\boldsymbol{x}^{(n+1)}$ is selected as

$$\boldsymbol{x}^{(n+1)}=\arg\max_{\boldsymbol{x}_+\in\mathcal{X}_{\rm cp}}\mathrm{TIMSER}_{n+1}(\boldsymbol{x}_+), \tag{31}$$

and $\mathcal{X}_1^*=\{\boldsymbol{x}^{(n+1)}\}$.

When $k\geq 2$, assume that the former $(k-1)$ best next points, i.e., $\mathcal{X}_{k-1}^*=\left\{\boldsymbol{x}^{(n+1)},\ldots,\boldsymbol{x}^{(n+k-1)}\right\}$, have been obtained and are taken as the fixed arguments of the k-point TIMSER, that is,

$$\mathrm{TIMSER}_{n+k}(\mathcal{X}_k^*,\boldsymbol{x}_+)=n_{\rm rp}\sum_{i=1}^{n_{\rm rp}}(p^{(i)})^2\pi_n(\boldsymbol{x}^{(i)})\left[\sigma_n^2(\boldsymbol{x}^{(i)})-\sigma_{n+k}^2(\boldsymbol{x}^{(i)};\mathcal{X}_k^*,\boldsymbol{x}_+)\right], \tag{32}$$

where $\sigma_{n+k}^2(\cdot;\mathcal{X}_k^*,\boldsymbol{x}_+)$ is a reminder that $\sigma_{n+k}^2(\cdot)$ is solely a function of $\boldsymbol{x}_+$ here. Then, the k-th best next point $\boldsymbol{x}^{(n+k)}$ can be selected as

$$\boldsymbol{x}^{(n+k)}=\arg\max_{\boldsymbol{x}_+\in\mathcal{X}_{\rm cp}}\mathrm{TIMSER}_{n+k}\left(\mathcal{X}_{k-1}^*,\boldsymbol{x}_+\right), \tag{33}$$

and $\mathcal{X}_k^*=\mathcal{X}_{k-1}^*\bigcup\{\boldsymbol{x}^{(n+k)}\}$.

Obviously, this heuristic approach reduces the original maximization problem of dimension $k\times d$ in Eq. (27) to k consecutive maximization problems of dimension d in Eqs. (31) and (33), which are much computationally cheaper to solve.

Another key issue in this heuristic approach is when to terminate the sequential selection process, i.e., how to determine the size, K, of batch of new samples added per iteration. Here, two different schemes are detailed as follows.

- *Prescribed scheme.* The batch size K is prescribed as a fixed value; then, sequentially increase k until K and solve the associated d-dimensional maximization problems in Eqs. (31) and (33), giving rise to $\mathcal{X}_K^*=\{\boldsymbol{x}^{(n+k)},k=1,\ldots,K\}$ readily.

- *Adaptive scheme.* The batch size K can be adaptively determined per iteration, thanks to the superior ability of k-TIMSER to measure the expected gain of adding each new point. Specifically, according to Eqs. (31) and (33), the individual expected gain of adding the k-th new point can be expressed as

$$G_{n+k} = \begin{cases} \mathrm{TIMSER}_{n+k}(\mathcal{X}_k^*) - \mathrm{TIMSER}_{n+k-1}(\mathcal{X}_{k-1}^*), & k > 1, \\ \mathrm{TIMSER}_{n+1}\left(\mathcal{X}_1^*\right), & k = 1, \end{cases} \tag{34}$$

  which generally reduces with the increasing of k. If the G_{n+k} itself or the $\frac{G_{n+k}}{G_{n+1}}$ is too minimal, adding $\boldsymbol{x}^{(n+k)}$ is not useful to further reduce the TIMSE in expectation. Hence, the sequential increase of k shall be stopped. In this way, the K is identified as

$$K = \inf\left\{k \in \mathbb{N} : \frac{G_{n+k}}{G_{n+1}} \leq \varepsilon_{\mathrm{G}} \bigcup G_{n+k} \leq 10^{-10}\right\} \tag{35}$$

  where the tolerance ε_{G} is set as 0.2 here.

Algorithm 1 provides the pseudo code of k-TIMSER-based multi-point enrichment process at a single iteration. The 'for-loop' in Lines 4-7 only involves some simple matrix operations and can be fastly run with the 'parfor' syntax in the MATLAB to support parallel computing.

Algorithm 1 k-TIMSER-based multi-point enrichment process at an iteration

Input: The Kriging $\widehat{\mathcal{M}}_n(\boldsymbol{x})$; the candidate pool $\mathcal{X}_{\mathrm{cp}} = \{\boldsymbol{x}_+^{(i)}, i = 1, \ldots, n_{\mathrm{cp}}\}$; the K in the prescribed scheme or the ε_{G} in the adaptive scheme.

1: *Initialize*: $k = 1$, and the initial batch of new samples is set as $\mathcal{X}_0^* = \{\}$.
2: The Kriging $\widehat{\mathcal{M}}_n(\boldsymbol{x})$ provides posterior mean $\mu_n(\boldsymbol{x})$, variance $\sigma_n^2(\boldsymbol{x})$ and covariance $c_n(\boldsymbol{x}, \boldsymbol{x}')$ at all samples in the $\mathcal{X}_{\mathrm{rp}}$. ▷ Eqs. (B.7), (B.8), (B.9)
3: **while** true **do**
4: **for** $i = 1 : n_{\mathrm{cp}}$ **do**
5: Compute the look-ahead variances $\left\{\sigma_{n+k}^2(\boldsymbol{x}^{(j)}; \mathcal{X}_{k-1}^*, \boldsymbol{x}_+^{(i)})\right\}_{j=1}^{n_{\mathrm{rp}}}$ at $\mathcal{X}_{\mathrm{rp}}$. ▷ Eq. (29)
6: Compute $\mathrm{TIMSER}_{n+k}(\mathcal{X}_{k-1}^*, \boldsymbol{x}_+^{(i)})$. ▷ Eqs. (30) and (32)
7: **end for**
8: Select the k-th best next point $\boldsymbol{x}^{(n+k)}$ from $\mathcal{X}_{\mathrm{cp}}$. ▷ Eqs. (31) or (33)
9: Determine the K value in the adaptive scheme according to ε_{G}. ▷ Eqs. (34) and (35)
10: **if** $k \geq K$ **then**
11: Break;
12: **else**
13: Update: $\mathcal{X}_k^* = \mathcal{X}_{k-1}^* \bigcup \{\boldsymbol{x}^{(n+k)}\}$, $\mathcal{X}_{\mathrm{cp}} = \mathcal{X}_{\mathrm{cp}} \setminus \{\boldsymbol{x}^{(n+k)}\}$, and $k = k + 1$.
14: **end if**
15: **end while**

Output: The $\mathcal{X}_K^* = \left\{\boldsymbol{x}^{(n+k)}, k = 1, \ldots, K\right\}$ obtained at this iteration.

4. Parallel active learning reliability analysis based on k-TIMSER

Apart from the learning function k-TIMSER, another two ingredients of parallel active learning reliability analysis are outlined here. Section 4.1 states the initial ED. Section 4.2 presents the convergence criterion. Then, Section 4.3 details the implementation of parallel active learning reliability analysis based on k-TIMSER. Finally, some discussions are given in Section 4.4.

4.1. Initial experimental design

To obtain a well-behaved Kriging in the initial stage, the initial ED is preferred to be as space-filling as possible. First, according to the so-called 'four-sigma' rule, the upper and lower bounds of the sampling domain $\mathbb{X}_{\mathrm{s}}$ at each dimension are set as

$$x_l^{\pm} = F_{X_l}^{-1}\big(\Phi(\pm 4)\big),\ l = 1, \ldots, d, \tag{36}$$

where $F_{X_l}(\cdot)$ is the CDF of the l-th component, X_l, in $\boldsymbol{X}$. Then, the $\mathbb{X}_{\mathrm{s}}$ is obtained by the following tensorization

$$\mathbb{X}_{\mathrm{s}} = \prod_{l=1}^{d} \left[x_l^-, x_l^+\right]. \tag{37}$$

Finally, the Latin centroidal Voronoi tessellation sampling method [54] is used to generate a set of n_0 uniform samples within $\mathbb{X}_{\mathrm{s}}$, denoted as $\mathcal{X}_{n_0} = \left\{\boldsymbol{x}^{(i)}, i = 1, \ldots, n_0\right\}$, with the size set as $n_0 = \max(d + 1, 10)$.

4.2. Convergence criterion

A hybrid convergence criterion that combines two different ones is considered here, so as to secure the robustness of termination of active learning reliability analysis process.

First, recall that the TIMSE H_n acts as an epistemic uncertainty measure of $\tilde{P}_{f,n}$, the metric $\frac{H_n}{\tilde{P}_{f,n}}$ can be used to specify a convergence criterion. When $\frac{H_n}{\tilde{P}_{f,n}}$ falls below a small tolerance, the accuracy of $\tilde{P}_{f,n}$ is considered satisfactory, that is,

$$\frac{H_n}{\tilde{P}_{f,n}} \leq \varepsilon_{\mathrm{H}}, \tag{38}$$

where the setting of ε_{H} shall account for the distinction between static and dynamic reliability problems in terms of the order of magnitude of Kriging variance. In static case, $\varepsilon_{\mathrm{H}} = \min\{10^{-3}, 10^{-2} \times \max_{i \leq n} \frac{H_n}{\tilde{P}_{f,n}}\}$; then, in dynamic case, $\varepsilon_{\mathrm{H}} = \min\{1, 5 \times 10^{-2} \times \max_{i \leq n} \frac{H_n}{\tilde{P}_{f,n}}\}$.

Second, denote $\tilde{\mathcal{Y}}_{\mathrm{rp},n} = \{\mu_n(\boldsymbol{x}^{(i)}), i = 1, \ldots, n_{\mathrm{rp}}\}$ as the Kriging means evaluated at the $\mathcal{X}_{\mathrm{rp}}$. Then, the upper and lower confidence bounds of $\tilde{\mathcal{Y}}_{\mathrm{rp},n}$ can be expressed as $\tilde{\mathcal{Y}}^{\pm}_{\mathrm{rp},n} = \{\mu_n(\boldsymbol{x}^{(i)}) \pm \alpha\sigma_n(\boldsymbol{x}^{(i)}), i = 1, \ldots, n_{\mathrm{rp}}\}$, where the coefficient α is usually taken as $1.96 = \Phi^{-1}(97.5\%)$. Further, performing the PDEM on $\tilde{\mathcal{Y}}_{\mathrm{rp},n}$, $\tilde{\mathcal{Y}}^{+}_{\mathrm{rp},n}$ and $\tilde{\mathcal{Y}}^{-}_{\mathrm{rp},n}$ will give rise to three failure probability estimates $\tilde{P}_{f,n}$, $\tilde{P}^{+}_{f,n}$, and $\tilde{P}^{-}_{f,n}$, respectively. In this way, the second convergence criterion is defined from the confidence bound of $\tilde{P}_{f,n}$ such that

$$\frac{\left|\tilde{P}^{+}_{f,n} - \tilde{P}^{-}_{f,n}\right|}{\tilde{P}_{f,n}} \leq \varepsilon_{\mathrm{B}}, \tag{39}$$

where the tolerance ε_{B} is taken as 1% here.

Finally, the hybrid convergence criterion stipulates that only when both Eq.(38) and Eq. (39) are satisfied simultaneously, the active learning workflow is converged, that is,

$$\frac{H_n}{\tilde{P}_{f,n}} \leq \varepsilon_{\mathrm{H}} \bigcap \frac{\left|\tilde{P}^{+}_{f,n} - \tilde{P}^{-}_{f,n}\right|}{\tilde{P}_{f,n}} \leq \varepsilon_{\mathrm{B}}. \tag{40}$$

4.3. Implementation

Fig. 4 presents the workflow of the proposed method, which is summarized as follows.

(1) **Partition of probability space**.
Generate a representative point set $\mathcal{X}_{\mathrm{rp}} = \{\boldsymbol{x}^{(i)}, i = 1, \ldots, n_{\mathrm{rp}}\}$, and compute their corresponding assigned probabilities $\mathcal{P}_{\mathrm{rp}} = \{p^{(i)}, i = 1, \ldots, n_{\mathrm{rp}}\}$; see Section 2.1. Then, the $\mathcal{X}_{\mathrm{rp}}$ is taken as the initial candidate pool $\mathcal{X}_{\mathrm{cp}}$ in the following sequential enrichment process.

(2) **Initial ED**.
Generate an initial input dataset $\mathcal{X}_{n_0} = \{\boldsymbol{x}^{(i)}, i = 1, \ldots, n_0\}$ of size $n_0 = \max(10, d + 1)$; then, evaluate the computational model $\mathcal{M}(\cdot)$ on $\mathcal{X}_{n_0}$ to obtain the corresponding responses $\mathcal{Y}_{n_0} = \{y^{(i)}, i = 1, \ldots, n_0\}$, forming the initial ED $\mathcal{D}_{n_0} = \{\mathcal{X}_{n_0}, \mathcal{Y}_{n_0}\}$; see Section 4.1. Let $n_{\mathrm{iter}} = 1$, and $n = n_0$.

(3) **Training of Kriging**.
Train a Kriging $\widehat{\mathcal{M}}_n(\boldsymbol{x})$ based on the current ED $\mathcal{D}_n$, where the kernel parameters $\boldsymbol{\theta}$ are optimized according to the maximum likelihood estimation method; see Eq. (B.5).

(4) **Failure probability estimation**.
The three failure probability estimates, $\tilde{P}_{f,n}$, $\tilde{P}^{+}_{f,n}$ and $\tilde{P}^{-}_{f,n}$, are computed by performing the PDEM on the corresponding Kriging predictions $\tilde{\mathcal{Y}}_{\mathrm{rp},n}$, $\tilde{\mathcal{Y}}^{+}_{\mathrm{rp},n}$ and $\tilde{\mathcal{Y}}^{-}_{\mathrm{rp},n}$, respectively. Meanwhile, in the computation of $\tilde{P}_{f,n}$, an estimate of h_{r}, denoted as $\tilde{h}_{\mathrm{r}}$, is obtained by substituting $\tilde{\mathcal{Y}}_{\mathrm{rp},n}$ into Algorithm A.1.

(5) **Convergence criterion**.
If the hybrid convergence criterion in Eq. (40) is satisfied, skip to Step 8; otherwise, continue to Step 6.

(6) **Learning function**.
Substitute $\tilde{h}_{\mathrm{r}}$ (Step 4) into the expression of k-TIMSER in Eqs. (30) and (32); then, select a batch of K best next points, $\mathcal{X}^{*}_{K} = \{\boldsymbol{x}^{(n+k)}, k = 1, \ldots, K\}$, from the candidate pool $\mathcal{X}_{\mathrm{cp}}$, where the K value can be identified either by the traditional prescribed scheme by or newly-developed adaptive scheme, as detailed in Algorithm 1.

(7) **Enrichment**.
Evaluate the actual computational model $\mathcal{M}(\cdot)$ on $\mathcal{X}^{*}_{K}$ in parallel, resulting in the corresponding new responses $\mathcal{Y}^{*}_{K} = \{y^{(n+k)}, i = 1, \ldots, K\}$. Then, conduct the following updates: $\mathcal{D}_{n+K} = \mathcal{D}_n \bigcup \{\mathcal{X}^{*}_{K}, \mathcal{Y}^{*}_{K}\}$, $\mathcal{X}_{\mathrm{cp}} = \mathcal{X}_{\mathrm{cp}} \setminus \mathcal{X}^{*}_{K}$, and $n \leftarrow n + K$.

