Efficient subset simulation with active learning Kriging model for low failure probability prediction

Zhang, Xiaodong; Quek, Ser Tong

Published in:
Probabilistic Engineering Mechanics

Link to article, DOI:
10.1016/j.probengmech.2022.103256

Publication date:
2022

Document Version
Peer reviewed version

Journal Pre-proof

Efficient subset simulation with active learning Kriging model for low failure probability prediction

Xiaodong Zhang, Ser Tong Quek

PII: S0266-8920(22)00035-2
DOI: https://doi.org/10.1016/j.probengmech.2022.103256
Reference: PREM 103256

To appear in: *Probabilistic Engineering Mechanics*

Received date : 4 March 2022
Accepted date : 13 March 2022

Please cite this article as: X. Zhang and S.T. Quek, Efficient subset simulation with active learning Kriging model for low failure probability prediction, *Probabilistic Engineering Mechanics* (2022), doi: https://doi.org/10.1016/j.probengmech.2022.103256.

Efficient subset simulation with active learning Kriging model for low failure probability prediction

Xiaodong ZHANG [a,b] and Ser Tong QUEK [a,*]

[a] Department of Civil and Environmental Engineering, National University of Singapore, 1 Engineering Drive 2, Singapore 117576, Singapore.

[b] Department of Wind Energy, Technical University of Denmark, Frederiksborgvej 399, 4000, Roskilde, Denmark

* Corresponding author: ceeqst@nus.edu.sg, Tel: +65 6516 2263, Fax: +65 6779 1635

Abstract

For low failure probability prediction, subset simulation can reduce the number of simulations significantly compared to the traditional MCS method for a target prediction error limit. To further reduce the computational effort for cases where the performance function evaluation is tedious and time-consuming, the performance function is approximated by a sequentially updated (instead of a global) Kriging model. For this purpose, an active learning technique with a new learning and stopping criterion is employed to efficiently select points to train the computationally cheaper Kriging model at each simulation level, which is used to estimate the intermediate threshold and generate a new simulation sample. The updated Kriging model at the final subset simulation level is used to compute the conditional failure probability. The failure probability is estimated based on an initial simulation sample size N, and an updated N is computed and employed to obtain the final failure probability within a desired bound on the variability. The efficiency (in terms of the number of expensive evaluations using the actual performance function) and prediction error (represented by the mean square error (MSE)) of the proposed method are benchmarked using several examples. The method is shown to be more efficient (using lesser expensive evaluations) with smaller MSE for problems having low failure probabilities compared with selected existing methods.

Keywords: active learning, Kriging model, subset simulation, low failure probability

Highlights

- Active learning technique with new learning and stopping criteria is employed to facilitate the use of small number of expensive evaluations for updating the Kriging model at each simulation level. The computationally expensive training points are re-used at subsequent levels for better efficiency.
- Kriging model is used to determine the intermediate threshold for the generation of a new simulation sample as well as the estimation of the conditional failure probability at the final subset simulation level with a low computational cost.
- Simulation sample size is estimated based on a desired limit in the variability of the failure probability.
- Method yields better results than those from the method of Ling et al. [14] in terms of efficiency and prediction error.

1. Introduction

In reliability analysis, the failure probability P_F is calculated by the integral

$$P_F = P\{\boldsymbol{\theta} \in F\} = \int_{\boldsymbol{\theta} \in F} f(\boldsymbol{\theta}) d\boldsymbol{\theta} \tag{1}$$

in which $\boldsymbol{\theta}$ is an n_d-dimensional random vector with a joint probability density function (pdf) f; F is the failure domain defined by $G(\boldsymbol{\theta}) \le 0$, where $G(.)$ is the performance function. The calculation of the integral in Eq. (1) is tedious when the integration domain F is complex (highly nonlinear and/or discontinuous) and/or implicit (represented through a numerical model as such the finite element model).

The direct Monte Carlo Simulation (MCS) method evaluates the integral via random sampling to yield an unbiased estimate of P_F given by

$$\hat{P}_F = \frac{1}{N} \sum_{n=1}^{N} I(\boldsymbol{\theta}_n) \tag{2}$$

where $\boldsymbol{\theta}_n$ denotes a random vector sampled from f, N is the sample size, and $I(.)$ is an indicator function taking on a value of 0 or 1 depending on whether $\boldsymbol{\theta}_n$ falls in the safe or failure domain, respectively. Based on the strong law of large numbers, $\hat{P}_F$ converges almost surely to P_F, and an estimate of the coefficient of variation (c.o.v.) of $\hat{P}_F$, δ_{MCS}, is given by [1]

$$\delta_{MCS} = \sqrt{\frac{1 - P_F}{P_F N}} \tag{3}$$

Generally, δ_{MCS} does not depend on the dimension of $\boldsymbol{\theta}$ or the complexity of $G(.)$, making the MCS approach simple, versatile, and attractive in solving many structural reliability analysis problems. However, for low failure probability (e.g., $P_F < 10^{-4}$) prediction, N needs to be large (e.g., $N > 10^6$) to ensure that δ_{MCS} is sufficiently small.

Many researchers thus focused on methods that can achieve the same accuracy with much smaller N, such as importance sampling [2], univariate dimension reduction [3], directional simulation [4], and subset simulation (SS) [5]. SS is a popular and efficient method that computes low failure

probability as the product of a sequence of high conditional failure probabilities, where the latter can be estimated using a small sample.

Despite the small N used in these methods, the cost to compute the performance function for each sample can be high if $G(.)$ is complex (such as being highly nonlinear, comprising a set of functions, multiple constraints, and/or difficult to be explicitly expressed). This leads to the use of computationally cheaper surrogate models to approximate $G(.)$, which can be built using methods such as support vector machine (SVM), Kriging, and neural network. In some of these methods (other than SS), direct MCS is used to evaluate P_F, albeit using an approximated and 'easy to compute' limit state function (LSF), but the curse of low P_F remains.

Using a random sample to train a surrogate model to approximate the LSF or performance function directly for low failure probability prediction is inefficient and may not produce a good model. This is due to a much higher number of points falling in the safe domain compared to the number in the failure domain such that the model may not be optimal for extrapolating into the failure domain. As such, active learning techniques have recently been adopted to efficiently select the training points. Pedroni and Zio [6] employed SS and an artificial neural network to estimate the importance sampling density for computing the failure probability. Pan and Dias [7] proposed using a SVM model enriched by an active learning method to approximate $G(.)$ before using MCS to estimate the failure probability. Similarly, Echard et al. [8] proposed using the Kriging model enriched by an active learning method to replace $G(.)$ before employing MCS (denoted as AK-MCS method). Subsequently, Echard et al. [9] coupled importance sampling with the active learning Kriging model to further improve its efficiency. Lelièvre et al. [10] addressed the drawbacks of AK-MCS, namely the inaccuracy associated with small failure probability and non-parallelized computation, by using sequential MCS and a clustering technique.

As SS is more efficient than the crude MCS, a natural improvement to the AK-MCS method is the AK-SS method proposed by Huang et al. [11], which used the Kriging model enriched by an active learning method to approximate $G(.)$ before applying SS. AK-MCS and AK-SS both

approximate $G(.)$ using a single updated Kriging model over the entire domain with active learning technique, which may not always be suitable for low failure probability prediction. This is because the active learning technique may not necessarily add points close to the limit state function, resulting in the latter not being well approximated at where it matters most. Even though increasing the simulation sample size could possibly ensure enough simulation points fall in the failure domain, this will substantially increase the computational cost.

Most recent efforts concentrated on improving the accuracy further in these SS coupled with machine learning reliability computational techniques by using a surrogate model at each simulation level. Papadopoulos et al. [12] used a neural network, which was trained at subdomains defined by the bounds of the data, to obtain the surrogate model at each simulation level. Cui and Ghosn [13] applied Kriging to approximate the performance function at each simulation level, and the local Kriging models are trained based on a clustering algorithm. However, the choice of the number of clusters remains to be addressed for practical applications. Ling et al. [14] coupled SS with an active learning Kriging model for failure probability estimation, where a new Kriging model is used at each simulation level. The SS used is different from that proposed by Au and Beck [5], and the active learning framework adopted can be improved in terms of efficiency and prediction error. To improve the method of SS with a surrogate model for low failure probability prediction, the intermediate threshold at each simulation level should be more accurately calculated, the simulation sample should be appropriately generated, and the limit state function should be well approximated to reduce the prediction error. The number of evaluations using the actual (not the surrogate) performance function for training the surrogate model, which is costly, should be as small as possible to increase efficiency.

The objective of this paper is to develop an improved algorithm for SS in conjunction with an active learning Kriging model to predict low failure probability using a small number of costly performance function evaluations. An active learning framework with new learning and stopping criteria is used to efficiently update the Kriging model, which is used to determine the intermediate threshold and generate a simulation sample for the next simulation level. The intermediate threshold is calculated with a predefined conditional failure probability (e.g., 0.1). The conditional failure

probability at the last simulation level is computed using the final updated Kriging model. The training set at each simulation level is updated and passed to the next simulation level, to approximate the performance function over a wider domain. The simulation sample size is estimated based on maintaining a specified limit on the prediction variability. The performance of the proposed procedure is investigated statistically using a variety of examples, including the dynamic response of an offshore drilling riser system, and compared with results from some existing methods.

2. Background methodologies

2.1 Subset simulation (SS)

The key ideas of SS [5] are summarized herein. Given a failure event F, and a sequence of other "failure" events, $F_1, F_2, \ldots, F_m$, such that $F_1 \supset F_2 \supset \ldots \supset F_m = F$, the failure probability is given by

$$P_F = P(F_m) = P(F_1)\prod_{i=1}^{m-1} P(F_{i+1} \mid F_i) \tag{4}$$

where $P(F_1)$ is estimated by MCS as in Eq. (2) and denoted by $\hat{P}_1$ with an initial simulation sample $\left\{\boldsymbol{\theta}_k^{(0)}: k = 1, 2, \ldots, N\right\}$. The conditional failure probability is calculated as:

$$P(F_{i+1} \mid F_i) \approx \hat{P}_{i+1} = \frac{1}{N}\sum_{k=1}^{N} I_{F_{i+1}}\left(\boldsymbol{\theta}_k^{(i)}\right) \tag{5}$$

where $\left\{\boldsymbol{\theta}_k^{(i)}: k = 1, 2, \ldots, N\right\}$ is the Markov chain sample at the ith conditional level, $1 \le i \le m-1$. For ease of implementation, only $\hat{P}_m$ is evaluated by Eq. (5) with the target threshold, while all other intermediate probabilities are fixed at 0.1 for simplicity (that is, $\hat{P}_i = 0.1$ for $i = 1, 2, \ldots, m\text{-}1$), from which the corresponding intermediate thresholds y_i are determined. The simulation sample $\boldsymbol{\Theta}^{(i)} = \{\boldsymbol{\theta}_k^{(i)}: k = 1, 2, \ldots, N; i = 1, 2, \ldots, m\text{-}1\}$ is generated using the modified Metropolis algorithm [15], with each of the N_c Markov chains generating N_s ($= N/N_c$) points. For the case of $P_i = 0.1$, $N_c = 0.1*N$ and $N_s = 10$. The advantage of SS over direct MCS lies in the controlled generation of simulation samples in the intermediate failure domain, thus increasing its efficiency associated with high failure probability. The failure probability is estimated as

$$\hat{P}_F = \prod_{i=1}^{m} \hat{P}_i \tag{6}$$

The theoretical lower bound of the c.o.v. of $\hat{P}_F$, when the conditional probabilities at different levels are uncorrelated, is given by

$$\delta_{ss} = \sqrt{\sum_{i=1}^{m} \delta_i^2} \tag{7}$$

where δ_1 is the c.o.v. of $\hat{P}_1$, given theoretically by

$$\delta_1 = \sqrt{\frac{1-P_1}{P_1 N}} \tag{8}$$

and δ_i, $2 \le i \le m$, is the c.o.v. of $\hat{P}_i$, given theoretically by

$$\delta_i = \sqrt{\frac{1-P_i}{P_i N}(1+\gamma_i)} \tag{9}$$

in which γ_i is the correlation factor [5].

2.2 Active learning with Kriging model (AK)

Given the role of the Kriging model in this paper, the key equations are summarized herein and more details can be found in [8,11,16]. Kriging is a Gaussian process regression method which generates a spatial interpolation function based on a covariance or variogram model derived from the data. Generally used for a prediction purpose, the model comprises a regression expression and a deviation term described by a random process. Treating the dependent variable as a stochastic process $Y(\mathbf{x})$, it can be mathematically expressed as

$$Y(\mathbf{x}) = \sum_{k=1}^{p} \beta_k f_k(\mathbf{x}) + z(\mathbf{x}) \tag{10}$$

where the first term represents the trend, and $f_k(\mathbf{x})$ and β_k with $k = 1, 2, \ldots, p$ are the basis functions and regression coefficients respectively. For convenience, denote $\mathbf{F}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_p(\mathbf{x})]$ and $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_p]$. The random process $z(\mathbf{x})$ is assumed to have mean zero, and the covariance between $z(\mathbf{x}(i))$ and $z(\mathbf{x}(j))$ is given by

$$V\left(\mathbf{x}^{(i)},\mathbf{x}^{(j)}\right)=\sigma_z{}^2R\left(\tilde{\boldsymbol{\theta}},\mathbf{x}^{(i)},\mathbf{x}^{(j)}\right) \tag{11}$$

where σ_z^2 is the process variance and $R_{ij} = R(\boldsymbol{\theta}, \mathbf{x}^{(i)}, \mathbf{x}^{(j)})$ is the correlation computed from the correlation model $\mathbf{R}$ with parameter $\tilde{\boldsymbol{\theta}}$. The deterministic response, denoted by $y(\mathbf{x})$, is treated as a realization of $Y(\mathbf{x})$.

For a design of experiments $[\mathbf{x}^{(1)},\mathbf{x}^{(2)},\ldots,\mathbf{x}^{(n)}]$, $\mathbf{Y}=\left[Y\left(\mathbf{x}^{(1)}\right),Y\left(\mathbf{x}^{(2)}\right),\ldots,Y\left(\mathbf{x}^{(\mathrm{n})}\right)\right]^{\mathrm{T}}$, the Kriging estimator at an unknown point $\mathbf{x}^{(*)}$ is given by

$$\hat{y}\left(\mathbf{x}^{(*)}\right)=\mathbf{F}\left(\mathbf{x}^{(*)}\right)^{\mathrm{T}}\hat{\boldsymbol{\beta}}+\mathbf{r}\left(\mathbf{x}^{(*)}\right)^{\mathrm{T}}\hat{\boldsymbol{\gamma}} \tag{12}$$

where $\mathbf{r}(\mathbf{x}^{(*)})$ is a vector with components $r_i=R\left(\boldsymbol{\theta},\mathbf{x}^{(*)},\mathbf{x}^{(i)}\right)$ and $\hat{\boldsymbol{\gamma}}$ is computed as $\hat{\boldsymbol{\gamma}}=\mathbf{R}^{-1}\left(\mathbf{Y}-\mathbf{F}\hat{\boldsymbol{\beta}}\right)$. The *MSE* (which is the same as the Kriging variance since the Kriging estimator is theoretically unbiased) of the predictor is

$$\sigma^2=\sigma_z{}^2\left(1+\mathbf{u}^{\mathrm{T}}\left(\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{F}\right)^{-1}\mathbf{u}-\mathbf{r}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{r}\right) \tag{13}$$

where $\mathbf{u}=\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{r}-\mathbf{F}(\mathbf{x}^{(*)})$. Computations can be performed using the Kriging Toolbox DACE [17,18] in MATLAB.

One challenge in using a surrogate model to approximate the performance function is in employing a small number of training points to yield a model with a small prediction error. Techniques such as Latin hypercube sampling (LHS) as well as other design of experiment methods may be used instead of conventional random sampling. However, the sample size of LHS is difficult to pre-determine to obtain a surrogate model within the desired prediction error. To improve the surrogate model trained using a sample from LHS, additional sampling points may be employed but the challenge is how to select such points to achieve maximum impact. Recently, the active learning technique has been proposed which incorporates information from previously used points to sequentially locate where a new training point is to be added to increase the accuracy of the model [19]. Starting with a few initial training points, a query strategy or learning criterion locates where the trained model has the maximum uncertainty to add a new training point. The process is repeated until a stopping criterion is met, where different criteria have been proposed.

AK-MCS [8] iteratively adds training points close to the LSF to train the Kriging model so that it can approximate the LSF for classifying sample points that are in the safe or failure domain. Two learning functions have been proposed, namely, the expected feasibility function (*EFF*), which originated from the EGRA method [20], and the *U* function.