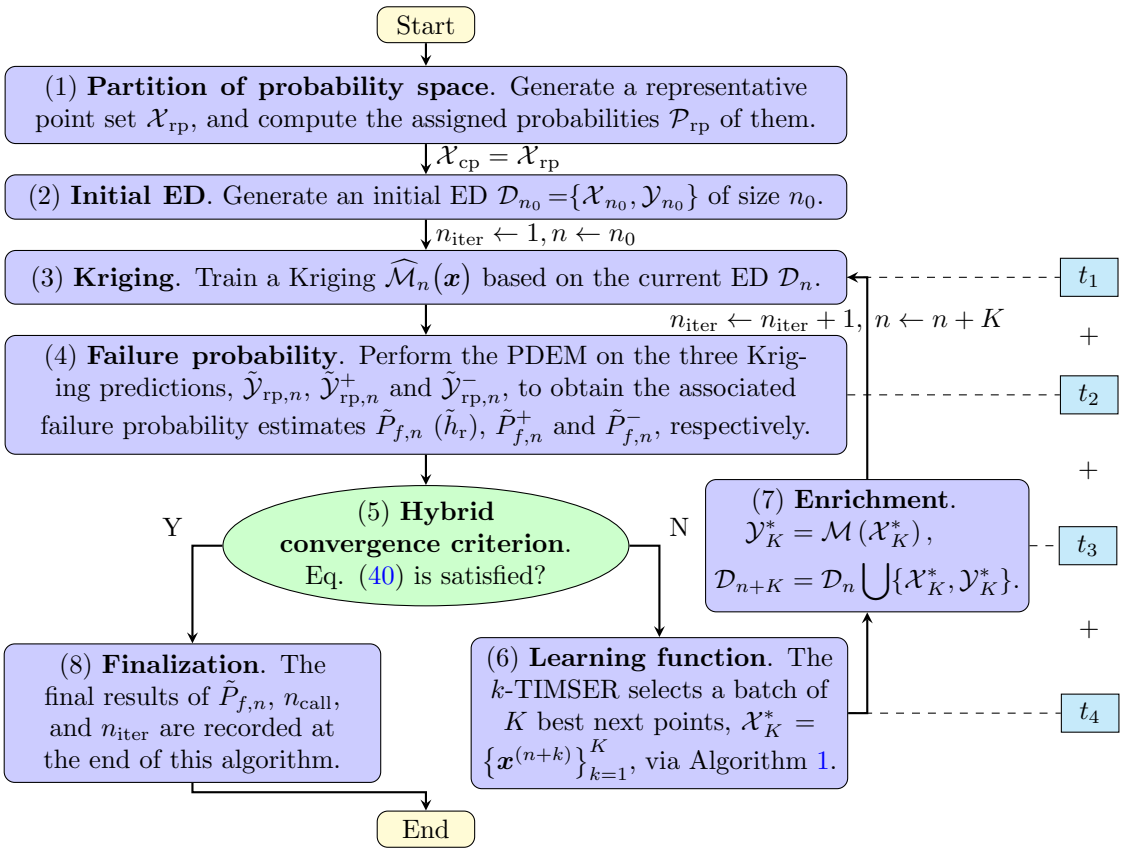


Figure 4: Flowchart of k-TIMSER-based parallel active learning reliability method

(8) **End**.
The results of the failure probability estimate $\tilde{P}_{f,n}$, the total number $n_{\rm call}$ of calls to computational model, and the total number $n_{\rm iter}$ of iterations are recorded at the end of this algorithm.

Remark 3. *The total computational time of the proposed method (Fig. 4), as well as any other parallel active learning reliability method, can be roughly calculated as*

$$t_{\rm c} = n_{\rm iter} \times t_{\rm iter}, \tag{41}$$

where $n_{\rm iter}$ is the total number of iterations, and $t_{\rm iter}$ is the running time of a single iteration. As plotted as cyan boxes in Fig. 4, there holds $t_{\rm iter} = t_1 + t_2 + t_3 + t_4$, with t_1, t_2, t_3, t_4 being the training time of Kriging, the time of reliability analysis, the evaluation time of computational model $\mathcal{M}(\cdot)$, and the time of multi-point enrichment process at an iteration, respectively. The t_1 and t_2 are relatively fixed, the t_3 depends on the $\mathcal{M}(\cdot)$ under consideration, and the t_4 is related to the learning function being used. If $\mathcal{M}(\cdot)$ is more time-intensive than the learning function, the t_3 will dominate $t_{\rm iter}$, i.e., $t_{\rm iter} \approx t_3$. In this way, the advantage of an active learning reliability method in terms of reducing $n_{\rm iter}$ will be readily converted into the advantage of reducing $t_{\rm c}$, as per Eq. (41). On the contrary, if $\mathcal{M}(\cdot)$ itself is very cheap to evaluate and the time t_4 of learning function dominates the $t_{\rm iter}$, i.e., $t_{\rm iter} \approx t_4$, it is probable that although the $n_{\rm iter}$ of an active learning reliability method is smaller, its $t_{\rm c}$ is inversely greater.

Remark 4. *It is admitted that the proposed k-TIMSER-based multi-point enrichment approach (Algorithm 1) needs slightly greater time, t_4, than the traditional practice of combining a pointwise learning function, e.g., U and EFF, with the K-means clustering strategy. Then, as per Remark 3, the advantage of the proposed approach in terms of the total computational time $t_{\rm c}$ will depend on the total number $n_{\rm iter}$ of iterations and the evaluation time t_3 of $\mathcal{M}(\cdot)$ under consideration. This will be further discussed in Section 5.*

4.4. Discussions

Recall that the main purpose of k-TIMSER is to select a sequence of new samples, so as to reduce dramatically the epistemic uncertainty measure of failure probability estimation, i.e., TIMSE. Then, let's clarify the distinction between the following two important quantities at each iteration of Fig. 4: (i) The expected reduction of TIMSE due to the batch of K best next points $\mathcal{X}^*_K$, which is quantified by the learning function

k-TIMSER in Step 6; (ii) The actual reduction of TIMSE brought by enriching the existing ED $\mathcal{D}_n$ with the pair of $\mathcal{X}_K^*$ and their actual computational model responses $\mathcal{Y}_K^*$ in Step 7.

According to the expressions of the learning function k-TIMSER in Eq. (28) and the TIMSE H_n in Eq. (18), these two quantities are computed as

$$G_K = \text{TIMSER}_{n+K}(\mathcal{X}_K^*) = H_n - n_{\text{rp}} \sum_{i=1}^{n_{\text{rp}}} (p^{(i)})^2 \pi_n(\boldsymbol{x}^{(i)}) \sigma_{n+k}^2(\boldsymbol{x}^{(i)})|_{k=K}, \tag{42}$$

$$\Delta H_K = H_n - H_{n+K} = H_n - n_{\text{rp}} \sum_{i=1}^{n_{\text{rp}}} (p^{(i)})^2 \pi_{n+K}(\boldsymbol{x}^{(i)}) \sigma_{n+K}^2(\boldsymbol{x}^{(i)}), \tag{43}$$

respectively, where $\sigma_{n+k}^2(\cdot)|_{k=K}$ is a reminder that it is given by the Kriging update formulas (Eq. (E.2)), distinguishing it from the actual variance $\sigma_{n+K}^2(\cdot)$ obtained by retraining the Kriging.

Then, the absolute difference between them is expressed as

$$\begin{aligned} |\Delta H_K - G_K| &= \left| n_{\text{rp}} \sum_{i=1}^{n_{\text{rp}}} (p^{(i)})^2 \pi_{n+K}(\boldsymbol{x}^{(i)}) \sigma_{n+K}^2(\boldsymbol{x}^{(i)}) - n_{\text{rp}} \sum_{i=1}^{n_{\text{rp}}} (p^{(i)})^2 \pi_n(\boldsymbol{x}^{(i)}) \sigma_{n+k}^2(\boldsymbol{x}^{(i)})|_{k=K} \right|, \\ &= n_{\text{rp}} \left| \sum_{i=1}^{n_{\text{rp}}} (p^{(i)})^2 \left[\pi_{n+K}(\boldsymbol{x}^{(i)}) \sigma_{n+K}^2(\boldsymbol{x}^{(i)}) - \pi_n(\boldsymbol{x}^{(i)}) \sigma_{n+k}^2(\boldsymbol{x}^{(i)})|_{k=K} \right] \right|, \\ &\leq n_{\text{rp}} \sum_{i=1}^{n_{\text{rp}}} (p^{(i)})^2 \left| \pi_{n+K}(\boldsymbol{x}^{(i)}) \sigma_{n+K}^2(\boldsymbol{x}^{(i)}) - \pi_n(\boldsymbol{x}^{(i)}) \sigma_{n+k}^2(\boldsymbol{x}^{(i)})|_{k=K} \right|, \\ &= n_{\text{rp}} \sum_{i=1}^{n_{\text{rp}}} (p^{(i)})^2 \xi_{\text{ad}}(\boldsymbol{x}^{(i)}), \end{aligned} \tag{44}$$

where $\xi_{\text{ad}}(\boldsymbol{x}^{(i)}) = \left| \pi_{n+K}(\boldsymbol{x}^{(i)}) \sigma_{n+K}^2(\boldsymbol{x}^{(i)}) - \pi_n(\boldsymbol{x}^{(i)}) \sigma_{n+k}^2(\boldsymbol{x}^{(i)})|_{k=K} \right|$.

In essence, the difference between ΔH_K and G_K is attributed to the distinction between the Kriging update formulas (Appendix E) and the actual retraining of Kriging (Appendix B) as follows.

- When a pair of K new samples and their actual computational model responses, i.e., $\{\mathcal{X}_K^*, \mathcal{Y}_K^*\}$, are added into the ED $\mathcal{D}_n$, a new Kriging $\widehat{\mathcal{M}}_{n+K}(\boldsymbol{x})$ can be trained based on the augmented ED $\mathcal{D}_{n+K} = \mathcal{D}_n \bigcup \{\mathcal{X}_K^*, \mathcal{Y}_K^*\}$. In essence, the retraining of Kriging comprises two consecutive steps. First, reoptimize the kernel parameters $\boldsymbol{\theta}$ according to Eq. (B.5). Second, recompute the remaining parameters, i.e., $\boldsymbol{\beta}$ and σ^2, based on $\boldsymbol{\theta}$; see Eqs. (B.3) and (B.4). In this way, the updated mean $\mu_{n+K}(\boldsymbol{x})$, variance $\sigma_{n+K}(\boldsymbol{x})$, and covariance $c_{n+K}(\boldsymbol{x}, \boldsymbol{x}')$ of the new Kriging $\widehat{\mathcal{M}}_{n+K}(\cdot)$ are provided in Eqs. (B.7), (B.8), and (B.9), respectively. Hence, the ΔH_K is interpreted as the actual gain incurred by $\{\mathcal{X}_K^*, \mathcal{Y}_K^*\}$.
- As shown in Eqs. (E.1), (E.2), and (E.3), the look-ahead posteriors of Kriging are directly obtained from the current posteriors, i.e., $\mu_n(\boldsymbol{x})$, $\sigma_n(\boldsymbol{x})$, and $c_n(\boldsymbol{x}, \boldsymbol{x}')$, of Kriging $\widehat{\mathcal{M}}_n(\cdot)$, without optimizing the kernel parameters $\boldsymbol{\theta}$. In essence, the Kriging update formulas only recompute both β and σ^2 based on the current values of $\boldsymbol{\theta}$ [32]. Hence, the G_K can be interpreted as the average value of possible reduction of TIMSE brought by $\mathcal{X}_K^*$, conditional on the Kriging assumption of their model responses.

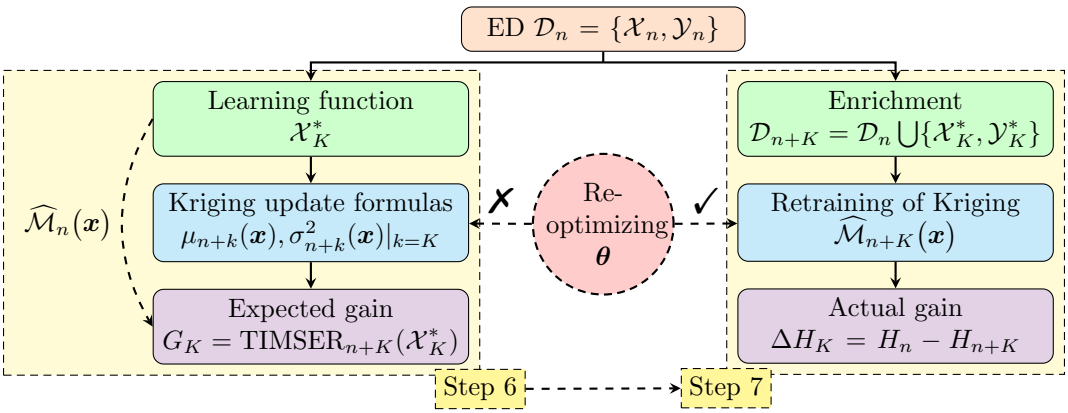


Figure 5: Difference of the actual gain ΔH_K and the expected gain G_K

In summary, the distinction between the Kriging update formulas and the actual retraining of Kriging lies in whether or not the kernel parameters $\boldsymbol{\theta}$ are re-optimized, as shown in Fig. 5. Hence, if the kernel parameters

$\boldsymbol{\theta}$ of Kriging only vary slightly in two consecutive iterations, the Kriging update formulas are nearly equivalent to the actual retraining of Kriging, that is, $\sigma^2_{n+k}(\boldsymbol{x})|_{k=K} \approx \sigma^2_{n+K}(\boldsymbol{x})$. Then, there holds

$$\xi_{\mathrm{ad}}(\boldsymbol{x}) \approx \left|\pi_{n+K}(\boldsymbol{x})\sigma^2_{n+K}(\boldsymbol{x}) - \pi_n(\boldsymbol{x})\sigma^2_{n+K}(\boldsymbol{x})\right| = \sigma^2_{n+K}(\boldsymbol{x})\left|\pi_{n+K}(\boldsymbol{x}) - \pi_n(\boldsymbol{x})\right|. \tag{45}$$

Further, substitute Eq. (45) into Eq. (44), yielding

$$|\Delta H_K - G_K| \le n_{\mathrm{rp}} \sum_{i=1}^{n_{\mathrm{rp}}} (p^{(i)})^2 \sigma^2_{n+K}(\boldsymbol{x}^{(i)}) \left|\pi_n(\boldsymbol{x}^{(i)}) - \pi_{n+K}(\boldsymbol{x}^{(i)})\right|. \tag{46}$$

Besides, recall that the actual gain ΔH_K in Eq. (43) can be further expanded as

$$\begin{aligned} \Delta H_K &= n_{\mathrm{rp}} \sum_{i=1}^{n_{\mathrm{rp}}} (p^{(i)})^2 \left[\pi_n(\boldsymbol{x}^{(i)})\sigma^2_n(\boldsymbol{x}^{(i)}) - \pi_{n+K}(\boldsymbol{x}^{(i)})\sigma^2_{n+K}(\boldsymbol{x}^{(i)})\right], \\ &\ge n_{\mathrm{rp}} \sum_{i=1}^{n_{\mathrm{rp}}} (p^{(i)})^2 \sigma^2_{n+K}(\boldsymbol{x}^{(i)}) \left|\pi_n(\boldsymbol{x}^{(i)}) - \pi_{n+K}(\boldsymbol{x}^{(i)})\right|, \end{aligned} \tag{47}$$

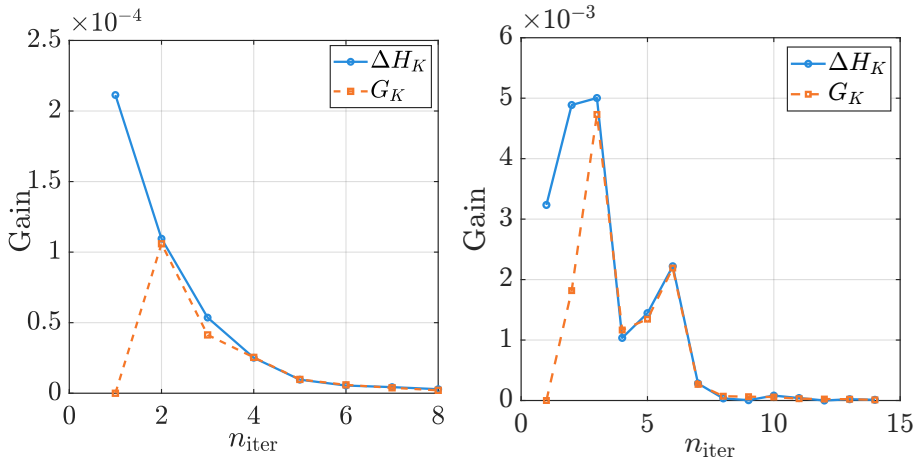
which is based on the fact that $\sigma^2_{n+K}(\boldsymbol{x}) \le \sigma^2_n(\boldsymbol{x}), \forall \boldsymbol{x} \in \mathbb{X}$.

Finally, substituting Eq. (47) into Eq. (46), there holds

$$|\Delta H_K - G_K| \le \Delta H_K \to 0, \tag{48}$$

which indicates that when the kernel parameters $\boldsymbol{\theta}$ of Kriging do not vary significantly with adding new samples, the $|\Delta H_K - G_K|$ converges to 0. This is a common scenario in the latter iterations of active learning reliability analysis.

For illustration, Fig. 6 presents the comparison between ΔH_K and G_K (prescribed scheme, $K = 5$) in both the truss example (Section 5.2) and the reinforced concrete frame example (Section 5.3). Due to only a few training samples in the several initial iterations, the kernel parameters $\boldsymbol{\theta}$ of Kriging vary significantly with adding new samples, giving rise to notable difference between ΔH_K and G_K. By comparison, very minor fluctuation of $\boldsymbol{\theta}$ exists in the latter iterations, and the H_{n+K} converges gradually to 0. Then, a good agreement between ΔH_K and G_K is witnessed.



(a) planar truss example (b) reinforced concrete frame example

Figure 6: Comparison between ΔH_K and G_K in two examples

It is worth noting that since Step 6 is prior to Step 7 during the workflow of Fig. 4, it is necessary to formulate the G_k in Step 6, serving as an estimate of ΔH_K, during the active learning workflow. Fortunately, apart from several initial iterations, this practice performs very well during the latter iterations of active learning process. Hence, this is very reassuring that the batches of new samples selected by k-TIMSER are optimal to reducing the TIMSE at most at the latter iterations.

5. Numerical examples

The proposed method is testified on three examples with varying complexity. The MCS is conducted to provide the failure probability estimate $\widehat{P}_f^{\mathrm{MCS}}$ for reference. In the k-TIMSER, both the prescribed scheme and

the adaptive scheme are considered to determine the size, K, of batch of new points added per iteration. The K values in the prescribed scheme are set as 1, 5, 10, 15 and 20, respectively, while the K value in the adaptive scheme is determined according to Eq. (35).

Several existing parallel active learning reliability methods are conducted for comparison, including the parallel AK-MCS [39], the ALR module in Uqlab [14], and the PABQ [55]. Specifically, in the parallel AK-MCS [39], the K-means clustering method is used with the classical AK-MCS to favor parallel computing. The default setups of the ALR module in Uqlab [14] are considered here, consisting of the learning function U, the polynomial-chaos Kriging model, the SuS-based reliability algorithm, the convergence criterion based on the bound of reliability index β, and the K-means clustering strategy. The PABQ comprises the following ingredients [55]: the Kriging model, the Bayesian inference of failure probability estimation, the learning function called upper-bound variance contribution, and the K-means clustering strategy. Note that the results from other existing reliability methods available in the literature will also be reported in the tables.

All reliability methods are run on an Intel i9-14900KF CPU processor at 3.2 GHz with 64 GB RAM and 20 CPU cores. Three performance metrics are recorded for each reliability method, namely the total number n_{iter} of iterations, the total number n_{call} of calls to computational model, and the final failure probability estimate $\tilde{P}_f$. Due to computational time considerations, the maximum value of n_{call} is set as 200 in those active learning reliability methods. Then, the relative error of $\tilde{P}_f$ with respect to $\widehat{P}_f^{\text{MCS}}$ is computed as

$$\delta_{\tilde{P}_f}^{(1)} = \left|\tilde{P}_f - \widehat{P}_f^{\text{MCS}}\right| / \widehat{P}_f^{\text{MCS}} \times 100\%, \tag{49}$$

which represents the overall error of a reliability method. Besides, the pure PDEM [10] is conducted to provide the second reference failure probability $\widehat{P}_f^{\text{PDEM}}$ for the proposed k-TIMSER. Then, another two error metrics are separately computed as

$$\delta_{\tilde{P}_f}^{(2)} = \left|\tilde{P}_f - \widehat{P}_f^{\text{PDEM}}\right| / \widehat{P}_f^{\text{PDEM}} \times 100\%, \tag{50}$$

$$\delta_{\tilde{P}_f}^{(3)} = \left|\widehat{P}_f^{\text{PDEM}} - \widehat{P}_f^{\text{MCS}}\right| / \widehat{P}_f^{\text{MCS}} \times 100\%, \tag{51}$$

which correspond to Kriging-caused error and PDEM-induced error in the k-TIMSER, respectively.

Those active learning reliability methods are repeated 10 times to account for the randomness arising from sampling and/or initial ED. Then, the average values of the four performance metrics, i.e., $\mathbb{E}\left[n_{\text{iter}}\right]$, $\mathbb{E}\left[n_{\text{call}}\right]$, $\mathbb{E}\left[\tilde{P}_f\right]$, and $\mathbb{E}\left[\delta_{\tilde{P}_f}\right]$, are listed in the tables. Additionally, the sample coefficient of variation (COV) of $\tilde{P}_f$, denoted as $\text{COV}[\tilde{P}_f]$, is calculated to reflect the variation of $\tilde{P}_f$ in those active learning reliability methods. The average results of the total computational time t_{c} are reported in the last two numerical examples for comparison.

Remark 5. *As per Section 2.3, two categories of error, i.e., PDEM-caused error $\delta_{\tilde{P}_f}^{(3)}$ in Eq. (51) and Kriging-induced error $\delta_{\tilde{P}_f}^{(2)}$ in Eq. (50), need to be reduced to ensure the accuracy of $\tilde{P}_f$ in k-TIMSER. (i) The 'over-kill' setting of PDEM is used to reduce $\delta_{\tilde{P}_f}^{(3)}$, where the size n_{rp} of $\mathcal{X}_{\text{rp}}$ is set as $\mathcal{O}(10^3)$, rather than the usual $\mathcal{O}(10^2)$. (ii) The k-TIMSER is defined to dramatically reduce the TIMSE, acting as the upper bound of $\delta_{\tilde{P}_f}^{(2)}$. By doing so, the overall error $\delta_{\tilde{P}_f}^{(1)}$ of k-TIMSER in Eq. (49) is expected to be reduced to a favorable level.*

5.1. A four-branch function

The four-branch function is a common benchmark in structural reliability analysis [24, 13, 38, 55], which represents a series system consisting of four distinct failure domains. The P_f is defined as

$$P_f = \mathbb{P}\left(Y \geq \frac{\sqrt{b}}{2}\right) = \int_{\frac{\sqrt{b}}{2}}^{+\infty} f_Y(y)\mathrm{d}y, \tag{52}$$

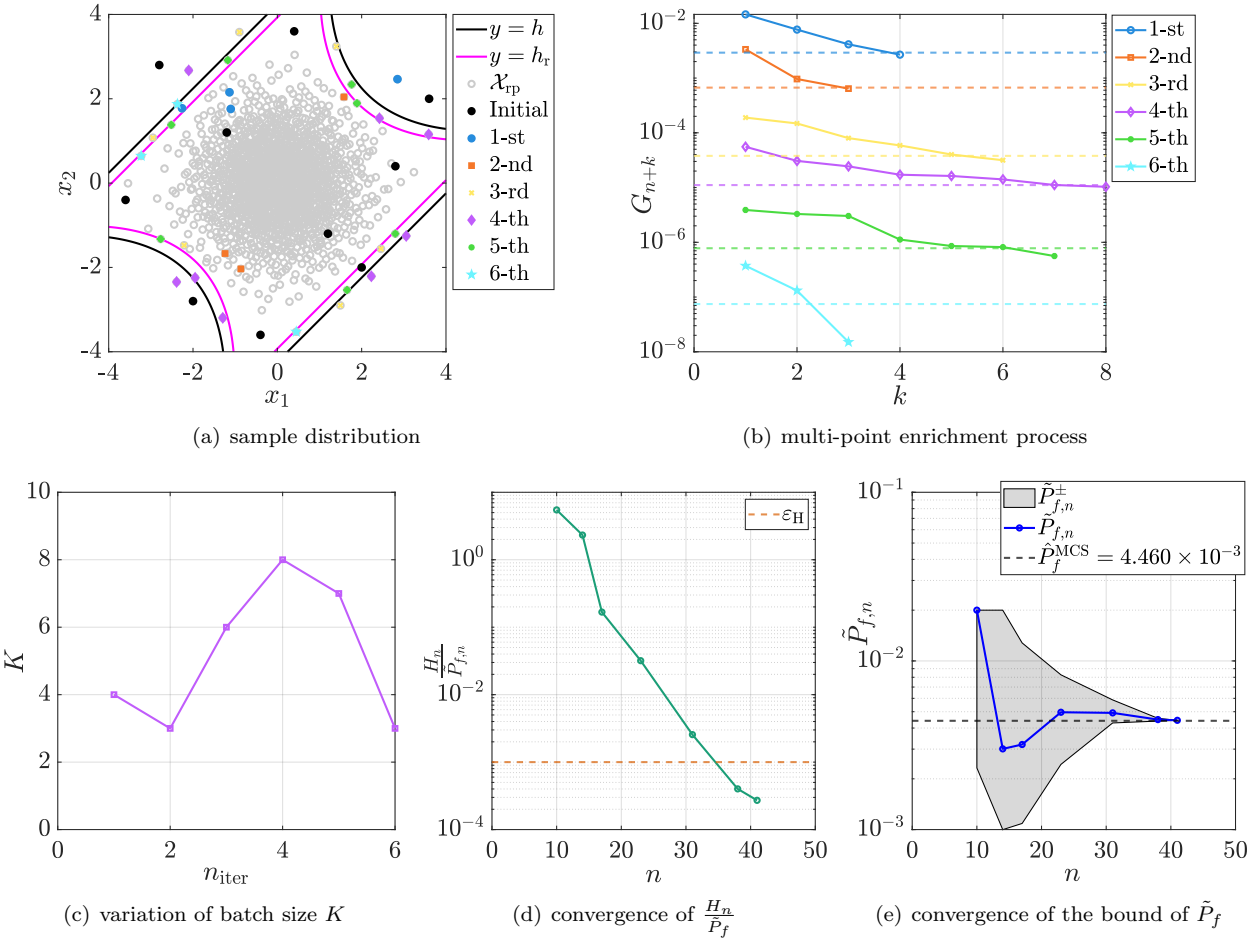
where

$$Y = \mathcal{M}(\boldsymbol{X}) = \max \begin{Bmatrix} -\frac{(X_1 - X_2)^2}{10} + \frac{X_1 + X_2}{\sqrt{2}} + \frac{b}{\sqrt{2}} - a \\ -\frac{(X_1 - X_2)^2}{10} - \frac{X_1 + X_2}{\sqrt{2}} + \frac{b}{\sqrt{2}} - a \\ X_1 - X_2 \\ X_2 - X_1 \end{Bmatrix}, \tag{53}$$

where $\boldsymbol{X} = \{X_1, X_2\}$ is a vector of two independent, standard Gaussian random variables; a and b are two constants. Two cases with distinct values of a and b are often considered in the existing literature: Case 1: $a = 3$ and $b = 6$; Case 2: $a = 3$ and $b = 7$.

5.1.1. Case 1: $a = 3$ and $b = 6$

Fig. 7 illustrates one run of k-TIMSER (adaptive scheme) in the four-branch function (Case 1). As marked as black circles in Fig. 7(a), the samples in the initial ED are scattered in the entire probability space, facilitating Kriging to discover the four component failure domains as early as possible. Then, after the first two iterations, the batches of new samples added by k-TIMSER in the latter iterations are mostly located around the actual ROI. Fig. 7(b) details the multi-point enrichment process in the adaptive scheme. Basically, the G_{k+k} value brought by adding the k-th new point in the latter iterations is smaller than that in the former ones, implying the gradual decrease of information gain brought by new samples. Fig. 7(c) shows that the number K of new samples added per iteration increases in the first several iterations and reduces subsequently. This behavior avoids the waste of adding some 'useless' new samples at the initial and final iterations. Figs. 7(d) and 7(e) presents the hybrid convergence criterion in Eq. (40). The $\frac{H_n}{\tilde{P}_{f,n}}$ reduces dramatically with the increasing of n_{iter}, justifying the good ability of k-TIMSER to reduce the epistemic uncertainty of $\tilde{P}_f$. Consequently, $\tilde{P}_{f,n}$ gradually converges to the referenced value $\widehat{P}_f^{\text{MCS}}$.



(a) sample distribution

(b) multi-point enrichment process

(c) variation of batch size K

(d) convergence of $\frac{H_n}{\tilde{P}_f}$

(e) convergence of the bound of $\tilde{P}_f$

Figure 7: The k-TIMSER (adaptive scheme) in the four-branch function (Case 1)

Table 1 gives comparative results of different reliability methods in the four-branch function (Case 1). In comparison with the classical AK-MCS, those parallel active learning reliability methods reduce n_{iter} remarkably. In the prescribed scheme of k-TIMSER, the n_{iter} value reduces slowly but the n_{call} value increases significantly when the prescribe batch size $K \geq 10$. This implies that the blind increase of K will cause k-TIMSER to add too many 'useless' new samples. By comparison, in the adaptive scheme of k-TIMSER, the average value of n_{call} is 41.1, which is much smaller than that of the prescribed scheme ($K = 20$). Hence, the adaptive scheme achieves a good balance between n_{iter} and n_{call}.

The relative error $\delta^{(3)}_{\tilde{P}_f}$ of the pure PDEM is only 0.193%, implying that the 'over-kill' setting of PDEM yields favorable accuracy of failure probability estimate. Then, both the overall error $\delta^{(1)}_{\tilde{P}_f}$ and the Kriging-induced error $\delta^{(2)}_{\tilde{P}_f}$ (in parenthesis) of k-TIMSER are listed in Table 1. On the whole, the $\delta^{(2)}_{\tilde{P}_f}$ is always smaller than 0.03%, indicating that the Kriging-induced error has been sufficiently reduced by the multi-point enrichment process. Then, the $\delta^{(1)}_{\tilde{P}_f}$ is dominated by the remaining PDEM-caused error $\delta^{(3)}_{\tilde{P}_f}$, say $< 0.2\%$. Obviously, such a

minimal value of $\delta^{(1)}_{\tilde{P}_f}$ demonstrates the good accuracy of k-TIMSER.

Table 1: Reliability results in the four-branch function (Case 1)

Method		$\mathbb{E}[n_{\text{iter}}]$	$\mathbb{E}[n_{\text{call}}]$	$\mathbb{E}[\tilde{P}_f](\times 10^{-3})$	$\text{COV}[\tilde{P}_f](\%)$	$\mathbb{E}[\delta_{\tilde{P}_f}](\%)$	References
MCS	-	-	10^6	4.460	-	-	[39]
AK-MCS	-	115	126	4.416	-	-	[24]
APCK-MCS (K-means)	$K=6$	15.4	98.4	4.458	1.500	-	[39]
RBIK (K-medoids)	$K=6$	17.7	110	4.429	0.070	-	[42]
	$K=8$	13.7	111.9	4.428	0.070	-	[42]
	$K=10$	12.1	121.4	4.429	0.070	-	[42]
AK-MCS (K-means)	$K=6$	21	132	4.420	1.480	-	[44]
AK-MCS (Combination)[a]	$K=6$	19	120	4.420	2.180	-	[44]
AK-KB^n (K-means)	$K=3$	22.5	74.6	4.419	-	-	[41]
	$K=6$	11.5	73.1	4.411	-	-	[41]
P-AK-MCS[b]	$K=4$	15.6	70.4	4.490	-	-	[50]
	$K=8$	8.8	74.4	4.560	-	-	[50]
AK-MCS-EU[c]	Adaptive	12	53	4.430	-	-	[46]
PABQ (K-means)	$K=6$	6.6	43.6	4.440	2.530	-	[55]
	$K=10$	5.2	52	4.400	2.220	-	[55]
	$K=15$	4.6	64.8	4.440	1.350	-	[55]
	$K=20$	4.1	71	4.440	1.290	-	[55]
ALR in UQLAB	$K=5$	20.3	106.5	4.571	1.424	3.509	-
	$K=10$	10.2	102	4.590	2.216	3.930	-
	$K=15$	8	115	4.541	2.493	2.985	-
	$K=20$	6.9	128	4.558	1.782	3.212	-
PDEM	-	-	2×10^3	4.451	-	0.193	-
k-TIMSER	$K=1$	28.7	37.7	4.450	0.038	0.199 (3.277×10^{-2})	-
	$K=5$	7	40	4.450	0.024	0.197 (1.550×10^{-2})	-
	$K=10$	4.8	48	4.450	0.016	0.196 (9.963×10^{-3})	-
	$K=15$	4.1	56.5	4.451	0.020	0.195 (9.488×10^{-3})	-
	$K=20$	3.9	68	4.451	0.001	0.194 (1.132×10^{-12})	-
	Adaptive	6.5	41.1	4.450	0.033	0.197 (2.196×10^{-2})	-

[a] Combination of K-means and K-medoids clustering techniques.
[b] This is based on pseudo learning function.
[c] This is based on Kriging believer strategy.