EFF is computed through an optimization algorithm which balances between the global search over the whole domain and the local search close to the limit state function [8]. The expression for *EFF* is given by

$$\begin{aligned} EFF(\boldsymbol{\theta}) = &\left(\hat{G}(\boldsymbol{\theta})-\alpha\right)\left[2\Phi\left(\frac{\alpha-\hat{G}(\boldsymbol{\theta})}{\sigma_{\hat{G}}(\boldsymbol{\theta})}\right)-\Phi\left(\frac{(\alpha-\varepsilon)-\hat{G}(\boldsymbol{\theta})}{\sigma_{\hat{G}}(\boldsymbol{\theta})}\right)-\Phi\left(\frac{(\alpha+\varepsilon)-\hat{G}(\boldsymbol{\theta})}{\sigma_{\hat{G}}(\boldsymbol{\theta})}\right)\right] \\ &-\sigma_{\hat{G}}(\boldsymbol{\theta})\left[2\phi\left(\frac{\alpha-\hat{G}(\boldsymbol{\theta})}{\sigma_{\hat{G}}(\boldsymbol{\theta})}\right)-\phi\left(\frac{(\alpha-\varepsilon)-\hat{G}(\boldsymbol{\theta})}{\sigma_{\hat{G}}(\boldsymbol{\theta})}\right)-\phi\left(\frac{(\alpha+\varepsilon)-\hat{G}(\boldsymbol{\theta})}{\sigma_{\hat{G}}(\boldsymbol{\theta})}\right)\right] \\ &-\varepsilon\left[\Phi\left(\frac{(\alpha+\varepsilon)-\hat{G}(\boldsymbol{\theta})}{\sigma_{\hat{G}}(\boldsymbol{\theta})}\right)-\Phi\left(\frac{(\alpha-\varepsilon)-\hat{G}(\boldsymbol{\theta})}{\sigma_{\hat{G}}(\boldsymbol{\theta})}\right)\right] \end{aligned} \tag{14}$$

where Φ is the standard normal cumulative distribution function, ϕ is the standard normal pdf, $\hat{G}(\boldsymbol{\theta})$ is the predicted response at location $\boldsymbol{\theta}$ and $\sigma_{\hat{G}}(\boldsymbol{\theta})$ is the predicted standard deviation. The value of *EFF* indicates how well the actual performance function value at location $\boldsymbol{\theta}$ is expected to satisfy $\hat{G}(\boldsymbol{\theta}) = \alpha$ over the region $\alpha \pm \varepsilon$. A large *EFF* value implies a high level of uncertainty in the predicted performance function value [20]. Hence, $\boldsymbol{\theta}$ corresponding to the maximum *EFF* is selected as the point to be added to the training set, until the stopping criterion of $\max(EFF(\boldsymbol{\Theta})) \leq 0.001$ is met (see Table 1). The empirical value of 0.001 was recommended by Echard et al. [8].

The other learning function *U* has been defined as [8]:

$$U(\boldsymbol{\theta}) = \frac{\left|\hat{G}(\boldsymbol{\theta})\right|}{\sigma_{\hat{G}}(\boldsymbol{\theta})}. \tag{15}$$

It measures the number of standard deviations that the predicted $\hat{G}(\boldsymbol{\theta})$ is from 0. Using a stopping criterion of $\min(U(\boldsymbol{\Theta})) \geq 2$ (see Table 1) as recommended by Echard et al. [8] ensures that only

additional points that have absolute performance values smaller than twice the standard deviation are appended into the set of training data as they have higher contributions to the failure probability. The threshold value of 2 is associated with points having an outlier probability of at most 1-$\Phi(2)$ = 0.023 on either side of the LSF.

Table 1 Definition of learning function and stopping criterion for AK-MCS

	EFF	*U*
Learning function	max($EFF(\boldsymbol{\Theta})$)	min($U(\boldsymbol{\Theta})$)
Stopping criterion	max($EFF(\boldsymbol{\Theta})$) $\leqslant$ 0.001	min($U(\boldsymbol{\Theta})$) $\geqslant$ 2

To train a Kriging model with a large simulation sample size is computationally challenging. The AK-MCS [8] and AK-SS [11] reduce the computational effort in evaluating the performance function values but do not overcome the limitation of requiring a large simulation sample size associated with a low failure probability, and the reasons are given as follows:

Firstly, the AK-MCS [8] is similar with MCS but requires much less performance function evaluations. For a problem with a target failure probability P_F, its simulation sample size N should be larger than $100/P_F$ to have a prediction c.o.v. of $\hat{P}_F$ smaller than 0.1 (see Eq. (3)). Secondly, the Kriging model is a regression model, which has limitations for extrapolation, and enough training points should come from the failure domain to approximate the limit state function over the entire domain of interest. A large simulation sample size is required to have enough points in the failure domain, as the sample are generated using the MCS method. The problem is compounded by the fact that the failure probability is not known a priori. Starting with a small N, the simulation sample size to use is increased iteratively with random sampling, which is also computationally costly.

Even training a Kriging model with a large N is possible, the active learning procedures provided in AK-MCS [8] do not guarantee a new training point is selected, which could result a failure of the algorithm.

Taking advantage of the SS method in dealing with small failure probabilities, the AK-SS method [11] evaluates failure probability with SS using the trained Kriging model from AK-MCS. Since AK-SS uses the trained Kriging model from AK-MCS, the above-mentioned issues remain unaddressed.

It should be pointed out that the target failure probabilities of most of the examples in AK-MCS [8] and AK-SS [11] are larger than 10^{-4}, and the only example with $P_F < 10^{-4}$ in AK-SS [11] is estimated with different initial training sample size and stopping criterion used with other high failure probability examples.

2.3 Improved active learning Kriging model with subset simulation (I-AK-SS)

Ling et al. [14] proposed further improvement by approximating the performance function at each simulation level with a corresponding surrogate model instead of a global model. However, unlike SS [5], which uses a fixed conditional failure probability to calculate the intermediate threshold, Ling et al. [14] uses an estimated intermediate threshold to calculate conditional failure probability at each simulation level. The first intermediate threshold y_1 is determined from 1 out of only 10 training points used. At simulation level i, the Kriging model K_i is updated with points selected based on the active learning function

$$U(\boldsymbol{\theta}) = \frac{\left|\hat{G}(\boldsymbol{\theta}) - y_i\right|}{\sigma_{\hat{G}}(\boldsymbol{\theta})} \quad . \tag{16}$$

The model is used to compute the conditional failure probability associated with this threshold, whereas most researchers fixed the conditional failure probability at 0.1 to find the threshold. MCMC algorithm is applied to generate a new sample, which is evaluated using K_i to obtain y_{i+1}. The process is repeated until y_i is in the failure domain.

There are three improvements which can be made. First, in Ling et al. [14], the intermediate threshold is determined before updating the Kriging model. The first intermediate threshold is calculated using only 10 points and thus may compromise the accuracy of the method. Second, at each simulation level, the Kriging model is updated using Eq. (16) for the calculation of conditional

failure probability with extra training points, which leads to more evaluations using the computationally expensive actual performance functions. Third, the simulation sample size is a key factor for balancing the accuracy and efficiency of the method, but no proper way of determining the sample size is given. All three aspects are addressed in this paper.

3. Proposed method

The proposed method in this paper approximates the performance function at each simulation level using a Kriging model, updated with a new active learning technique. At simulation level i (i = 1, 2, ..., m), the Kriging model M_i is updated with training points selected from the simulation sample $\boldsymbol{\Theta}^{(i-1)}$ using the proposed active learning technique. To determine the intermediate threshold based on a fixed conditional failure probability of 0.1, M_i is used to compute the performance function values. MCMS is employed to generate the simulation sample $\boldsymbol{\Theta}^{(i)}$ and the process is repeated until y_i is in the failure domain. The conditional failure probability at the final level $\hat{P}_m$ is evaluated using M_m associated with the target threshold. The failure probability $\hat{P}_F$ and its associated c.o.v. (Eqs. (6) and (7)) are estimated. The whole procedure is repeated with an increased simulation sample size N (estimated using Eq. (18) given below) until the desired prediction variability (c.o.v. ≤ 0.05) is achieved. The proposed method is detailed below.

1) Generate the simulation sample and form the training set.

   Generate the simulation sample, $\boldsymbol{\Theta}^{(0)} = \{\boldsymbol{\theta}_k^{(0)}: k = 1, 2, \ldots, N\}$, following f. Next, select n_0 training points, $\mathbf{X} = \{\mathbf{x}_k: k = 1, 2, \ldots, n_0\}$, by LHS and evaluate the associated performance function values $\mathbf{Y} = G(\mathbf{X})$ to form the training set S $=\{\mathbf{X}, \mathbf{Y}\}$. The number of evaluations N_{call} using the ***actual*** performance function is n_0. As active learning technique will be used later to add training points, a small value of n_0, with a good spread of selected points based on LHS, is used.

2) Train the Kriging model at current simulation level i with active learning technique.

   Train the Kriging model M_i with training set S. The DACE toolbox [17,18] is used for this purpose. On the choice of new training points (which are selected from the simulation sample

$\boldsymbol{\Theta}^{(i-1)}$) to include for improving the Kriging model, one criterion is to consider a point which contributes most to the uncertainty in the model.