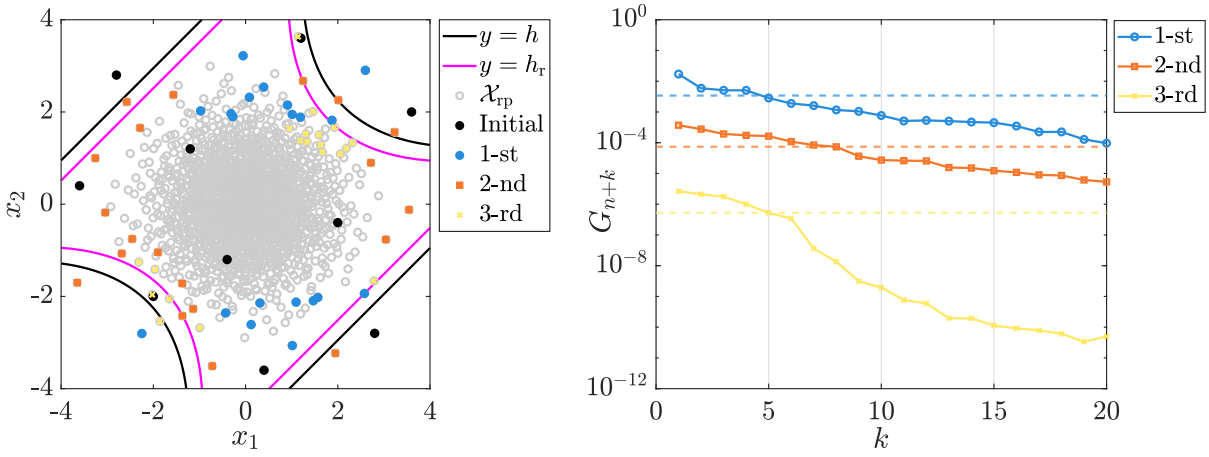
5.1.2. Case 2: $a = 3$ and $b = 7$

To shed light on the distinction between the adaptive scheme and the prescribed scheme in k-TIMSER, Fig. 8 presents the performance of k-TIMSER (prescribed scheme, $K = 20$) in the four-branch function (Case 2), and two unfavorable observations are outlined here. (i) As marked as blue circles in Fig. 8(a), the batch of 20 new samples added in the first iteration are overlapped and are far away from the actual ROI, due to the inferior performance of Kriging at the initial stage. (ii) As plotted as yellow circles in Fig. 8(a), most of the 20 new samples added in the last iteration are outside the ROI. Hence, there is a waste of new samples in these two iterations. To mitigate the two hurdles, it is feasible to timely stop the multi-point enrichment process according to the average gain induced by adding each new sample; see Fig. 8(b). This is exactly the momentum of devising the adaptive scheme. Obviously, in comparison with the traditional prescribe scheme, the adaptive scheme could avoid some wastes of computational model evaluations, as shown in Fig. 7.

Table 2 lists the results of different reliability methods for the four-branch function (Case 2). The existing ISKRA, the PA-BFPL, the ALR in Uqlab, and the PABQ provide adequate accuracy of failure probability estimates. Then, the proposed k-TIMSER needs much smaller values of n_{iter} and n_{call}. Based on the 'over-kill' setting, the pure PDEM comes with a much smaller value of $\delta^{(3)}_{\tilde{P}_f}$, say 0.134%. Then, thanks to the multi-point enrichment process, the Kriging-induced error $\delta^{(2)}_{\tilde{P}_f}$ (in parenthesis) of k-TIMSER is very minimal, say $< 0.02\%$. In this way, the k-TIMSER comes with a comparable value of $\delta_{\tilde{P}_f}$ to the pure PDEM, but only needing approximately 2.5% of n_{call}.

5.2. A two-dimensional truss

The second example considers a two-dimensional truss under vertical concentrated loads (Fig. 9), which is also a benchmark reliability problem in existing literature [14, 39]. This truss comprises a total of 23 bars and



(a) sample distribution (b) multi-point enrichment process

Figure 8: The k-TIMSER (prescribed scheme, $K = 20$) in the four-branch function (Case 2)

Table 2: Reliability results in the four-branch function (Case 2)

Method		$\mathbb{E}[n_{\text{iter}}]$	$\mathbb{E}[n_{\text{call}}]$	$\mathbb{E}[\tilde{P}_f](\times 10^{-3})$	$\text{COV}[\tilde{P}_f](\%)$	$\mathbb{E}[\delta_{\tilde{P}_f}](\%)$	References
MCS	-	-	10^6	2.233	-	-	[24]
AK-MCS	-	85	96	2.233	-	-	[24]
ISKRA(KB)	$K = 12$	7.68	92.16	2.230	1.500	-	[38]
ISKRA(K-means)	$K = 12$	9.62	115.44	2.215	1.500	-	[38]
	$K = 5$	7.5	42.5	2.130	3.070	-	[56]
PA-BFPL	$K = 12$	5.3	61.6	2.240	1.590	-	[56]
	$K = 15$	4.4	61	2.220	1.080	-	[56]
	$K = 5$	10.1	55.5	1.910	7.760	14.715	-
ALR in	$K = 10$	6.9	69	1.977	11.586	13.471	-
UQLAB	$K = 15$	6.9	98.5	2.150	10.637	8.026	-
	$K = 20$	6.4	118	2.178	10.257	7.585	-
	$K = 5$	7.7	43.5	2.182	1.255	2.274	-
PABQ	$K = 10$	5.1	51	2.213	1.690	1.367	-
	$K = 15$	4.5	62.5	2.210	1.144	1.280	-
	$K = 20$	4.3	76	2.223	1.359	1.001	-
PDEM	-	-	2×10^3	2.230	-	0.134	-
	$K = 1$	21.1	30.1	2.230	0.031	0.138 (2.038×10^{-2})	-
	$K = 5$	6.4	37	2.230	0.031	0.137 (1.905×10^{-2})	-
k-TIMSER	$K = 10$	4	40	2.230	0.002	0.135 (1.423×10^{-3})	-
	$K = 15$	3.7	50.5	2.230	0.008	0.135 (2.821×10^{-3})	-
	$K = 20$	3.2	54	2.230	0.003	0.136 (1.649×10^{-3})	-
	Adaptive	5.6	35.3	2.230	0.027	0.137 (1.704×10^{-2})	-

13 nodes. The random vector $\boldsymbol{X}$ consists of 10 random variables, that is, $\boldsymbol{X} = \{E_1, E_2, A_1, A_2, P_1, \ldots, P_6\}$, where E_1 and A_1 are Young's modulus and cross section of horizontal bars, respectively; E_2 and A_2 are Young's modulus and cross section of diagonal bars, respectively; $P_1, \ldots, P_6$ are vertical loads applied on the upper nodes from left to right. The statistical information of those random variables is given in Table 3.

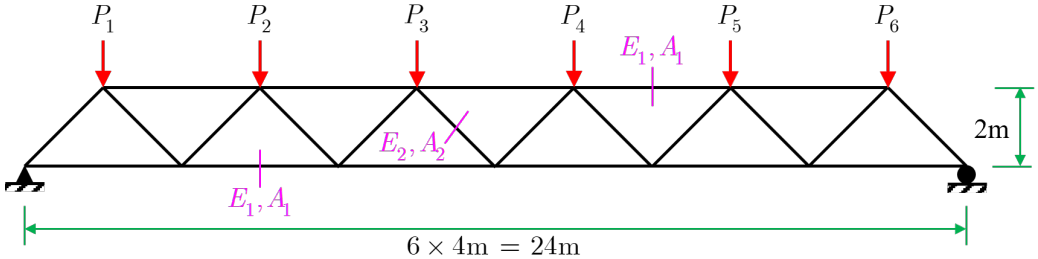


Figure 9: A two-dimensional truss

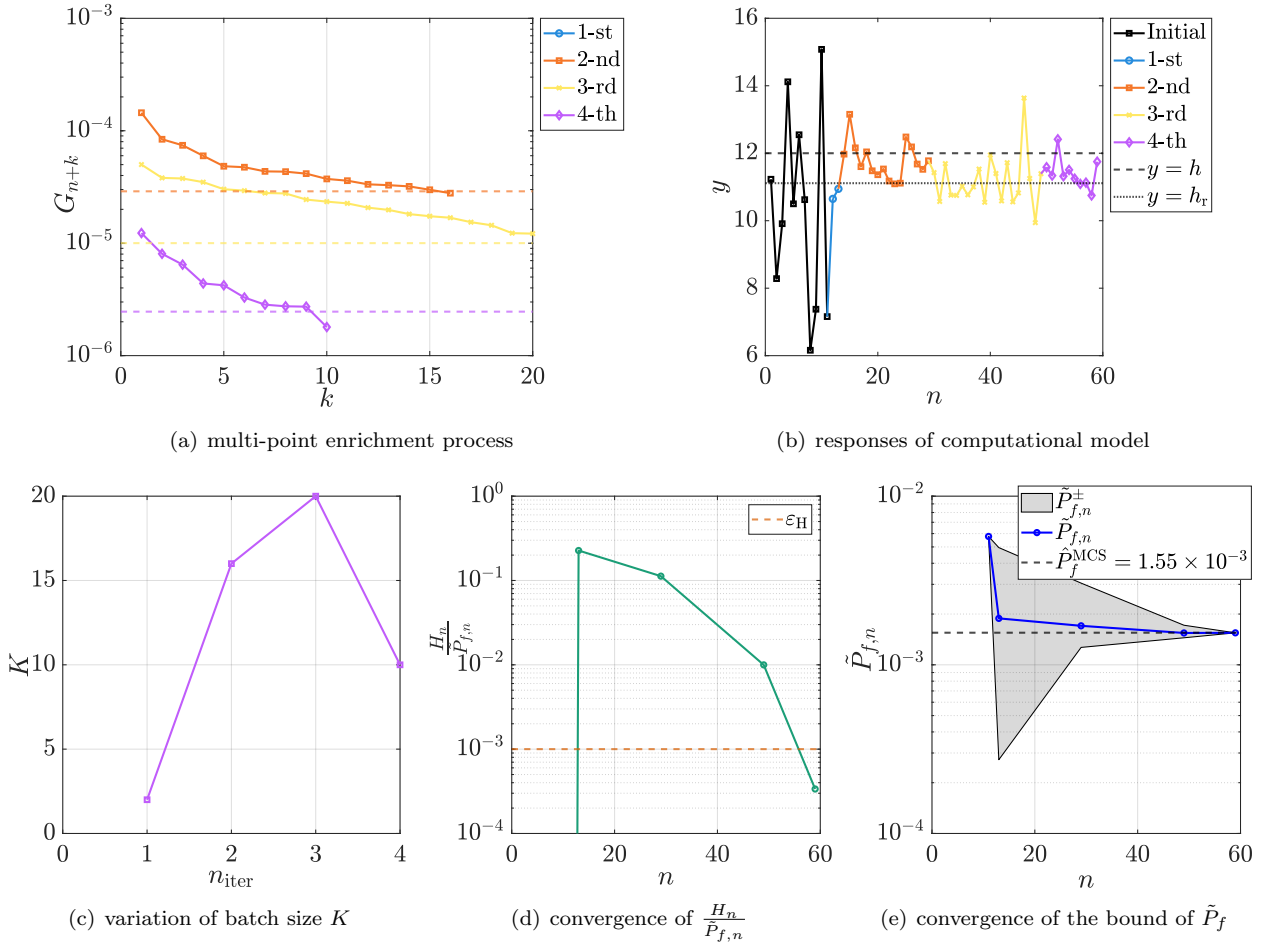
Finite element analysis of this truss is conducted with an in-house MATLAB code. Of interest is the mid-span deflection $y = U(\boldsymbol{x})$, and the failure threshold is set as 12 mm. Hence, the P_f is defined as

$$P_f = \mathbb{P}(Y \geq 12) = \int_{12}^{+\infty} f_Y(y)\mathrm{d}y \tag{54}$$

Table 3: Random variables in the two-dimensional truss [39]

Variable	Unit	Distribution	Mean	Standard deviation
$E_1 - E_2$	Pa	Lognormal	2.1×10^{11}	2.1×10^{10}
A_1	m^2	Lognormal	2.0×10^{-3}	2.0×10^{-4}
A_2	m^2	Lognormal	1.0×10^{-3}	1.0×10^{-4}
$P_1 - P_6$	N	Gumbel	5×10^4	7.5×10^3

Fig. 10 presents one run of the proposed k-TIMSER (adaptive scheme) in the planar truss example. The G_{n+k} value of adding the k-th new sample in each iteration of multi-point enrichment process is illustrated in Fig. 10(a). Due to the inferior performance of Kriging in the first iteration, the corresponding values of G_{n+k} are of minor magnitude and are invisible in the logarithmic coordinates. Then, the G_{n+k} values in the remaining iterations are gradually reduced, and most of the new samples added by k-TIMSER are located in the actual ROI; see Fig. 10(b). As a result, the TIMSE H_n reduces significantly as the n_{iter} increases, and a good accordance between $\tilde{P}_f$ and $\widehat{P}_f^{\text{MCS}}$ is finally achieved; see Figs. 10(d) and 10(e).



(a) multi-point enrichment process

(b) responses of computational model

(c) variation of batch size K

(d) convergence of $\frac{H_n}{\tilde{P}_{f,n}}$

(e) convergence of the bound of $\tilde{P}_f$

Figure 10: The k-TIMSER (adaptive scheme) in the planar truss example

Table 4 gives a comparison of different reliability methods for this planar truss example. In comparison with the existing parallel AK-MCS, the ALR module in Uqlab and the PABQ, the proposed k-TIMSER exerts fair advantage in terms of n_{iter} and n_{call}. Since the TIMSE H_n is dramatically reduced (Fig. 10(d)), the Kriging-induced error $\delta^{(2)}_{\tilde{P}_f}$ is always smaller than 0.004 %. Then, the overall error $\delta^{(1)}_{\tilde{P}_f}$ is always smaller than 0.17 %. In comparison with the pure PDEM, the k-TIMSER provides comparable accuracy of $\tilde{P}_f$ at the cost of only 3.5%

of n_{call}. Besides, since a single run of finite element analysis of this truss is relatively cheap, the time t_{iter} of a single iteration is dominated by the time t_4 of learning function. In this way, although the k-TIMSER needs smaller value of n_{iter} than the ALR module in Uqlab, its t_{c} is greater than that of the ALR module in Uqlab.

Table 4: Reliability results in the planar truss example

Method		$\mathbb{E}[n_{\text{iter}}]$	$\mathbb{E}[n_{\text{call}}]$	$\mathbb{E}[\tilde{P}_f](\times 10^{-3})$	$\text{COV}[\tilde{P}_f](\%)$	$\mathbb{E}[\delta_{\tilde{P}_f}](\%)$	$\mathbb{E}[t_{\text{c}}(\text{s})]$	References
MCS	-	-	10^6	1.550	-	-	-	[39]
AK-MCS	$K=1$	272	283	1.520	-	1.935	-	[39]
	$K=6$	26	162	1.530	-	1.290	-	[39]
ALR in UQLAB	$K=5$	14.6	79	1.558	1.870	1.505	148.3	-
	$K=10$	7.4	75	1.574	2.455	1.992	29.2	-
	$K=15$	6.3	90.5	1.568	2.118	1.816	31.8	-
	$K=20$	4.8	87	1.589	3.036	2.767	29.2	-
PABQ	$K=5$	14.9	79.5	1.360	5.553	12.243	230.1	-
	$K=10$	8.2	82	1.377	3.303	11.174	119.6	-
	$K=15$	6	85	1.427	4.487	8.310	89.1	-
	$K=20$	5.1	92	1.428	6.361	9.342	79.1	-
PDEM	-	-	2×10^3	1.547	-	0.164	215.3	-
k-TIMSER	$K=1$	43.1	53.1	1.548	0.005	0.161 (4.641×10^{-3})	348.2	-
	$K=5$	10	56	1.547	0.003	0.163 (1.258×10^{-3})	92.7	-
	$K=10$	5.9	60	1.547	0.001	0.164 (1.008×10^{-4})	58.9	-
	$K=15$	4.5	63.5	1.547	0.004	0.162 (2.133×10^{-3})	52.3	-
	$K=20$	4	71	1.547	0.001	0.164 (2.004×10^{-12})	49.6	-
	Adaptive	5.7	61.5	1.547	0.001	0.164 (1.981×10^{-4})	54.2	-

5.3. Seismic reliability analysis of a spatial reinforced concrete frame

The third example considers a practical six-story, three-bay reinforced concrete frame under bidirectional seismic ground motion excitation, with both geometric layout and reinforcement details displayed in Fig. 11. Finite element model of this frame is built using the OpenSees software [57]. Both beams and columns are modeled by the force-based elements with fiber-discretized cross section. The uniaxial consecutive relationships of rebar and concrete are described by Steel 01 and Concrete 01 models, respectively. The thickness of concrete slab at each floor is set as 110 mm, and its weight is applied on the surrounding beams. Rayleigh damping is specified, with damping ratio 5%. The related material parameters are viewed as independent random variables, with their statistical information given in Table 5.

Table 5: Random variables in the reinforced concrete frame

Variables	Description	Distribution	Mean	COV
f_{cc}(MPa)	Confined concrete compressive strength	Lognormal	35	0.1
ε_{cc}	Confined concrete strain at maximum strength	Lognormal	0.005	0.05
f_{cu}(MPa)	Confined concrete crushing strength	Lognormal	25	0.1
ε_{cu}	Confined concrete strain at crushing strength	Lognormal	0.02	0.05
f_c(MPa)	Unconfined concrete compressive strength	Lognormal	27	0.1
ε_c	Unconfined concrete strain at maximum strength	Lognormal	0.002	0.05
f_u(MPa)	Unconfined concrete crushing strength	Lognormal	10	0.1
ε_u	Unconfined concrete strain at crushing strength	Lognormal	0.006	0.05
f_y(MPa)	Yield strength of rebar	Lognormal	400	0.1
E_0(GPa)	Elastic modulus of rebar	Lognormal	200	0.1
b	Strain hardening ratio of rebar	Lognormal	0.007	0.05

Figs. 12(a) and 12(b) present the time histories of the amplitude-normalized N-S and W-E components of El-Centro accelerogram, respectively. The two horizontal components are applied on the x- and z-directions of this frame, respectively. Then, the corresponding amplitude coefficients are taken as Gaussian random variables, with the means both 4 and the COVs both 0.1.

Figs. 13(a) and 13(b) display the hysteretic curves of concrete and rebar at the end section of corner column in the bottom floor, respectively. Then, Fig. 13(c) illustrates the inter-story force-deformation curve of the bottom column in the z-direction. Clearly, strong nonlinearity is observed in both material- and structure-levels.

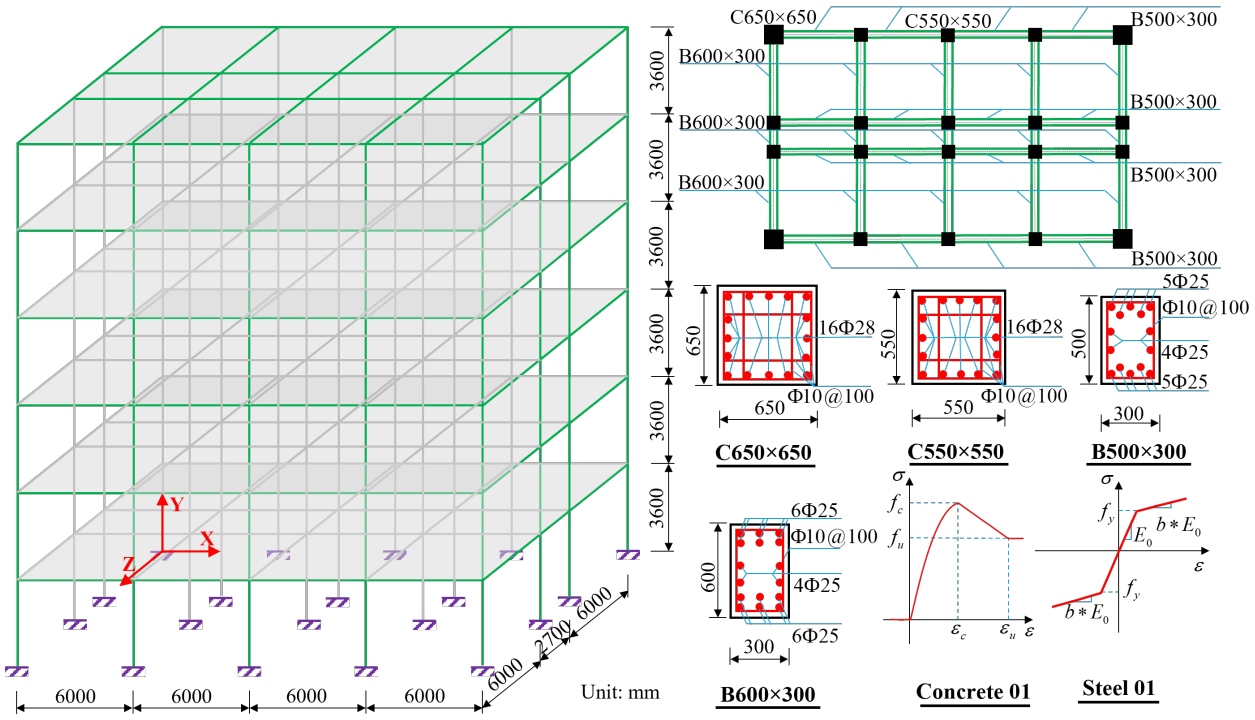
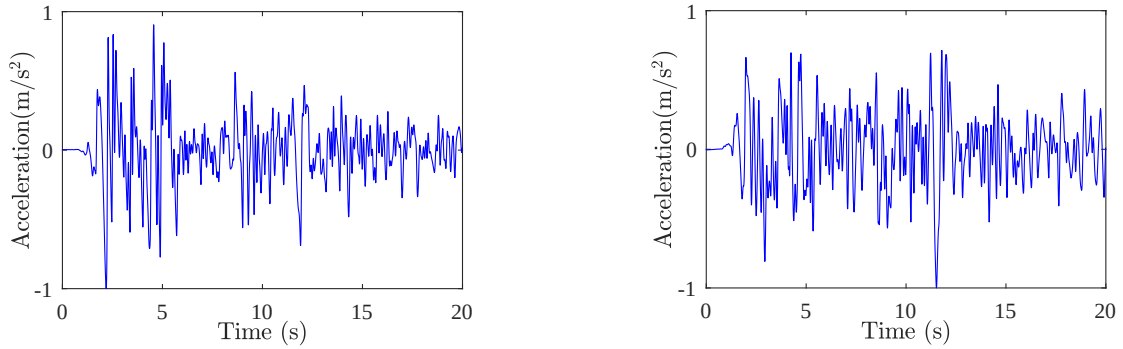
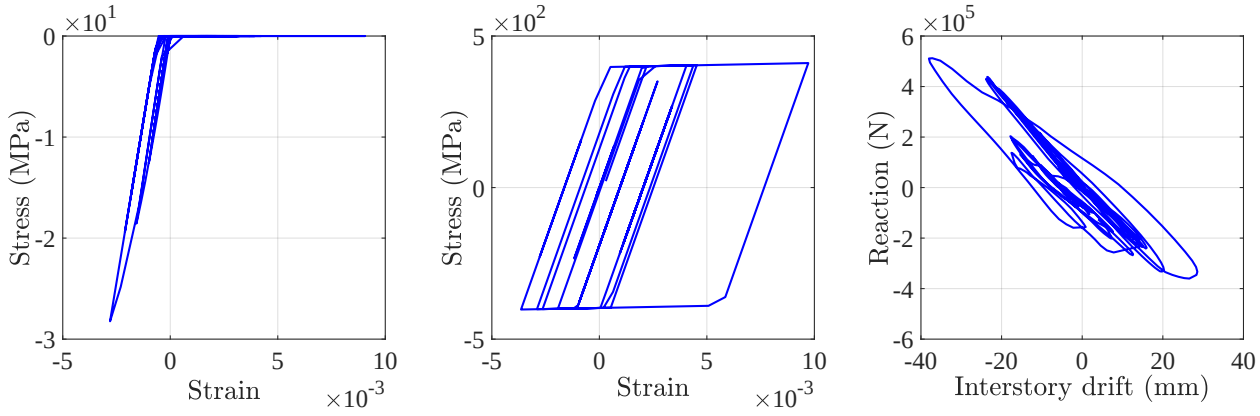


Figure 11: A six-story three-bay reinforced concrete frame



(a) N-S component (b) W-E component

Figure 12: Amplitude-normalized El-Centro accelerograms in two orthogonal directions



(a) stress-strain curve of concrete (b) stress-strain curve of rebar (c) hysteretic curve of the first floor

Figure 13: Nonlinear hysteretic behaviors of spatial reinforced concrete frame

Of interest are the inter-story drifts, $U_{i,j}(\boldsymbol{X},t), i=1,\ldots,6, j=1,2$, between the i- and $(i-1)$-th floor at the j-th direction. The maximum allowable inter-story drift is set as 72 mm. Then, the *system* failure probability

P_f is defined as

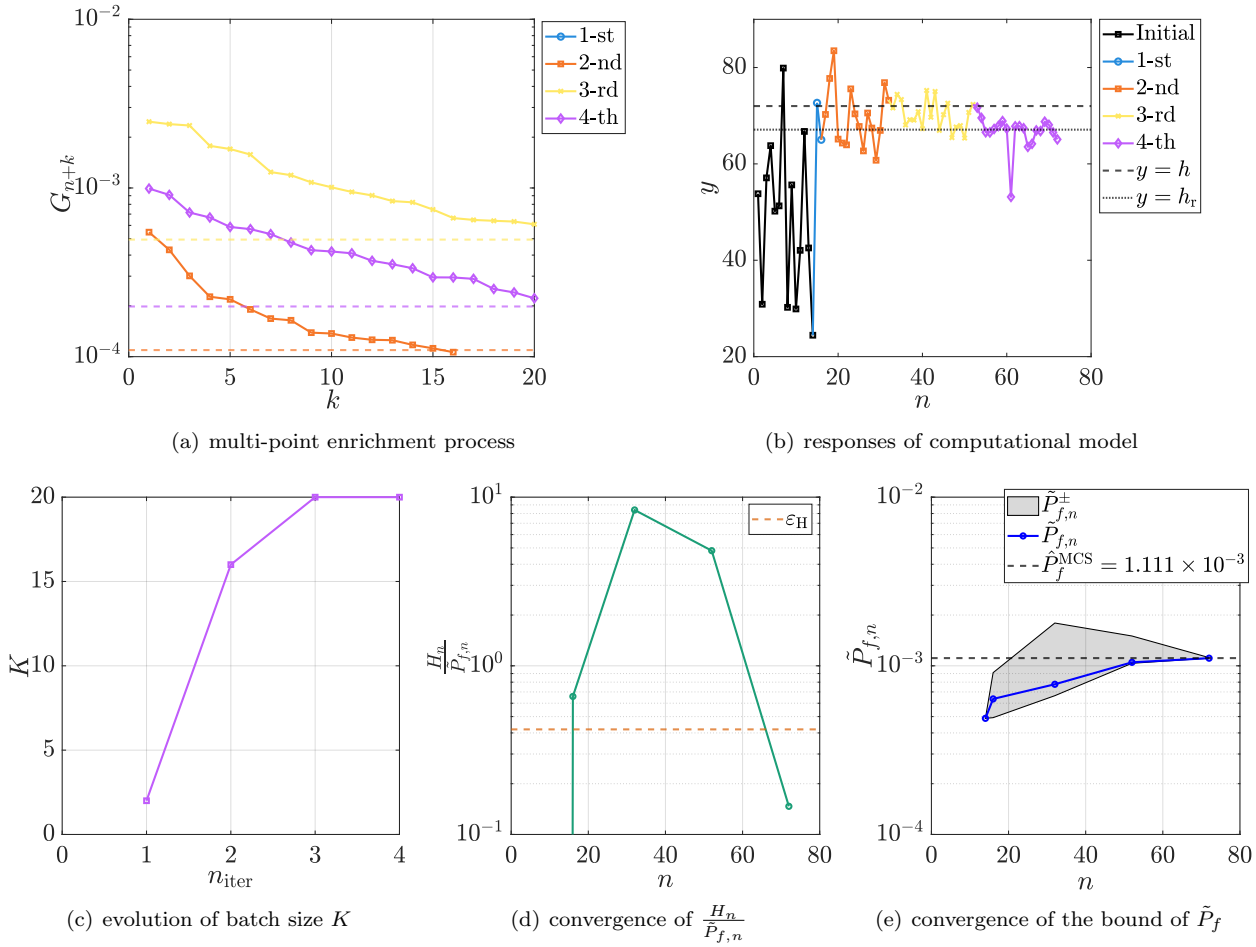
$$P_f = \mathbb{P}\left(\bigcup_{i=1}^{6}\bigcup_{j=1}^{2}(\exists t \in [0, 20\mathrm{s}], |U_{i,j}(\boldsymbol{X}, t)| \geq 72)\right), \tag{55}$$

which can be further recast as

$$P_f = \mathbb{P}(Y \geq 72) = \int_{72}^{+\infty} f_Y(y)\mathrm{d}y, \tag{56}$$

with $Y = \max\limits_{1\leq i\leq 6}\left(\max\limits_{1\leq j\leq 2}\left(\max\limits_{t\in[0,20\mathrm{s}]}|U_{i,j}(\boldsymbol{X}, t)|\right)\right)$.

Fig. 14 illustrates one run of the proposed k-TIMSER (adaptive scheme) in the reinforced concrete frame example. The G_k values in the first iteration are invisible in the logarithmic coordinate of Fig. 14(a), due to their minimal magnitude. Then, at least 16 new samples are added at each subsequent iteration, most of which are located in the ROI; see Fig. 14(b). Consequently, both substantial reduction of TIMSE H_n and good accordance between $\tilde{P}_f$ and $\hat{P}_f^{\mathrm{MCS}}$ are achieved at the end of this algorithm; see Figs. 14(d) and 14(e).



(a) multi-point enrichment process

(b) responses of computational model

(c) evolution of batch size K

(d) convergence of $\frac{H_n}{\tilde{P}_{f,n}}$

(e) convergence of the bound of $\tilde{P}_f$

Figure 14: The k-TIMSER (adaptive scheme) in the reinforced concrete frame example

Table 6 lists the results of different reliability methods in the reinforced concrete frame example. The parallel AK-MCS, the ALR module in Uqlab, and the PABQ all need at least 200 runs of finite element analysis, rendering them very time-consuming. On the contrary, the proposed k-TIMSER achieves better accuracy of $\tilde{P}_f$ at the cost of only 88 runs of finite element analysis at most. In the k-TIMSER, the Kriging-induced error $\delta_{\tilde{P}_f}^{(3)}$ is at most 0.005%, and its overall error $\delta_{\tilde{P}_f}^{(1)}$, say around 0.1%, is very close to that of the pure PDEM with 'over-kill' setting. Although increasing the size n_{rp} of $\mathcal{X}_{\mathrm{rp}}$ may further reduce the PDEM-caused error, such a minimal value of $\delta_{\tilde{P}_f}^{(1)}$ in the k-TIMSER is sufficient for practical engineering problems. More importantly, since the time, t_3, of finite-element analysis of this reinforced concrete frame is much greater than that, t_4, of k-TIMSER, the running time, t_{iter}, of an iteration is governed by t_3. Then, the remarkable advantage of k-TIMSER in terms of reducing n_{iter} is readily converted to that of reducing the total computational time t_{c}. It is observed that the t_{c} of the proposed k-TIMSER is only at most 35.7% of other active learning reliability methods. Finally, the results of Tables 4 and 6 indicate that the advantage of the total computational time

t_c becomes very evident in the proposed k-TIMSER, when computationally-expensive reliability problems are considered.

All the four parallel active learning reliability methods in Table 6 are faced with both the approximation error $\delta^{(3)}_{\tilde{P}_f}$ due to the reliability estimation algorithm and the Kriging-induced error $\delta^{(2)}_{\tilde{P}_f}$. For example, the approximation error in the ALR module in Uqlab results from the Markov chain Monte Carlo sampling, and the approximation error in the parallel AK-MCS stems from the Monte Carlo sampling. Although the $\delta^{(3)}_{\tilde{P}_f}$ in the parallel AK-MCS is theoretically minimal, its Kriging-induced error is relatively significant, due to so vast number of Monte Carlo samples generated in the 'over-kill' setting. This is the reason why the overall error of the parallel AK-MCS, say 10%, is greater than that of the ALR module in Uqlab, say 5%, and the proposed k-TIMSER, say 0.1%. This also justifies the usage of more advanced reliability estimation algorithms, e.g., SuS or PDEM, rather than the crude MCS, in the ALR framework [13, 14]. Further, in the 'over-kill' setting, the sample size is generally $\mathcal{O}(10^{5-8})$ in the parallel AK-MCS, the ALR module in Uqlab, and PABQ, while the n_{rp} is generally $\mathcal{O}(10^3)$ in PDEM. Such relatively small sample size of PDEM is the main reason why the Kriging-induced error can be reduced to such a minimal level in the k-TIMSER.