Each point k of the simulation sample $\boldsymbol{\Theta}^{(i-1)} = \{\boldsymbol{\theta}_k^{(i-1)}: k = 1, 2, \ldots, N\}$ is a potential training point. As such for each point, evaluate the performance value $\widehat{G}(\boldsymbol{\theta}_k)$ using the trained Kriging model (see Eq. (12)) and the associated mean square error (*MSE*) $\sigma_{\widehat{G}}^2(\boldsymbol{\theta}_k)$ (see Eq. (13)). Given that the value of $\widehat{G}(\boldsymbol{\theta}_k)$ for each k is different, a normalized measure $C(\boldsymbol{\theta}_k)$ of its variability is adopted, given by

$$C(\boldsymbol{\theta}) = \frac{\sqrt{MSE}}{\left|\widehat{G}(\boldsymbol{\theta})\right|} = \frac{\sigma_{\widehat{G}}(\boldsymbol{\theta})}{\left|\widehat{G}(\boldsymbol{\theta})\right|} \tag{17}$$

The maximum $C(\boldsymbol{\theta}_k)$ over the N candidate points is used to determine whether an additional training point needs to be selected. A threshold of 0.1 is used in this paper. No new point will be added to the training set if (a) $\max\left(C\left(\boldsymbol{\Theta}^{(i-1)}\right)\right) \leq 0.1$, or (b) majority of points in the failure domain remains unchanged (99% points in the failure domain remain the same after 10 new training points are added). If neither of the criteria is satisfied, the point with the highest C, say $\{\mathbf{x}_c, \mathbf{y}_c\}$, where $\mathbf{y}_c = G(\mathbf{x}_c)$, is added to the current training set. Since $\mathbf{y}_c$ is computed using the actual performance function, update $N_{call} = N_{call}+1$. This will yield a re-trained Kriging model M_i and the process is repeated until one of the criteria is met. Note that C at the last level is identical to the reciprocal of the U function proposed in [8], where a threshold value of 2 (giving a reciprocal value of 0.5) was proposed.

3) Calculate the intermediate threshold.

   If the intermediate threshold y_i corresponding to the conditional failure probability p_0 is to be determined, the procedure is to rank $\hat{G}\left(\boldsymbol{\Theta}^{(i-1)}\right)$ in ascending order and use the average of the (Np_0)th and (Np_0+1)th value as y_i. A conditional failure probability with a fixed value $p_0 = 0.1$, which is not recalculated based on the intermediate threshold associated with $p_0 = 0.1$, is adopted.

4) Generate the simulation sample and select the training set for the next simulation level.

If y_i is positive (i.e., the associated random vector is within the safe domain), use the modified Metropolis algorithm to generate a new simulation sample $\boldsymbol{\Theta}^{(i)} = \{\boldsymbol{\theta}_k^{(i)}: k = 1, 2, \ldots, N\}$. Note that the generated points are evaluated using the Kriging model M_i (instead of the actual performance function) for acceptance and rejection, based on whether they are within the 'failure' domain at the current simulation level. Each of the N_c ($=Np_0$) Markov chains is used to generate N_s ($= 1/p_0$) points to get a total of N points. All the training points selected from the first to the current simulation level (where the $G(\mathbf{x})$ values have already been computed using the actual performance function) are kept as the training points for the next simulation level. Repeat steps 2-4 until y_i is smaller than 0 (associated with the failure domain of the original problem).

5) Calculate the failure probability and its corresponding c.o.v..

The conditional failure probability corresponding to the target threshold of 0 is calculated using Eq. (5) via the updated Kriging model. The failure probability and the corresponding c.o.v. δ_{SS} are estimated based on Eqs. (6) and (7) respectively. If δ_{SS} greater than the target c.o.v. $\tilde{\delta}_{SS}$ (set as 0.05 in this paper), go to step 2 and repeat each step thereafter but with a larger sample size N. Assuming that $\hat{P}_i$ and γ_i do not change much as N increases, the new sample size is estimated as

$$N \geq \frac{1}{\tilde{\delta}_{SS}^2}\left(\frac{1-\hat{P}_1}{\hat{P}_1} + \sum_{i=2}^{m}\frac{1-\hat{P}_i}{\hat{P}_i}\left(1+\gamma_i\right)\right), \tag{18}$$

and N is rounded up to the next value of 1000 to ensure that N_c is an integer. For computational efficiency, all the existing training points are kept and used where appropriate.

6) End of simulation.

The computation ends when δ_{SS} is not more than the target c.o.v. $\tilde{\delta}_{SS}$.

The flow chart of the proposed method is summarized in Fig. 1.

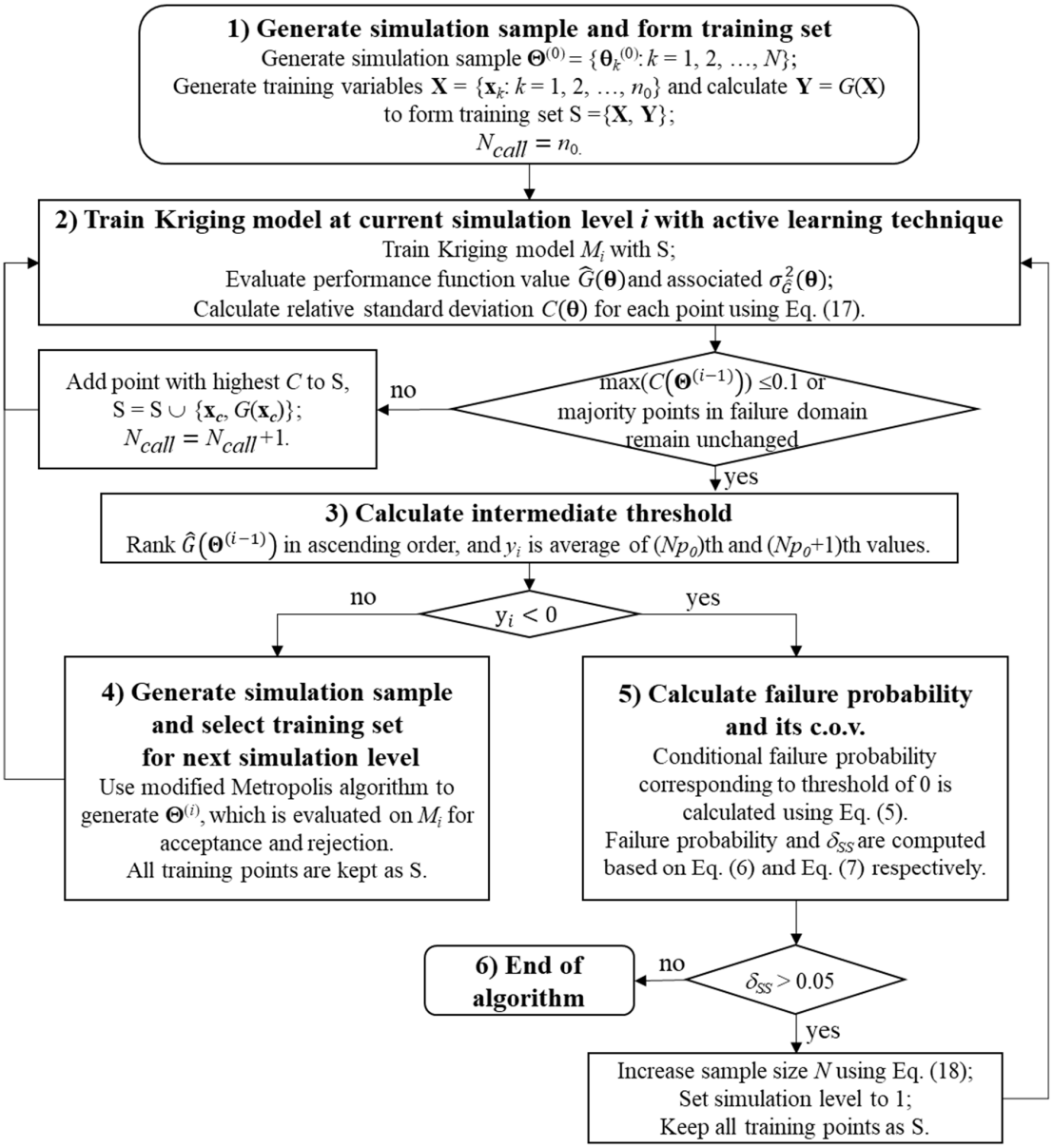

Fig. 1 Flow chart of proposed method

4. Key points to note

4.1 The general framework

The proposed method adopts the procedure commonly used in most SS algorithms, where the conditional failure probability is set to 0.1 and the associated intermediate threshold is determined. The Kriging model used to determine the threshold at each simulation level is based on an updated

model trained with points selected by the proposed active learning technique. In the proposed method, a large sample size N is used in conjunction with the updated Kriging model to give a good estimate of the intermediate threshold (including the first threshold) while the probability is fixed at 0.1. The learning criterion in the proposed method selects the point with the largest uncertainty, which is quantified by the relative standard deviation. The active learning algorithm stops when all points have small prediction variability or the majority of the points in the failure domain remain unchanged. All training points are kept and re-used at subsequent levels. This gives more points in the training with little additional computational cost to yield a Kriging model that is fitted with data covering a wide domain. This results in a better-quality sample generated by MCMC for the final simulation level to determine the conditional failure probability. The simulation sample size is updated to achieve a theoretical prediction variability limit.