Table 6: Reliability results in the reinforced concrete frame example

Method		$\mathbb{E}[n_{\mathrm{iter}}]$	$\mathbb{E}[n_{\mathrm{call}}]$	$\mathbb{E}[\tilde{P}_f](\times 10^{-3})$	$\mathrm{COV}[\tilde{P}_f](\%)$	$\mathbb{E}[\delta_{\tilde{P}_f}](\%)$	$\mathbb{E}[t_c](\mathrm{s})$
MCS	-	-	2×10^5	1.111	-	-	8.1×10^6
AK-MCS	$K=5$	>39	>200	9.778	8.099	11.991	$>4.7\times10^3$
	$K=10$	>20	>200	9.860	9.744	11.252	$>3.0\times10^3$
	$K=15$	>14	>200	1.022	4.053	8.023	$>2.5\times10^3$
	$K=20$	>11	>200	1.027	4.721	7.544	$>2.3\times10^3$
ALR in UQLAB	$K=5$	>39	>200	1.083	3.786	3.550	$>4.6\times10^3$
	$K=10$	>20	>200	1.103	3.822	3.185	$>3.5\times10^3$
	$K=15$	>14	>200	1.113	6.008	5.172	$>3.4\times10^3$
	$K=20$	>11	>200	1.099	4.761	5.323	$>2.1\times10^3$
PABQ	$K=5$	>39	>200	1.059	32.035	25.104	$>3.7\times10^3$
	$K=10$	>20	>200	1.009	5.147	9.287	$>2.7\times10^3$
	$K=15$	>14	>200	0.986	15.359	12.724	$>2.5\times10^3$
	$K=20$	>11	>200	1.026	12.655	12.043	$>2.4\times10^3$
PDEM	-	-	2×10^3	1.110	-	0.102	1.3×10^5
k-TIMSER	$K=1$	58.5	71.5	1.110	0.003	0.106 (1.825×10^{-3})	3.4×10^3
	$K=5$	13	74	1.110	0.002	0.104 (1.687×10^{-3})	8.3×10^2
	$K=10$	7.4	78	1.110	0.001	0.102 (3.637×10^{-4})	8.1×10^2
	$K=15$	5.4	80	1.110	0.012	0.108 (5.695×10^{-3})	7.1×10^2
	$K=20$	4.7	88	1.110	0.001	0.103 (7.673×10^{-4})	7.5×10^2
	Adaptive	5.7	78.8	1.111	0.002	0.104 (1.426×10^{-3})	7.1×10^2

5.4. Discussions

Fig. 15 summarizes the results of both the prescribed scheme and the adaptive scheme of k-TIMSER above. On the one hand, the n_{iter} values needed by the prescribed scheme, marked as solid lines in Fig. 15, are generally reduced with the increasing of the prescribed batch size K. However, the downward trend varies with the reliability problems at hand. For example, in the four-branch function (Sections 5.1.1 and 5.1.2), the n_{iter} is stalled after the K exceeds 5, implying that further increasing K will not reduce n_{iter} remarkably. By contrast, in the reinforced concrete frame example (Section 5.3), the n_{iter} is stalled after the K exceeds 15. Therefore, the turning point that corresponds to the shift from rapid to slight decrease of n_{iter} can be viewed as the optimal value of K in the prescribed scheme. Unfortunately, this optimal value of K is unknown *a priori*. On the other hand, the average K values needed by the adaptive scheme, plotted as dashed lines in Fig. 15, are observed to be close to the so-called 'optimal' batch size to some extent. This implies that the adaptive scheme allows to account for the peculiar characteristics of different reliability problems, achieving fair savings of computational model evaluations.

6. Concluding remarks

A cost-effective parallel active learning reliability method is developed based on the multi-point look-ahead paradigm. First, the overall error of active learning reliability analysis is decomposed into two separate parts, i.e., the approximation error due to reliability estimation algorithm and the surrogate-induced error arising

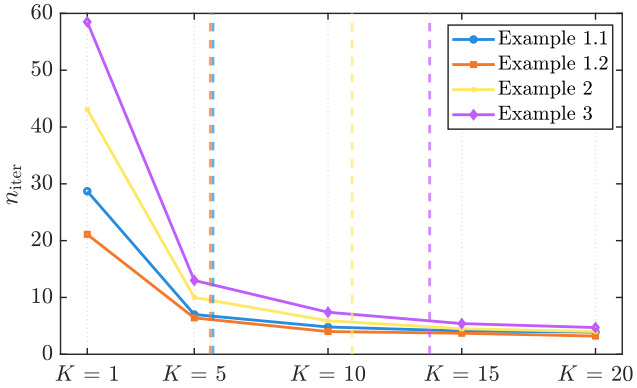


Figure 15: Comparison between the prescribed scheme and the adaptive scheme in the k-TIMSER

from the limited ED. Then, a metric called TIMSE is defined to measure the epistemic uncertainty of Kriging-estimated failure probability in the framework of PDEM. With the aim of reducing TIMSE, three key ingredients are developed: (i) a multi-point learning function called k-TIMSER; (ii) a hybrid convergence criterion; (iii) an adaptive scheme of identifying the rational size of batch of new samples added per iteration. The salient feature that distinguishes the proposed method from most of existing parallel active learning reliability methods lies in that the multi-point enrichment process is conducted based on the learning function k-TIMSER itself, without resorting to any additional parallel enrichment strategy. The proposed method is testified on three examples, where comparisons are made against several existing ones. Finally, some concluding remarks are given as follows.

- In the framework of PDEM, the metric TIMSE is proved as the upper bound of Kriging-induced error. Hence, it is a fair measure of epistemic uncertainty of Kriging-based failure probability estimation, due to only very limited evaluations of computational model in the ED.
- Thanks to the elegant definition of TIMSE, the resulting learning function k-TIMSER is analytically tractable, which allows quantifying the global gain of adding a batch of $k(>1)$ new points on the reduction of TIMSE in expectation.
- The number of new samples added by the k-TIMSER per iteration can be identified either by a traditional prescribed scheme or by a newly-developed adaptive scheme. In comparison with the prescribed scheme, the adaptive scheme gains a fair balance between the total computational time and the computing resource consumption.
- Thanks to the core role of TIMSE, the four key ingredients of parallel active learning reliability analysis, i.e., the PDEM, the Kriging, the learning function k-TIMSER (and its parallelization), and the hybrid convergence criterion, are assembled in a consistent way. Hence, the proposed multi-point look-ahead paradigm is theoretically rigorous and numerically elegant.

It is admitted that the proposed method may underperform in small failure probabilities, say $\leq \mathcal{O}(10^{-5})$, due to the limitation of PDEM. Besides, more reasonable, i.e., less conservative, measure of epistemic uncertainty about Kriging-based failure probability estimate can be defined to develop more fair form of learning function. They will be further investigated in the near future.

CRediT authorship contribution statement

Tong Zhou: Conceptualization, Formal analysis, Methodology, Software, Validation, Visualization, Writing - original draft. **Tong Guo**: Writing - review & editing. **Chao Dang**: Conceptualization, Writing -review & editing. **Lei Jia**: Writing - review & editing. **You Dong**: Funding acquisition, Writing -review & editing.

Conflict of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability

Data will be made available on request.

Acknowledgements

The support of the National Natural Science Foundation of China (Grant Nos. 52408222, 52078448), the Research Grants Council of Hong Kong (Grant Nos. PolyU 15225722 and PolyU 15221521) are highly appreciated.

Appendix A. Iterative identification of h_{r}

First, according to Eq. (15), the h_{r} is exactly the infimum of all infeasible values of the boundary η, that is,

$$h_{\mathrm{r}} = \inf\left\{\eta \in \mathbb{R} : \frac{P_f - \sum_{i=1}^{n_{\mathrm{rp}}} P_f^{(i)} \mathbb{1}\left(\mathcal{M}\left(\boldsymbol{x}^{(i)}\right) \geq \eta\right)}{P_f} \geq \varepsilon_{\mathrm{r}}\right\}, \tag{A.1}$$

where $\inf\{\cdot\}$ is the infimum of a set.

Therefore, the h_{r} can be sequentially identified through trial and error, as detailed in Algorithm A.1.

Algorithm A.1 Procedure of identifying h_{r}

Input: $\mathcal{Y}_{\mathrm{rp}} = \{y^{(i)}, i = 1, \ldots, n_{\mathrm{rp}}\}$, $\mathcal{P}_{\mathrm{rp}} = \{p^{(i)}, i = 1, \ldots, n_{\mathrm{rp}}\}$, and the failure threshold h.

1: Initialize: $\eta \leftarrow h$, $\eta_{\mathrm{inc}} \leftarrow h/50$.
2: Solve the GDEE based on both $\mathcal{Y}_{\mathrm{rp}}$ and $\mathcal{P}_{\mathrm{rp}}$, resulting in $\{P_f^{(i)}, i = 1, \ldots, n_{\mathrm{rp}}\}$ and P_f. ▷ Eqs. (7) - (10)
3: **while** $\frac{P_f - \sum_{i=1}^{n_{\mathrm{rp}}} P_f^{(i)} \mathbb{1}\left(y^{(i)} \geq \eta\right)}{P_f} \geq \varepsilon_{\mathrm{r}}$ **do** ▷ Eq. (A.1)
4: $\quad \eta \leftarrow \eta - \eta_{\mathrm{inc}}$;
5: **end while**
6: $h_{\mathrm{r}} = \eta$.

Output: h_{r}.

Appendix B. Basics of Kriging

Kriging assumes the response of $\mathcal{M}(\boldsymbol{x})$ as one realization of a underlying Gaussian process, expressed as [17]

$$\mathcal{M}(\boldsymbol{x}) \approx \widehat{\mathcal{M}}_n(\boldsymbol{x}) = \boldsymbol{\beta}^{\top} \boldsymbol{f}(\boldsymbol{x}) + \sigma^2 Z(\boldsymbol{x}), \tag{B.1}$$

where the trend function $\boldsymbol{\beta}^{\top} \boldsymbol{f}(\boldsymbol{x}) = \beta_0 + \sum_{l=1}^{d} \beta_l x_l$, with $\boldsymbol{f}(\boldsymbol{x}) = \{1, x_1, \ldots, x_d\}$ a set of basis functions and $\boldsymbol{\beta} = \{\beta_l, l = 0, \ldots, d\}$ a set of unknown coefficients. σ^2 is the variance of Gaussian process; $Z(\boldsymbol{x})$ is a zero-mean, unit-variance Gaussian process fully described by a correlation function $R(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{\theta})$ with parameters $\boldsymbol{\theta}$. The Matern-5/2 kernel function is adopted here, expressed as [17]

$$R(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{\theta}) = \prod_{l=1}^{d} \left(1 + \sqrt{5}\frac{|x_l - x_l'|}{\theta_l} + \frac{5}{3}\left(\frac{|x_l - x_l'|}{\theta_l}\right)^2\right) \exp\left(-\sqrt{5}\frac{|x_l - x_l'|}{\theta_l}\right), \tag{B.2}$$

where $\boldsymbol{x}$ and $\boldsymbol{x}'$ are two realizations of $\boldsymbol{X}$; the kernel parameters $\boldsymbol{\theta} = \{\theta_l > 0, l = 1, \ldots, d\}$.

Suppose that an ED $\mathcal{D}_n = \{\mathcal{X}_n, \mathcal{Y}_n\}$ consists of a set of n input samples $\mathcal{X}_n = \{\boldsymbol{x}^{(i)}, i = 1, \ldots, n\}$ and the associated computational model responses $\mathcal{Y}_n = \{y^{(i)} = \mathcal{M}\left(\boldsymbol{x}^{(i)}\right), i = 1, \ldots, n\}$, both $\boldsymbol{\beta}$ and σ^2 are estimated as

$$\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{F}^{\top} \boldsymbol{R}^{-1} \boldsymbol{F}\right)^{-1} \boldsymbol{F}^{\top} \boldsymbol{R}^{-1} \mathcal{Y}_n, \tag{B.3}$$

$$\widehat{\sigma}^2 = \frac{1}{n}(\mathcal{Y}_n - \boldsymbol{F}\boldsymbol{\beta})^{\top} \boldsymbol{R}^{-1} (\mathcal{Y}_n - \boldsymbol{F}\boldsymbol{\beta}), \tag{B.4}$$

where $\boldsymbol{F} \coloneqq [f_j(\boldsymbol{x}^{(i)})]_{1 \leq i \leq n, 1, \leq j \leq d+1}$ is an information matrix; $\boldsymbol{R} \coloneqq [R(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}; \boldsymbol{\theta})]_{1 \leq i,j \leq n}$ is a matrix of correlations between all points in $\mathcal{X}_n$. Both $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma}^2$ depend on $\boldsymbol{\theta}$, which can be estimated as [17]

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta} \in \Theta} \widehat{\sigma}^2 |\boldsymbol{R}|^{\frac{1}{n}}, \tag{B.5}$$

where Θ is the support of $\boldsymbol{\theta}$.

Then, the Kriging prediction conditional on $\mathcal{D}_n$ still follows a Gaussian process, that is,

$$\widehat{\mathcal{M}}_n\left(\boldsymbol{x}\right) \sim \mathcal{GP}(\mu_n(\boldsymbol{x}), c_n(\boldsymbol{x}, \boldsymbol{x}')), \tag{B.6}$$

with its mean $\mu_n(\boldsymbol{x})$, variance $\sigma_n^2(\boldsymbol{x})$ and covariance $c_n(\boldsymbol{x},\boldsymbol{x}')$ expressed as [17]

$$\mu_n(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})^\top \widehat{\boldsymbol{\beta}} + \boldsymbol{r}(\boldsymbol{x})^\top \boldsymbol{R}^{-1}\left(\mathcal{Y}_n - \boldsymbol{F}\widehat{\boldsymbol{\beta}}\right), \tag{B.7}$$

$$\sigma_n^2(\boldsymbol{x}) = \widehat{\sigma}^2\left(1 - \boldsymbol{r}(\boldsymbol{x})^\top \boldsymbol{R}^{-1}\boldsymbol{r}(\boldsymbol{x}) + \boldsymbol{u}(\boldsymbol{x})^\top\left(\boldsymbol{F}^\top \boldsymbol{R}^{-1}\boldsymbol{F}\right)^{-1}\boldsymbol{u}(\boldsymbol{x})\right), \tag{B.8}$$

$$c_n(\boldsymbol{x},\boldsymbol{x}') = \widehat{\sigma}^2\left(R(\boldsymbol{x},\boldsymbol{x}') - \boldsymbol{r}(\boldsymbol{x})^\top \boldsymbol{R}^{-1}\boldsymbol{r}(\boldsymbol{x}') + \boldsymbol{u}(\boldsymbol{x})^\top\left(\boldsymbol{F}^\top \boldsymbol{R}^{-1}\boldsymbol{F}\right)^{-1}\boldsymbol{u}(\boldsymbol{x}')\right), \tag{B.9}$$

respectively, where the subscript n implies that these quantities condition on $\mathcal{D}_n$; $\boldsymbol{r}(\boldsymbol{x}) = \left[R(\boldsymbol{x},\boldsymbol{x}^{(1)}),\ldots,R(\boldsymbol{x},\boldsymbol{x}^{(n)})\right]^\top$; $\boldsymbol{u}(\boldsymbol{x}) = \boldsymbol{F}^\top \boldsymbol{R}^{-1}\boldsymbol{r}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x})$.

The $\mu_n(\boldsymbol{x})$ is a natural estimate of $\widehat{\mathcal{M}}_n(\boldsymbol{x})$, while $\sigma_n^2(\boldsymbol{x})$ can be viewed as a local measure of epistemic uncertainty of $\widehat{\mathcal{M}}_n(\boldsymbol{x})$, due to only very limited computational model evaluations in $\mathcal{D}_n$.

Appendix C. Four existing pointwise learning functions

The U and EFF are two best-known learning functions in the active learning-based simulation methods. First, the U function is expressed as [24]

$$U(\boldsymbol{x}) = \frac{|\mu_n(\boldsymbol{x})|}{\sigma_n(\boldsymbol{x})}, \tag{C.1}$$

whereby the best next point is selected as $\boldsymbol{x}^{(n+1)} = \arg\min_{\boldsymbol{x}} U(\boldsymbol{x})$.

Second, the EFF is expressed as [23]

$$\begin{aligned}\mathrm{EFF}(\boldsymbol{x}) = {}& \mu_n(\boldsymbol{x})\left[2\Phi\left(-\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right) - \Phi\left(\frac{-\varepsilon-\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right) - \Phi\left(\frac{\varepsilon-\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right)\right] \\ & - \sigma_n(\boldsymbol{x})\left[2\phi\left(-\frac{\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right) - \phi\left(\frac{-\varepsilon-\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right) - \phi\left(\frac{\varepsilon-\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right)\right] \\ & + \varepsilon\left[\Phi\left(\frac{\varepsilon-\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right) - \Phi\left(\frac{-\varepsilon-\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right)\right],\end{aligned} \tag{C.2}$$

where $\varepsilon = 2\sigma_n(\boldsymbol{x})$ here. Then, $\boldsymbol{x}^{(n+1)} = \arg\max_{\boldsymbol{x}} \mathrm{EFF}(\boldsymbol{x})$.