There are differences between the proposed method and that in Ling et al. [14]. In the method of Ling et al. [14], the first intermediate threshold is calculated based on 10 training points; the threshold at each level (except the first) is determined using the Kriging model from the previous level with a fixed conditional failure probability 0.1, and the conditional failure probability is recomputed using the respective improved Kriging model with the threshold; the learning function and the stopping criterion are fundamentally different from the proposed method; and the simulation sample size is not updated.

4.2 Learning function and stopping criteria

The Kriging model is used to calculate the intermediate threshold with limited prediction error and classify the simulation sample into safe and failure domain at each simulation level.

The learning function C given in Eq. (17) selects the point with maximum uncertainty to be added to the training set, which yields a desired Kriging model. The training set at the current simulation level is passed to subsequent levels. The points added by the learning algorithm make improvements for the Kriging model over the domain of the current simulation sample.

Two stopping criteria are provided for the proposed method, and the active learning stops when either one of the criteria is satisfied. The first stopping criterion $\max\left(C\left(\boldsymbol{\Theta}^{(i-1)}\right)\right) \leq 0.1$ limits the Kriging model prediction error for the simulation sample, which reduces the error of intermediate threshold estimation. The threshold 0.1 yields a small prediction error of the proposed method. If a value of 0.2 is chosen, the learning algorithm will stop with fewer additional points whereas if the value is 0.05, more points will be added. The smaller threshold 0.1 will not select unnecessary points with the help of the second stopping criterion. The second stopping criterion requires that 99% of the points in the failure domain remain the same after 10 new training points are added, which means the active learning stops when it does not improve the Kriging model in terms of classifying the simulation sample into safe and failure domains. The choice of parameters in the stopping criteria, i.e., 0.1, 99% and 10, is a matter of balance between prediction error and computational cost and might not be optimal for all the applications.

4.3 Simulation sample size

The choice of the initial training set size at the first simulation level is intended to be small ($n_0 = n_d+12$) as (a) the actual performance will be used to compute the values at these points which are expensive, (b) these few points are selected using LHS thus ensuring an appropriate broad representation over the domain, (c) additional points will be included using an active learning algorithm to ensure that they are optimally selected to improve the model, and (d) n_0 should be at least larger than the dimension of the problem.

To minimize computational cost, the Kriging model is used to compute performance function values for estimating (a) the threshold at each simulation level corresponding to a fixed probability value of 0.1, and (b) the conditional failure probability at the final simulation level. In terms of computing the final conditional failure probability with an order of magnitude of 0.1, a sample with size $N = 10^4$ would produce a good initial estimate. High N will produce a more accurate threshold for the generation of the simulation sample, but with more training points selected. Given the much lower cost in the determination of the threshold and conditional failure probability compared to the

much higher overall computational cost (mostly in using the actual performance function to evaluate N_{call} training points for the Kriging models), $N = 10^4$ is chosen for less prediction error with little additional cost compared $N = 10^3$.

From the initial estimate of the failure probability with $N = 10^4$, Eq. (18) can be used to determine an updated simulation sample size N to yield a theoretically acceptable c.o.v. of the estimator $\hat{P}_F$. The choice of the c.o.v. is again a balance of prediction error and computational cost. In this study, a small value of 0.05 is chosen.

5. Examples

5.1 General considerations

Numerical examples are shown in this section to compare the results with those obtained using other existing SS-based structural reliability methods. The examples cover five numerical problems and a real-world application, with low or high dimensional variables, linear or nonlinear performance functions, and continuous or discontinuous limit state functions.

To compare the performance amongst the SS, the method in Ling et al. [14], and the proposed method, the number of computations using the actual (not surrogate) performance function (which is expensive), N_{call}, and the mean square error (*MSE*) are examined in the examples. The *MSE* of the estimator $\hat{P}_F$, given by $MSE(\hat{P}_F) = E\left[(\hat{P}_F - P_F)^2\right]$, is a function of its variance, $V(\hat{P}_F)$, and bias, $B(\hat{P}_F) = E(\hat{P}_F) - P_F$; that is, $MSE(\hat{P}_F) = V(\hat{P}_F) + \left[B(\hat{P}_F)\right]^2$ [21]. For the unbiased case, the *MSE* and predictor variance are the same.

To estimate these statistics for each example, each method is repeatedly simulated 100 times. The mean of the computed failure probabilities from 100 runs using the direct MCS method is taken as the reference failure probability P_F to compare with. The mean estimated failure probabilities $E(\hat{P}_F)$ from each method, the prediction variance V, the square of the bias B^2 and the mean number of evaluations using the actual performance function $E(N_{call})$ for a specified simulation sample size N, are estimated accordingly.

For the proposed method, a fixed simulation sample size 10^4 is first used to provide one level of comparison, and it is next simulated with an updated simulation sample size to meet a target variability level. For a fair comparison, SS is simulated using the same initial simulation sample as the proposed method and the method in Ling et al. [14] is simulated using a fixed simulation sample size 10^4. The spread of the predicted failure probabilities is also summarized in a box plot, where the center line in the box represents the median, the top edge represents 75^{th} percentile, the bottom edge represents 25^{th} percentile, and the outliers are plotted as '+' symbol.

For the Kriging-based methods, the regression model used is linear for the high-dimensional problem, and zero-order polynomial for the other examples. The correlation function is assumed as Gaussian. The Kriging model in the proposed method is mainly intended to help efficiently calculate intermediate thresholds, generate simulation samples, and calculate the conditional failure probability at the final level. For Kriging model error estimation, please refer to [22,23]. An indication of the overall contribution of the active learning Kriging procedure to the *MSE* of the proposed method may be obtained by comparing the *MSE* from the SS method against that from the proposed method for the same initial simulation sample.

5.2 Series system with four branches

Consider a system comprising several subsystems or components, where the failure of the system is defined as the failure of any component. That is, the failure event is represented by the union of all the component failure events. Such formulation is applicable to many civil engineering problems. The first example [8, 24] is a system comprising four subsystems, two with linear and two with nonlinear performance functions. The performance function of this system with random variables θ_1 and θ_2 following the standard normal distribution is given by:

$$G(\theta_1,\theta_2)=\min\begin{Bmatrix} m+0.1(\theta_1-\theta_2)^2-\dfrac{(\theta_1+\theta_2)}{\sqrt{2}}; \\ m+0.1(\theta_1-\theta_2)(\theta_1-\theta_2)^2+\dfrac{(\theta_1+\theta_2)}{\sqrt{2}}; \\ (\theta_1-\theta_2)+\dfrac{n}{\sqrt{2}}; \\ (\theta_2-\theta_1)+\dfrac{n}{\sqrt{2}}. \end{Bmatrix} \tag{19}$$

where $m = 4.8$ and $n = 10$.

The computation starts in step 1 with the generation of a simulation sample $\boldsymbol{\Theta}^{(0)}$ with size $N = 10^4$ at simulation level $i = 1$. The training points, with an initial size of $n_0 = 14$ (section 4.3 discusses this choice), are evaluated using Eq. (19) to form the first training set. The Kriging model is trained and updated in step 2, where new training points are added. This results in 38 training points, where 24 new training points (shown in parenthesis in Table 2 and marked as "x" in Fig. 2(a)) have been added based on the active learning technique. The updated Kriging model (which is computationally cheaper) is then used to compute the performance values $\hat{G}\left(\boldsymbol{\Theta}^{(i-1)}\right)$ using the large set of simulation sample (of size $N = 10^4$) from which the intermediate threshold y_1 is estimated. For the conditional failure probability of 0.1, $y_1 = 3.197$ is obtained (step 3). As shown in Fig. 2(a), the surrogate model plotted at $y_1 = 3.197$ is in good agreement with that of the actual performance function. In step 4, a new simulation sample $\boldsymbol{\Theta}^{(1)}$ for level $i = 2$ is generated using the modified Metropolis algorithm conditioned on $y_1 = 3.197$. Steps 2 to 4 are repeated until y_i is smaller than 0 (Fig. 2(b) to 2(f)). The failure probability ($\tilde{P}_F = 1.416\times10^{-6}$) and the associated δ_{SS} are computed in step 5. As $\delta_{SS} > 0.05$, the sample size N is increased to 60,000 (based on Eq. (18)), and the entire simulation is repeated until $\delta_{SS} \leq 0.05$. For this entire simulation run, the final solution is 1.521×10^{-6} using a total of $N_{call} = 152$ expensive evaluations via the actual performance function. The results are summarized in Table 2 and Figs. 2 and 3.

Table 2 gives the threshold values and cumulative N_{call} at each of the six levels including the number of points added based on the active learning algorithm for the initial sample size N of 10^4 and

the next and final recommended sample size of 60,000. All the 125 training points from the first iteration ($N = 10^4$) are passed to the next iteration (with $N = 60{,}000$) which explains why the corresponding additional training points are small for the second iteration.

Table 2 Results at each simulation level with 2 simulation sample sizes (N)

N	i	1	2	3	4	5	6	$\tilde{P}_F$
10^4	P_i	0.1	0.1	0.1	0.1	0.1	0.1416	1.416×10^{-6}
	N_{call}	38 (24)	50 (12)	68 (18)	94 (26)	109 (15)	125 (16)	
	y_i	3.197	2.345	1.601	0.976	0.455	0	
60,000	P_i	0.1	0.1	0.1	0.1	0.1	0.1521	1.521×10^{-6}
	N_{call}	126 (1)	127 (1)	127 (0)	132 (5)	142(10)	152 (10)	
	y_i	3.109	2.314	1.602	0.977	0.424	0	

Figs. 2 and 3 show that the sample points are well distributed based on the proposed procedure. The points added by the active learning algorithm are close to the contours associated with the respective thresholds. The plot of the Kriging model at each simulation level shows that it can represent the actual performance function well, with some discrepancies at locations where the contributions to the failure probability are small (away from the simulation sample).

As discussed in section 5.1, 100 runs of this example are performed for comparison. The results using crude MCS, SS method [5], method of Ling et al. [14] and the proposed method are shown in Table 3. In Table 3, the $E(\hat{P}_F)$ of MCS results is taken as a reference failure probability to compute the prediction error for other methods; the result for the proposed method is shown at the last row of the table, and the SS [5] using the same initial simulation sample as the proposed method is shown at the second row. The results for the proposed method and the method in Ling et al. [14] using the fixed simulation size 10^4 are also compared. The box plot of the predicted failure probabilities from different methods is shown in Fig. 4. The figure shows that P_F predicted by the proposed method has less variability than the method of Ling et al. [14].

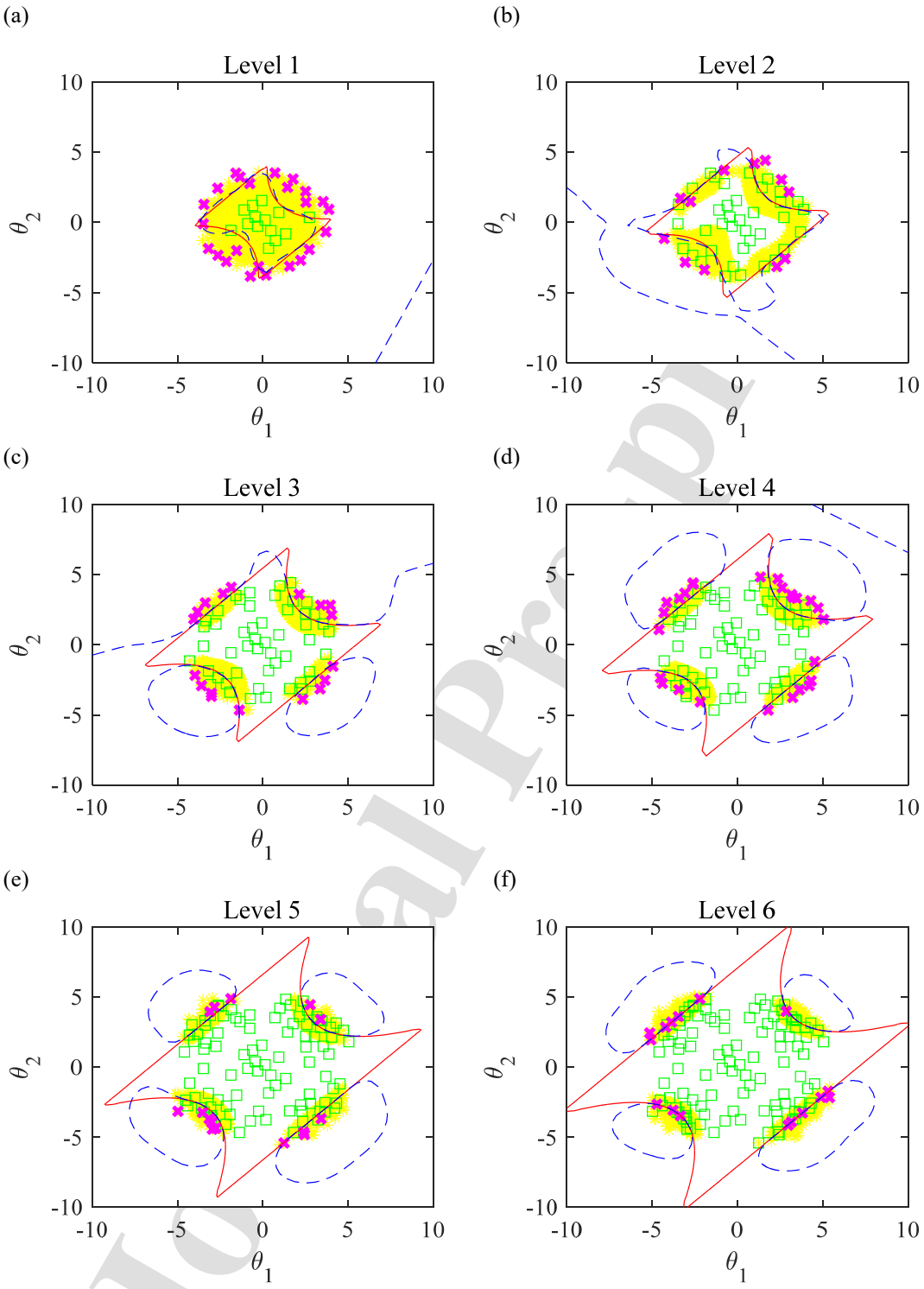

Fig. 2 Results of subset simulation levels for with $N = 10^4$ (yellow asterisk '⋆': simulation sample; green square: existing training set; magenta cross 'x': new training points; red solid line: performance function with y_i value; blue dashed line: Kriging model with y_i value).