The PIE and PEIF are two existing learning functions in the active learning-based PDEM methods. First, the PIE is expressed as [27]

$$\mathrm{PIE}(\boldsymbol{x}) = \left|\left(\ln(\sqrt{2\pi}\sigma_n(\boldsymbol{x})) + \frac{1}{2}\right)\left(1 - \Phi\left(\frac{h-\varepsilon-\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right)\right) + \frac{h-\varepsilon-\mu_n(\boldsymbol{x})}{2}\phi\left(\frac{h_\mathrm{r}-\varepsilon-\mu_n(\boldsymbol{x})}{\sigma_n(\boldsymbol{x})}\right)\right|, \tag{C.3}$$

where h is the failure threshold; $\varepsilon = \sigma_n(\boldsymbol{x})$ here. Then, $\boldsymbol{x}^{(n+1)} = \arg\max_{\boldsymbol{x}} \mathrm{PIE}(\boldsymbol{x})$.

Second, the PEIF is given by [28]

$$\mathrm{PEIF}(\boldsymbol{x}) = (\mu_n(\boldsymbol{x}) - h_\mathrm{r})\Phi\left(\frac{\mu_n(\boldsymbol{x})-h_\mathrm{r}}{\sigma_n(\boldsymbol{x})}\right) + \sigma_n(\boldsymbol{x})\phi\left(\frac{\mu_n(\boldsymbol{x})-h_\mathrm{r}}{\sigma_n(\boldsymbol{x})}\right), \tag{C.4}$$

whereby $\boldsymbol{x}^{(n+1)} = \arg\max_{\boldsymbol{x}} \mathrm{PEIF}(\boldsymbol{x})$.

Appendix D. Proof of Proposition 1

Proof. First, according to Eqs. (16) and (17), there exists

$$\mathbb{E}_n\left[\left(\widehat{P}_{f,n} - \tilde{P}_{f,n}\right)^2\right] = \mathbb{E}_n\left[\left(\sum_{i=1}^{n_\mathrm{rp}} \widehat{P}_{f,n}^{(i)} - \sum_{i=1}^{n_\mathrm{rp}} \tilde{P}_{f,n}^{(i)}\right)^2\right] = \mathbb{E}_n\left[\left(\sum_{i=1}^{n_\mathrm{rp}}\left(\widehat{P}_{f,n}^{(i)} - \tilde{P}_{f,n}^{(i)}\right)\right)^2\right]. \tag{D.1}$$

Based on the Cauchy-Schwarz inequality, Eq. (D.1) satisfies

$$\mathbb{E}_n\left[\left(\widehat{P}_{f,n} - \tilde{P}_{f,n}\right)^2\right] \le \mathbb{E}_n\left[n_\mathrm{rp}\sum_{i=1}^{n_\mathrm{rp}}\left(\widehat{P}_{f,n}^{(i)} - \tilde{P}_{f,n}^{(i)}\right)^2\right] = n_\mathrm{rp}\sum_{i=1}^{n_\mathrm{rp}}\mathbb{E}_n\left[\left(\widehat{P}_{f,n}^{(i)} - \tilde{P}_{f,n}^{(i)}\right)^2\right]. \tag{D.2}$$

Then, according to Eq. (13), $\mathbb{E}_n\left[\left(\widehat{P}_{f,n}^{(i)}-\tilde{P}_{f,n}^{(i)}\right)^2\right]$ can be expanded as

$$\begin{aligned}\mathbb{E}_n\left[\left(\widehat{P}_{f,n}^{(i)}-\tilde{P}_{f,n}^{(i)}\right)^2\right] &= \mathbb{E}_n\left[\left(\Gamma\left(\widehat{\mathcal{M}}_n(\boldsymbol{x}^{(i)})\right)\cdot p^{(i)}-\Gamma\left(\mu_n(\boldsymbol{x}^{(i)})\right)\cdot p^{(i)}\right)^2\right],\\ &= \left(p^{(i)}\right)^2\mathbb{E}_n\left[\left(\Gamma\left(\widehat{\mathcal{M}}_n(\boldsymbol{x}^{(i)})\right)-\Gamma\left(\mu_n(\boldsymbol{x}^{(i)})\right)\right)^2\right].\end{aligned} \tag{D.3}$$

Further, the first-order Taylor expansion of $\Gamma(\cdot)$ at $\mu_n(\boldsymbol{x})$ gives rise to

$$\Gamma\left(\widehat{\mathcal{M}}_n(\boldsymbol{x})\right)-\Gamma\left(\mu_n(\boldsymbol{x})\right)\approx\Gamma'\left(\mu_n(\boldsymbol{x})\right)\left(\widehat{\mathcal{M}}_n(\boldsymbol{x})-\mu_n(\boldsymbol{x})\right), \tag{D.4}$$

where $\Gamma'(\cdot)$ is the derivative of $\Gamma(\cdot)$. Hence, there exists

$$\left(\Gamma\left(\widehat{\mathcal{M}}_n(\boldsymbol{x})\right)-\Gamma\left(\mu_n(\boldsymbol{x})\right)\right)^2\approx\left[\Gamma'\left(\mu_n(\boldsymbol{x})\right)\right]^2\left(\widehat{\mathcal{M}}_n(\boldsymbol{x})-\mu_n(\boldsymbol{x})\right)^2, \tag{D.5}$$

then,

$$\mathbb{E}_n\left[\left(\Gamma\left(\widehat{\mathcal{M}}_n(\boldsymbol{x})\right)-\Gamma\left(\mu_n(\boldsymbol{x})\right)\right)^2\right]\approx\left[\Gamma'\left(\mu_n(\boldsymbol{x})\right)\right]^2\mathbb{E}_n\left[\left(\widehat{\mathcal{M}}_n(\boldsymbol{x})-\mu_n(\boldsymbol{x})\right)^2\right]=\left[\Gamma'\left(\mu_n(\boldsymbol{x})\right)\right]^2\sigma_n^2(\boldsymbol{x}). \tag{D.6}$$

Substitute Eq. (D.6) into Eq. (D.3), yielding

$$\mathbb{E}_n\left[\left(\widehat{P}_{f,n}^{(i)}-\tilde{P}_{f,n}^{(i)}\right)^2\right]\approx\left(p^{(i)}\right)^2\sigma_n^2(\boldsymbol{x}^{(i)})\left[\Gamma'\left(\mu_n(\boldsymbol{x}^{(i)})\right)\right]^2. \tag{D.7}$$

In this way, Eq. (D.2) can be further expressed as

$$\mathbb{E}_n\left[\left(\widehat{P}_{f,n}-\tilde{P}_{f,n}\right)^2\right]\le n_{\mathrm{rp}}\sum_{i=1}^{n_{\mathrm{rp}}}\left(p^{(i)}\right)^2\sigma_n^2(\boldsymbol{x}^{(i)})\left[\Gamma'\left(\mu_n(\boldsymbol{x}^{(i)})\right)\right]^2. \tag{D.8}$$

As plotted as the purple solid line in the top panel of Fig. 1(b), $\Gamma'(y)$ equals 0 when $y\in(-\infty,h_{\mathrm{r}}]$, and it achieves the maximum value when $y=h$. Generally, $\Gamma(y)$ varies slowly with y, and $\Gamma'(y)$ is often smaller than $\pi_n(\boldsymbol{x})$. Hence, Eq. (20) can be proved such that

$$\mathbb{E}_n\left[\left(\widehat{P}_{f,n}-\tilde{P}_{f,n}\right)^2\right]\le n_{\mathrm{rp}}\sum_{i=1}^{n_{\mathrm{rp}}}\left(p^{(i)}\right)^2\sigma_n^2(\boldsymbol{x}^{(i)})\pi_n(\boldsymbol{x}^{(i)})=H_n. \tag{D.9}$$

□

Appendix E. Multi-point Kriging update formulas

First, when providing an ED $\mathcal{D}_n=\{\mathcal{X}_n,\mathcal{Y}_n\}$ of size n, a Kriging $\widehat{\mathcal{M}}_n(\boldsymbol{x})$ can be trained, with its mean $\mu_n(\boldsymbol{x})$, variance $\sigma_n^2(\boldsymbol{x})$ and covariance $c_n(\boldsymbol{x},\boldsymbol{x}')$ given by Eqs. (B.7), (B.8) and (B.9), respectively.

Then, denote $\mathcal{X}_k^+=\{\boldsymbol{x}_+^{(1)},\ldots,\boldsymbol{x}_+^{(k)}\}^\top$ and $\mathcal{Y}_k^+=\{y_+^{(1)},\ldots,y_+^{(k)}\}^\top$ as a batch of $k(\ge1)$ new points and their computational model responses, respectively. When $\mathcal{D}_n$ is enriched with $\{\mathcal{X}_k^+,\mathcal{Y}_k^+\}$, the look-ahead mean, variance, and covariance of Kriging are given by [58]

$$\mu_{n+k}(\boldsymbol{x})=\mu_n(\boldsymbol{x})+c_n\left(\boldsymbol{x},\mathcal{X}_k^+\right)^\top\left(\mathcal{C}_k^+\right)^{-1}\left(\mathcal{Y}_k^+-\mu_n\left(\mathcal{X}_k^+\right)\right), \tag{E.1}$$

$$\sigma_{n+k}^2(\boldsymbol{x})=\sigma_n^2(\boldsymbol{x})-c_n\left(\boldsymbol{x},\mathcal{X}_k^+\right)^\top\left(\mathcal{C}_k^+\right)^{-1}c_n\left(\boldsymbol{x},\mathcal{X}_k^+\right), \tag{E.2}$$

$$c_{n+k}(\boldsymbol{x},\boldsymbol{x}')=c_n(\boldsymbol{x},\boldsymbol{x}')-c_n\left(\boldsymbol{x},\mathcal{X}_k^+\right)^\top\left(\mathcal{C}_k^+\right)^{-1}c_n\left(\boldsymbol{x}',\mathcal{X}_k^+\right), \tag{E.3}$$

which are called multi-point Kriging update formulas in [58]. $c_n\left(\boldsymbol{x},\mathcal{X}_k^+\right):=\left[c_n\left(\boldsymbol{x},\boldsymbol{x}_+^{(1)}\right),\ldots,c_n\left(\boldsymbol{x},\boldsymbol{x}_+^{(k)}\right)\right]^\top$ is a $k\times1$ vector of covariances between $\boldsymbol{x}$ and all points in $\mathcal{X}_k^+$; $\mathcal{C}_k^+:=\left[c_n\left(\boldsymbol{x}_+^{(i)},\boldsymbol{x}_+^{(j)}\right)\right]_{1\le i,j\le k}$ is a $k\times k$ matrix of covariances between all points in $\mathcal{X}_k^+$; $\mu_n(\mathcal{X}_k^+):=\left[\mu_n\left(\boldsymbol{x}_+^{(1)}\right),\ldots,\mu_n\left(\boldsymbol{x}_+^{(k)}\right)\right]^\top$ is the posterior means at all points in $\mathcal{X}_k^+$.

Appendix F. Proof of Proposition 2

Proof. In $\mathrm{TIMSER}_{n+k}(\mathcal{X}_k^+)$, the second term can be expanded as

$$\mathbb{E}_{\boldsymbol{U}_k^+}\left[\mathcal{H}_{n+k}(\mathcal{X}_k^+)\right] = n_{\mathrm{rp}} \sum_{i=1}^{n_{\mathrm{rp}}} (p^{(i)})^2 \sigma_{n+k}^2(\boldsymbol{x}^{(i)}) \mathbb{E}_{\boldsymbol{U}_k^+}\left[\Pi_{n+k}(\boldsymbol{x}^{(i)}; \mathcal{X}_k^+)\right], \tag{F.1}$$

which is based on the fact that $\sigma_{n+k}^2(\boldsymbol{x})$ is independent of $\boldsymbol{Y}_k^+$ (and further $\boldsymbol{U}_k^+$); see Eq. (E.2).

Then, according to Eq. (24), $\mathbb{E}_{\boldsymbol{U}_k^+}\left[\Pi_{n+k}(\boldsymbol{x}; \mathcal{X}_k^+)\right]$ is expressed as

$$\mathbb{E}_{\boldsymbol{U}_k^+}\left[\Pi_{n+k}(\boldsymbol{x}; \mathcal{X}_k^+)\right] = \int_{\mathbb{R}^k} \Phi\big(a(\boldsymbol{x}) + \boldsymbol{b}(\boldsymbol{x})^\top \mathcal{U}_k^+\big) f_{\boldsymbol{U}_k^+}(\mathcal{U}_k^+) \mathrm{d}\mathcal{U}_k^+, \tag{F.2}$$

where $f_{\boldsymbol{U}_k^+}(\mathcal{U}_k^+)$ is the joint PDF of $\boldsymbol{U}_k^+$.

The basic definition of $\Phi(\cdot)$ states that if $Z \sim \mathcal{N}(0,1)$ is a standard Gaussian random variable independent of $\boldsymbol{U}_k^+$, there exists

$$\Phi\big(a(\boldsymbol{x}) + \boldsymbol{b}(\boldsymbol{x})^\top \mathcal{U}_k^+\big) = \mathbb{P}\big(Z \le a(\boldsymbol{x}) + \boldsymbol{b}(\boldsymbol{x})^\top \mathcal{U}_k^+\big) = \mathbb{P}\big(Z \le a(\boldsymbol{x}) + \boldsymbol{b}(\boldsymbol{x})^\top \boldsymbol{U}_k^+ | \boldsymbol{U}_k^+ = \mathcal{U}_k^+\big). \tag{F.3}$$

Hence, according to the law of total probability, Eq. (F.2) is equivalent to

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{U}_k^+}\left[\Pi_{n+k}(\boldsymbol{x}; \mathcal{X}_k^+)\right] &= \int_{\mathbb{R}^k} \mathbb{P}\big(Z \le a(\boldsymbol{x}) + \boldsymbol{b}(\boldsymbol{x})^\top \boldsymbol{U}_k^+ | \boldsymbol{U}_k^+ = \mathcal{U}_k^+\big) f_{\boldsymbol{U}_k^+}(\mathcal{U}_k^+) \mathrm{d}\mathcal{U}_k^+, \\
&= \mathbb{P}\big(Z \le a(\boldsymbol{x}) + \boldsymbol{b}(\boldsymbol{x})^\top \boldsymbol{U}_k^+\big), \\
&= \mathbb{P}\big(Z - \boldsymbol{b}(\boldsymbol{x})^\top \boldsymbol{U}_k^+ \le a(\boldsymbol{x})\big), \\
&= \mathbb{P}(W \le a(\boldsymbol{x})),
\end{aligned} \tag{F.4}$$

where $W = Z - \boldsymbol{b}(\boldsymbol{x})^\top \boldsymbol{U}_k^+$ is a Gaussian random variable, with its mean and variance expressed as

$$\mu_W = \mathbb{E}\big[Z - \boldsymbol{b}(\boldsymbol{x})^\top \boldsymbol{U}_k^+\big] = 0 - \boldsymbol{b}(\boldsymbol{x})^\top \boldsymbol{0} = 0, \tag{F.5}$$

$$\begin{aligned}
\sigma_W^2 &= \mathbb{V}[Z - \boldsymbol{b}(\boldsymbol{x})^\top \boldsymbol{U}_k^+] = 1 + \boldsymbol{b}(\boldsymbol{x})^\top \mathcal{C}_k^+ \boldsymbol{b}(\boldsymbol{x}) = 1 + \frac{c_n(\boldsymbol{x}, \mathcal{X}_k^+)^\top (\mathcal{C}_k^+)^{-1}}{\sigma_{n+k}(\boldsymbol{x})} \mathcal{C}_k^+ \frac{(\mathcal{C}_k^+)^{-1} c_n(\boldsymbol{x}, \mathcal{X}_k^+)}{\sigma_{n+k}(\boldsymbol{x})} \\
&= 1 + \frac{c_n(\boldsymbol{x}, \mathcal{X}_k^+)^\top (\mathcal{C}_k^+)^{-1} c_n(\boldsymbol{x}, \mathcal{X}_k^+)}{\sigma_{n+k}^2(\boldsymbol{x})} = \frac{\sigma_n^2(\boldsymbol{x})}{\sigma_{n+k}^2(\boldsymbol{x})},
\end{aligned} \tag{F.6}$$

where $\mathbb{V}[\cdot]$ denotes the variance operator.