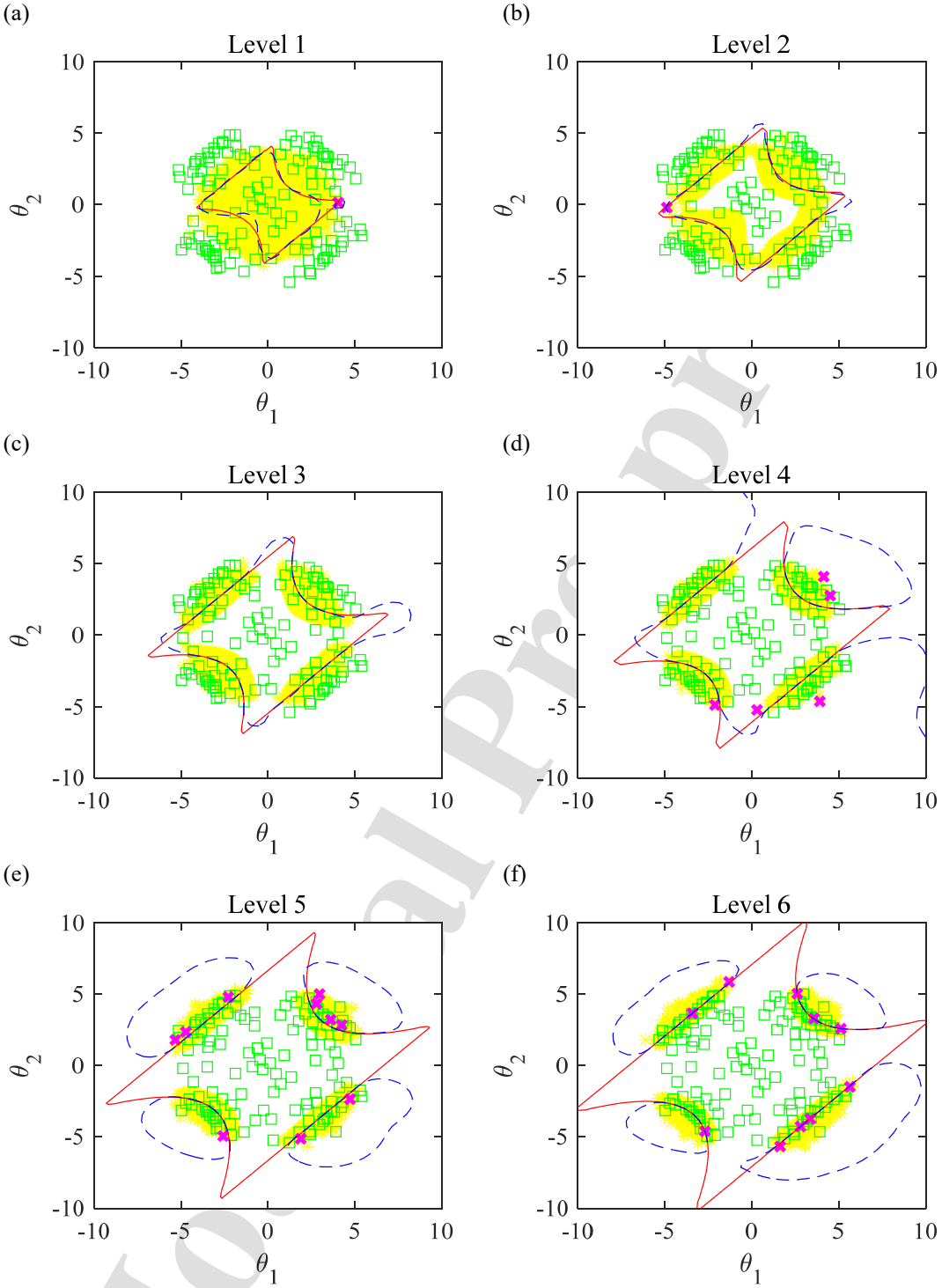

Fig. 3 Results of subset simulation levels for N = 60,000 (legend as in Fig. 2)

The *MSE* value from the proposed method is only 15% of the method of Ling et al. [14] (1.64×10^{-14} vs 1.10×10^{-13}) and N_{call} is less than half of Ling et al [14] (172 vs 385). Hence, the proposed method is computationally cheaper with less prediction error than the method of Ling et al. [14]. When using the same simulation sample size 10^4, the proposed method has *MSE* around 20% higher than that of Ling et al [14] (1.34×10^{-13} vs 1.10×10^{-13}) using only one-third N_{call} (127 vs 385). Compared to SS [5], the proposed method only uses 0.06% N_{call} (172 vs 305,981) with similar *MSE* (1.64×10^{-14} vs 1.69×10^{-14}), which indicates that the effect of the prediction error of the Kriging model on the failure probability prediction error is negligible.

Table 3 Comparison of results for series system with four branches example

Method	N	$E(\hat{P}_F)$	$E(N_{call})$	V	B^2(%MSE)	MSE
MCS		1.495×10^{-6}	10^8	1.55×10^{-14}		
SS [5]	60,740	1.478×10^{-6}	305,981	1.66×10^{-14}	3.01×10^{-16} (1.8)	1.69×10^{-14}
Ling et al. [14]	10^4	1.450×10^{-6}	385	1.08×10^{-13}	2.07×10^{-15} (1.9)	1.10×10^{-13}
Proposed	10^4	1.520×10^{-6}	127	1.33×10^{-13}	6.15×10^{-16} (0.5)	1.34×10^{-13}
	60,740	1.515×10^{-6}	172	1.60×10^{-14}	3.98×10^{-16} (2.4)	1.64×10^{-14}

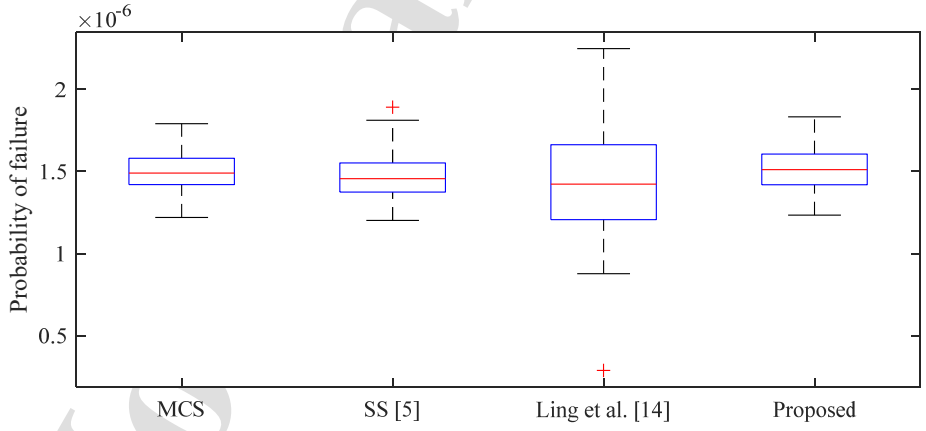

Fig. 4 Box plot of P_F for series system with four branches example

Theoretically, the bias should be independent of N, but small variations are observed in all methods as it is estimated numerically and constitutes only a small percentage (<3%) of the *MSE* (shown in parenthesis in Table 3). With regards to the initial sample size, as an illustration, if the simulation starts with a smaller N of 10^3 instead of 10^4, the final *MSE* obtained is slightly higher (1.67×10^{-14} vs 1.64×10^{-14}) and with a similar mean N_{call} (170 vs 172), which implies the results with initial N equals to 10^3 and 10^4 are similar.

No new training point is selected to update the Kriging model and the predicted $\hat{P}_F = 0$ when $N = 10^3$ and 10^4 using AK-MCS [8] and AK-SS [11]. As discussed in section 2.2, a large simulation sample size N is required for estimating low failure probability, and increasing N does not guarantee a new training point is selected following the active learning procedures in AK-MCS [8] and AK-SS [11]. With an initial $N = 10^7$, no new training point is selected ($\min(U(\boldsymbol{\Theta})) = 4.54$ using learning function U, and $\max(EFF(\boldsymbol{\Theta})) = 6.10\times10^{-5}$ using learning function *EFF*, see Table 1), and the predicted $\hat{P}_F = 0$. After the N is iteratively increased to 4×10^7, still no training point is selected, and the algorithm could not continue ("out of memory" error with a 16 GB RAM computer) as training Kriging model using DACE toolbox in MATLAB [8] with a sample size this large is infeasible with limited computational power. Further increasing N without changing the active learning procedure does not address the issue.

As AK-SS is based on AK-MCS with learning function U, it inherits the same limitation. The use of a global Kriging model works for problems with higher failure probability, where for the case of $m = 3$ and $n = 7$ (in Eq. (19)), AK-MCS and AK-SS predict a failure probability of 2.222×10^{-3}, where the required simulation sample size is relatively small. Fig. 5(b) shows that there are sufficient training points in both the safe and failure domains and the global Kriging model matches the actual LSF well to achieve a good estimation.

(a) (b)

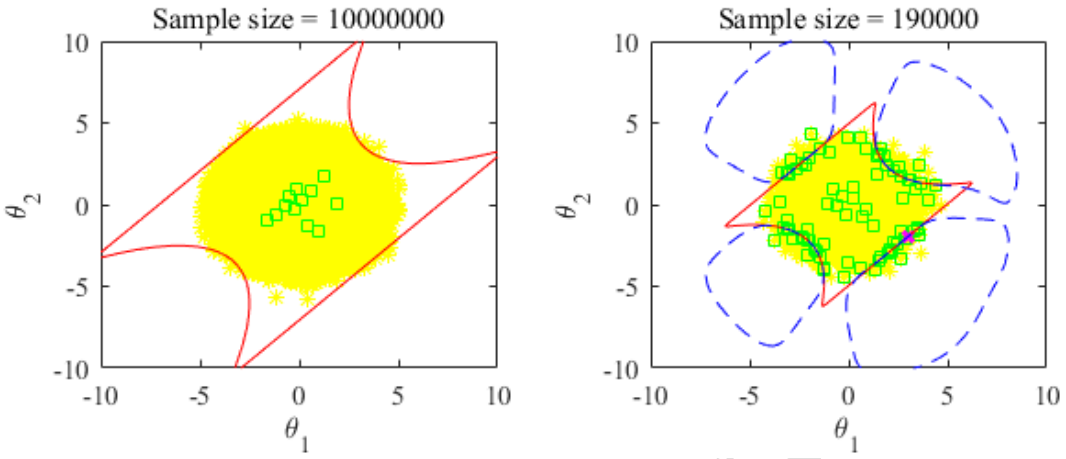

Fig. 5 AK-MCS for series system with four branches example (a) $m = 4.8$, $n = 10$, (b) $m = 3$, $n = 7$.