In this way, Eq. (F.4) is equivalent to

$$\mathbb{E}_n\left[\Pi_{n+k}(\boldsymbol{x}; \mathcal{X}_k^+)\right] = \Phi\left(\frac{a(\boldsymbol{x}) - \mu_W}{\sigma_W}\right) = \Phi\left(\frac{a(\boldsymbol{x}) - 0}{\frac{\sigma_n(\boldsymbol{x})}{\sigma_{n+k}(\boldsymbol{x})}}\right) = \Phi\left(\frac{\mu_n(\boldsymbol{x}) - h_{\mathrm{r}}}{\sigma_n(\boldsymbol{x})}\right) = \pi_n(\boldsymbol{x}), \tag{F.7}$$

which implies an interesting fact that in $\mathrm{TIMSER}_{n+k}(\mathcal{X}_k^+)$, the addition of $\mathcal{X}_k^+$ has no impact on $\mathbb{E}_n\left[\Pi_{n+k}(\boldsymbol{x}; \mathcal{X}_k^+)\right]$, but on the look-ahead variance $\sigma_{n+k}^2(\boldsymbol{x})$ solely; see Eq. (E.2).

Substitute Eq. (F.7) into Eq. (F.1), yielding

$$\mathbb{E}_{\boldsymbol{U}_k^+}\left[\mathcal{H}_{n+k}(\mathcal{X}_k^+)\right] = n_{\mathrm{rp}} \sum_{i=1}^{n_{\mathrm{rp}}} (p^{(i)})^2 \sigma_{n+k}^2(\boldsymbol{x}^{(i)}) \pi_n(\boldsymbol{x}^{(i)}). \tag{F.8}$$

Finally, substituting Eq. (F.8) into Eq. (26) proves the analytical expression of $\mathrm{TIMSER}_{n+k}(\mathcal{X}_k^+)$ in Eq. (28), that is,

$$\begin{aligned}
\mathrm{TIMSER}_{n+k}(\mathcal{X}_k^+) &= H_n - \mathbb{E}_{\boldsymbol{U}_k^+}\left[\mathcal{H}_{n+k}(\mathcal{X}_k^+)\right], \\
&= n_{\mathrm{rp}} \sum_{i=1}^{n_{\mathrm{rp}}} (p^{(i)})^2 \sigma_n^2(\boldsymbol{x}^{(i)}) \pi_n(\boldsymbol{x}^{(i)}) - n_{\mathrm{rp}} \sum_{i=1}^{n_{\mathrm{rp}}} (p^{(i)})^2 \sigma_{n+k}^2(\boldsymbol{x}^{(i)}) \pi_n(\boldsymbol{x}^{(i)}), \\
&= n_{\mathrm{rp}} \sum_{i=1}^{n_{\mathrm{rp}}} (p^{(i)})^2 \pi_n(\boldsymbol{x}^{(i)}) [\sigma_n^2(\boldsymbol{x}^{(i)}) - \sigma_{n+k}^2(\boldsymbol{x}^{(i)})].
\end{aligned} \tag{F.9}$$

□

References

[1] J. Li, J. Chen, Stochastic Dynamics of Structures, John Wiley & Sons, 2009.

[2] A. M. Hasofer, N. C. Lind, Exact and invariant second-moment code format., ASCE Journal of Engineering Mechanics 100 (1) (1974) 111–121. doi:10.1061/JMCEA3.0001848.

[3] K. Breitung, Asymptotic approximations for multinormal integrals, Journal of Engineering Mechanics 110 (3) (1984) 357–366. doi:10.1061/(ASCE)0733-9399(1984)110:3(357).

[4] C. Song, R. Kawai, Monte carlo and variance reduction methods for structural reliability analysis: A comprehensive review, Probabilistic Engineering Mechanics 73 (2023). doi:10.1016/j.probengmech.2023.103479.

[5] S. Engelund, R. Rackwitz, A benchmark study on importance sampling techniques in structural reliability, Structural Safety 12 (4) (1993) 255–276. doi:10.1016/0167-4730(93)90056-7.

[6] P. Koutsourelakis, H. Pradlwarter, G. Schuëller, Reliability of structures in high dimensions, part I: Algorithms and applications, Probabilistic Engineering Mechanics 19 (4) (2004) 409 – 417. doi:10.1016/j.probengmech.2004.05.001.

[7] S.-K. Au, J. Beck, Estimation of small failure probabilities in high dimensions by subset simulation, Probabilistic Engineering Mechanics 16 (4) (2001) 263–277. doi:10.1016/S0266-8920(01)00019-4.

[8] Y.-G. Zhao, Z.-H. Lu, Structural reliability: approaches from perspectives of statistical moments, John Wiley & Sons, 2021.

[9] T. Zhou, Y. Peng, Adaptive bayesian quadrature based statistical moments estimation for structural reliability analysis, Reliability Engineering and System Safety 198 (2020). doi:10.1016/j.ress.2020.106902.

[10] J. Li, J. Chen, The principle of preservation of probability and the generalized density evolution equation, Structural Safety 30 (1) (2008) 65–77. doi:10.1016/j.strusafe.2006.08.001.

[11] T. Zhou, Y. Peng, A two-stage point selection strategy for probability density evolution method-based reliability analysis, Structural and Multidisciplinary Optimization 65 (5) (2022). doi:10.1007/s00158-022-03244-7.

[12] J. Li, J. Chen, W. Fan, The equivalent extreme-value event and evaluation of the structural system reliability, Structural Safety 29 (2) (2007) 112–131. doi:10.1016/j.strusafe.2006.03.002.

[13] R. Teixeira, M. Nogal, A. O'Connor, Adaptive approaches in metamodel-based reliability analysis: A review, Structural Safety 89 (2021). doi:10.1016/j.strusafe.2020.102019.

[14] M. Moustapha, S. Marelli, B. Sudret, Active learning for structural reliability: Survey, general framework and benchmark, Structural Safety 96 (2022). doi:10.1016/j.strusafe.2021.102174.

[15] R. G. Ghanem, P. D. Spanos, Stochastic finite elements: a spectral approach, Springer-Verlag Inc, 1991.

[16] X. Zeng, R. Ghanem, Projection pursuit adaptation on polynomial chaos expansions, Computer Methods in Applied Mechanics and Engineering 405 (2023). doi:10.1016/j.cma.2022.115845.

[17] C. Lataniotis, S. Marelli, B. Sudret, The gaussian process modeling module in uqlab, Journal of Soft Computing in Civil Engineering 2 (3) (2018) 91–116. doi:10.22115/SCCE.2018.129323.1062.

[18] P. Pei, T. Zhou, One-step look-ahead policy for active learning reliability analysis, Reliability Engineering and System Safety 236 (2023). doi:10.1016/j.ress.2023.109312.

[19] T. Zhou, Y. Peng, An active-learning reliability method based on support vector regression and cross validation, Computers and Structures 276 (2023). doi:10.1016/j.compstruc.2022.106943.

[20] A. Roy, S. Chakraborty, Support vector machine in structural reliability analysis: A review, Reliability Engineering & System Safety 233 (2023). doi:10.1016/j.ress.2023.109126.

[21] T. Zhou, Y. Peng, J. Li, An efficient reliability method combining adaptive global metamodel and probability density evolution method, Mechanical Systems and Signal Processing 131 (2019) 592–616. doi:10.1016/j.ymssp.2019.06.009.

[22] T. Zhou, Y. Peng, Ensemble of metamodels-assisted probability density evolution method for structural reliability analysis, Reliability Engineering and System Safety 228 (2022). doi:10.1016/j.ress.2022.108778.

[23] B. Bichon, M. Eldred, L. Swiler, S. Mahadevan, J. McFarland, Efficient global reliability analysis for nonlinear implicit performance functions, AIAA Journal 46 (10) (2008) 2459–2468. doi:10.2514/1.34321.

[24] B. Echard, N. Gayton, M. Lemaire, Ak-mcs: An active learning reliability method combining kriging and monte carlo simulation, Structural Safety 33 (2) (2011) 145–154. doi:10.1016/j.strusafe.2011.01.002.

[25] M. Li, S. Shen, V. Barzegar, M. Sadoughi, C. Hu, S. Laflamme, Kriging-based reliability analysis considering predictive uncertainty reduction, Structural and Multidisciplinary Optimization 63 (6) (2021) 2721–2737. doi:10.1007/s00158-020-02831-w.

[26] T. Zhou, Y. Peng, Gaussian process regression based on deep neural network for reliability analysis in high dimensions, Structural and Multidisciplinary Optimization 66 (6) (2023). doi:10.1007/s00158-023-03582-0.

[27] T. Zhou, Y. Peng, A new active-learning function for adaptive polynomial-chaos kriging probability density evolution method, Applied Mathematical Modelling 106 (2022) 86–99. doi:10.1016/j.apm.2022.01.030.

[28] T. Zhou, Y. Peng, Reliability analysis using adaptive polynomial-chaos kriging and probability density evolution method, Reliability Engineering and System Safety 220 (2022). doi:10.1016/j.ress.2021.108283.

[29] C. Agrell, K. Dahl, Sequential bayesian optimal experimental design for structural reliability analysis, Statistics and Computing 31 (3) (2021). doi:10.1007/s11222-021-10000-2.

[30] J. Bect, D. Ginsbourger, L. Li, V. Picheny, E. Vazquez, Sequential design of computer experiments for the estimation of a probability of failure, Statistics and Computing 22 (3) (2012) 773–793. doi:10.1007/s11222-011-9241-4.

[31] C. Chevalier, D. Ginsbourger, J. Bect, E. Vazquez, V. Picheny, Y. Richet, Fast parallel kriging-based stepwise uncertainty reduction with application to the identification of an excursion set, Technometrics 56 (4) (2014) 455–465. doi:10.1080/00401706.2013.860918.

[32] T. Zhou, T. Guo, Y. Dong, Y. Peng, Polynomial chaos kriging-based structural reliability analysis via the expected margin volume reduction, Computers and Structures 287 (2023). doi:10.1016/j.compstruc.2023.107117.

[33] T. Zhou, T. Guo, Y. Dong, F. Yang, D. M. Frangopol, Look-ahead active learning reliability analysis based on stepwise margin reduction, Reliability Engineering and System Safety 243 (2024). doi:10.1016/j.ress.2023.109830.

[34] P. Wei, Y. Zheng, J. Fu, Y. Xu, W. Gao, An expected integrated error reduction function for accelerating bayesian active learning of failure probability, Reliability Engineering and System Safety 231 (2023). doi:10.1016/j.ress.2022.108971.

[35] T. Zhou, T. Guo, C. Dang, M. Beer, Bayesian reinforcement learning reliability analysis, Computer Methods in Applied Mechanics and Engineering 424 (2024). doi:10.1016/j.cma.2024.116902.

[36] T. Zhou, T. Guo, Y. Dong, Y. Peng, Structural reliability analysis based on probability density evolution method and stepwise truncated variance reduction, Probabilistic Engineering Mechanics 75 (2024). doi:10.1016/j.probengmech.2024.103580.

[37] R. T. Haftka, D. Villanueva, A. Chaudhuri, Parallel surrogate-assisted global optimization with expensive functions – a survey, Structural and Multidisciplinary Optimization 54 (1) (2016) 3 – 13. doi:10.1007/s00158-016-1432-3.

[38] Z. Wen, H. Pei, H. Liu, Z. Yue, A sequential kriging reliability analysis method with characteristics of adaptive sampling regions and parallelizability, Reliability Engineering and System Safety 153 (2016) 170 – 179. doi:10.1016/j.ress.2016.05.002.

[39] R. Schöbi, B. Sudret, S. Marelli, Rare event estimation using polynomial-chaos kriging, ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering 3 (2) (2017). doi:10.1061/AJRUA6.0000870.
[40] N. Lelièvre, P. Beaurepaire, C. Mattrand, N. Gayton, Ak-mcsi: A kriging-based method to deal with small failure probabilities and time-consuming models, Structural Safety 73 (2018) 1 – 11. doi:10.1016/j.strusafe.2018.01.002.
[41] J. Wang, Z. Cao, G. Xu, J. Yang, A. Kareem, An adaptive kriging method based on k-means clustering and sampling in n-ball for structural reliability analysis, Engineering Computations (Swansea, Wales) 40 (2) (2023) 378 – 410. doi:10.1108/EC-12-2021-0705.
[42] Z. Chen, G. Li, J. He, Z. Yang, J. Wang, A new parallel adaptive structural reliability analysis method based on importance sampling and k-medoids clustering, Reliability Engineering and System Safety 218 (2022). doi:10.1016/j.ress.2021.108124.
[43] G. Li, T. Wang, Z. Chen, J. He, X. Wang, X. Du, Rbik-ss: A parallel adaptive structural reliability analysis method for rare failure events, Reliability Engineering and System Safety 239 (2023). doi:10.1016/j.ress.2023.109513.
[44] Y. Xiong, S. Sampath, A fast-convergence algorithm for reliability analysis based on the ak-mcs, Reliability Engineering and System Safety 213 (2021). doi:10.1016/j.ress.2021.107693.
[45] F. A. Viana, R. T. Haftka, L. T. Watson, Sequential sampling for contour estimation with concurrent function evaluations, Structural and Multidisciplinary Optimization 45 (4) (2012) 615 – 618. doi:10.1007/s00158-011-0733-9.
[46] F. Yang, R. Kang, Q. Liu, C. Shen, R. Du, F. Zhang, A new active learning method for reliability analysis based on local optimization and adaptive parallelization strategy, Probabilistic Engineering Mechanics 75 (2024). doi:10.1016/j.probengmech.2023.103572.
[47] N.-C. Xiao, K. Yuan, H. Zhan, System reliability analysis based on dependent kriging predictions and parallel learning strategy, Reliability Engineering and System Safety 218 (2022). doi:10.1016/j.ress.2021.108083.
[48] Y. Meng, D. Zhang, B. Shi, D. Wang, F. Wang, An active learning kriging model with approximating parallel strategy for structural reliability analysis, Reliability Engineering and System Safety 247 (2024). doi:10.1016/j.ress.2024.110098.
[49] D. Zhan, J. Qian, Y. Cheng, Pseudo expected improvement criterion for parallel ego algorithm, Journal of Global Optimization 68 (3) (2017) 641 – 662. doi:10.1007/s10898-016-0484-7.
[50] Z. Zhao, Z.-H. Lu, Y.-G. Zhao, P-ak-mcs: Parallel ak-mcs method for structural reliability analysis, Probabilistic Engineering Mechanics 75 (2024). doi:10.1016/j.probengmech.2023.103573.
[51] C. Dang, M. Beer, Semi-bayesian active learning quadrature for estimating extremely low failure probabilities, Reliability Engineering and System Safety 246 (2024). doi:10.1016/j.ress.2024.110052.
[52] J. Chen, J. Yang, J. Li, A gf-discrepancy for point selection in stochastic seismic response analysis of structures with uncertain parameters, Structural Safety 59 (2016) 20–31. doi:10.1016/j.strusafe.2015.11.001.
[53] C. Chevalier, V. Picheny, D. Ginsbourger, Kriginv: An efficient and user-friendly implementation of batch-sequential inversion strategies based on kriging, Computational Statistics and Data Analysis 71 (2014) 1021 – 1034. doi:10.1016/j.csda.2013.03.008.
[54] V. J. Romero, J. V. Burkardt, M. D. Gunzburger, J. S. Peterson, Comparison of pure and "latinized" centroidal voronoi tessellation against various other statistical sampling methods, Reliability Engineering and System Safety 91 (10-11) (2006) 1266 – 1280. doi:10.1016/j.ress.2005.11.023.
[55] C. Dang, P. Wei, M. Faes, M. Valdebenito, M. Beer, Parallel adaptive bayesian quadrature for rare event estimation, Reliability Engineering and System Safety 225 (2022). doi:10.1016/j.ress.2022.108621.
[56] C. Dang, M. A. Valdebenito, M. G. Faes, P. Wei, M. Beer, Structural reliability analysis: A bayesian perspective, Structural Safety 99 (2022). doi:10.1016/j.strusafe.2022.102259.
[57] F. McKenna, M. Scott, G. Fenves, Nonlinear finite-element analysis software architecture using object composition, Journal of Computing in Civil Engineering 24 (1) (2010) 95–107. doi:10.1061/(ASCE)CP.1943-5487.0000002.
[58] C. Chevalier, D. Ginsbourger, X. Emery, Corrected kriging update formulae for batch-sequential data assimilation, Lecture Notes in Earth System Sciences 0 (2014) 119–122. doi:10.1007/978-3-642-32408-6_29.