5.3 *Modified Rastrigin function*

To test the versatility of the proposed method, the computationally challenging Rastrigin function [25] is used here where the performance function is highly nonlinear and non-convex with disconnected failure domains (see Fig. 6). The Rastrigin function was originally used as a difficult performance test example for optimization algorithms in the domain of mathematical optimization, as it is a non-linear multimodal function with many local minima and a large search space. The performance function is given by

$$G(\theta_1, \theta_2) = 25 - \sum_{i=1}^{2} \left(\theta_i^2 - 5\cos(2\pi\theta_i)\right) \tag{20}$$

where θ_1 and θ_2 are standard normal distributed random variables.

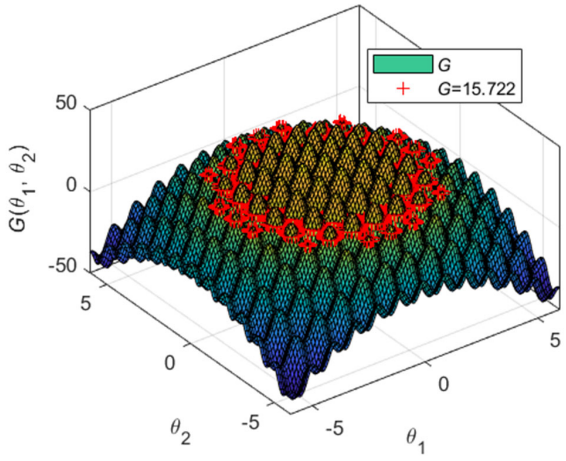

Fig. 6 Performance function and surface with threshold of 15.722 for modified Rastrigin function example (indicated as +)

In Fig. 6. the contour of the function at a threshold of 15.722 is illustrated in red color, and also plotted in a plan view (red color) in Fig. 7(a). This corresponds to the first level intermediate threshold simulated in the proposed method. In Fig. 7, The corresponding contour from the Kriging model is shown in dotted blue color where the match is very good (the blue and red curves virtually coincide) except outside the domain of simulation sample, which is not relevant. Despite the complexity of the performance function in terms of its nonlinearity and the discontinuity of the failure domains, Fig. 7 shows the suitability of the Kriging models at each simulation level.

The results are summarized in Table 4. The computational effort by the proposed method and that of Ling et al. [14] are similar (N_{call} = 633 vs 673) but the *MSE* from the proposed method is only 5% of Ling et al. [14] (1.05×10^{-11} vs 2.01×10^{-10}). If the proposed method does not update the simulation sample size, i.e., using a fixed size 10^4, the prediction error is ~15 times larger (1.62×10^{-10} vs 1.05×10^{-11}), with around 20% reduction of N_{call} (516 vs 633). The B^2 contributes 54% of the *MSE*, where the mean prediction (2.258×10^{-5}) deviates significantly from the reference value (3.197×10^{-5}). This demonstrates the necessity to update the simulation sample size as described in section 4.3. Comparing the proposed method with SS [5] using the same simulation sample size, the proposed

method uses 0.20% N_{call} (633 vs 315,206) with 10% more *MSE* (1.05×10^{-11} vs 9.57×10^{-12}), which partially results from the regression prediction error of the Kriging model.

The AK-MCS [8] with learning function *U* and AK-SS [11] are unable to update the training sample even with a large initial simulation sample size $N = 10^7$. The AK-MCS [8] with learning function *EFF* is able to select a few training points but the training of Kriging model fails due to the large sample size before the learning stopping criterion is met. For the method of Ling et al. [14], there are 23 out of 100 runs that do not predict reasonable results (giving $P_F < 10^{-7}$) due to the intermediate thresholds not being able to evolve close to the final threshold. This could be due to the low number of points used to determine the initial threshold y_1. In this case, y_1 calculated from the few points is associated with a small conditional failure probability. This will result in an insufficient number of points in the 'failure' domain to serve as chains for the generation of a new simulation sample. If the size of the initial set of 10 is increased, the efficiency of the method will decrease. No procedure for determining the size of the initial set is provided in Ling et al. [14].

Fig. 8 shows the box plot of P_F predicted by the different methods. The proposed method has similar variability with SS [5], whereas the method in Ling et al. [14] has much larger prediction variability.

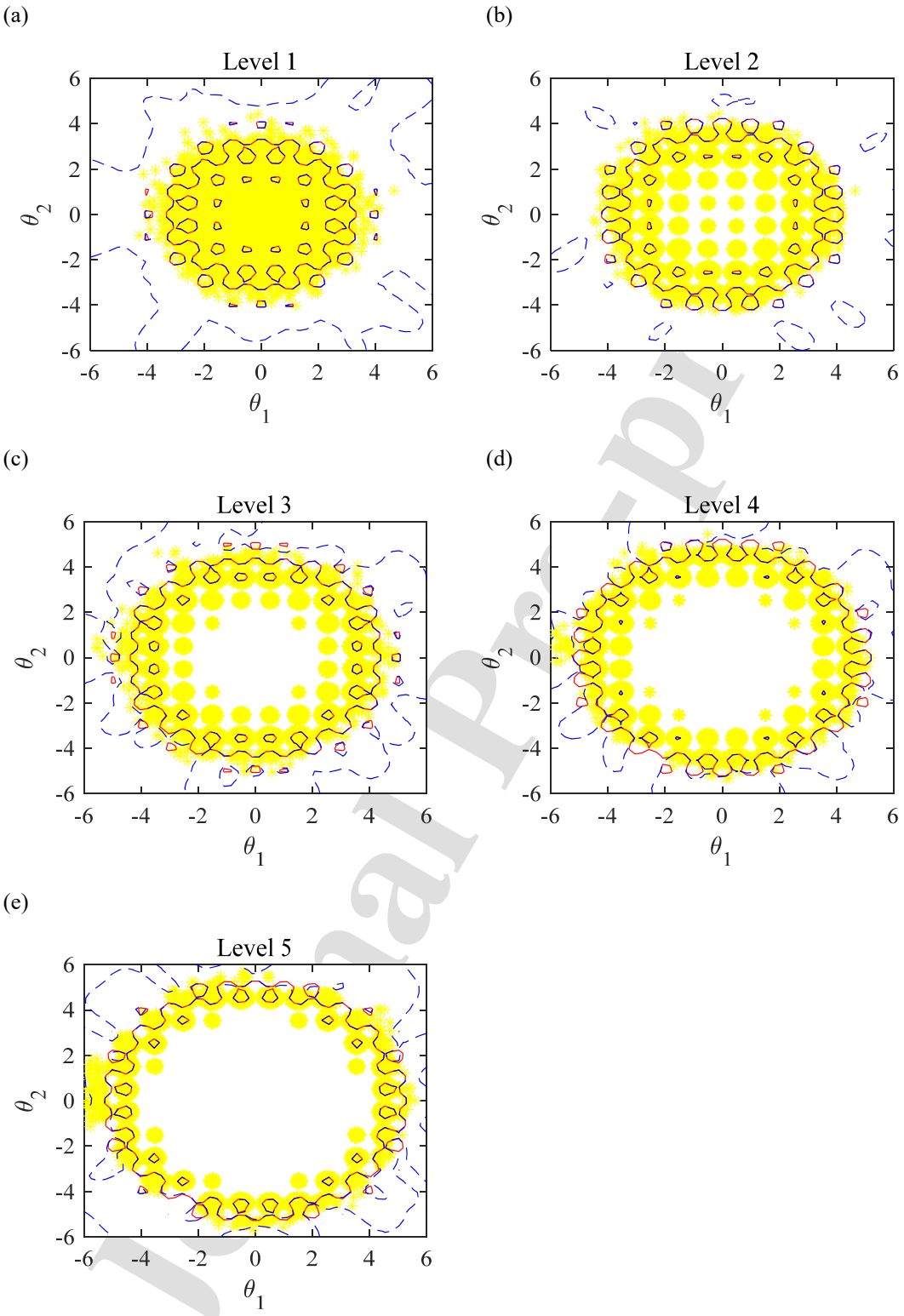

Fig. 7 Different simulation levels at last iteration for modified Rastrigin function example (legend as in Fig. 2)

Table 4 Comparison of results for modified Rastrigin function example

Method	N	$E(\hat{P}_F)$	$E(N_{call})$	V	B^2(%MSE)	MSE
MCS		3.197×10^{-5}	10^8	3.01×10^{-13}		
SS [5]	72,870	3.179×10^{-5}	315,206	9.53×10^{-12}	3.47×10^{-14} (0.4)	9.57×10^{-12}
Ling et al. [14]	10^4	2.393×10^{-5}	673	1.36×10^{-10}	6.48×10^{-11} (32)	2.01×10^{-10}
Proposed	10^4	2.258×10^{-5}	516	7.40×10^{-11}	8.82×10^{-11} (54)	1.62×10^{-10}
	72,870	3.132×10^{-5}	633	1.01×10^{-11}	4.31×10^{-13} (4.1)	1.05×10^{-11}

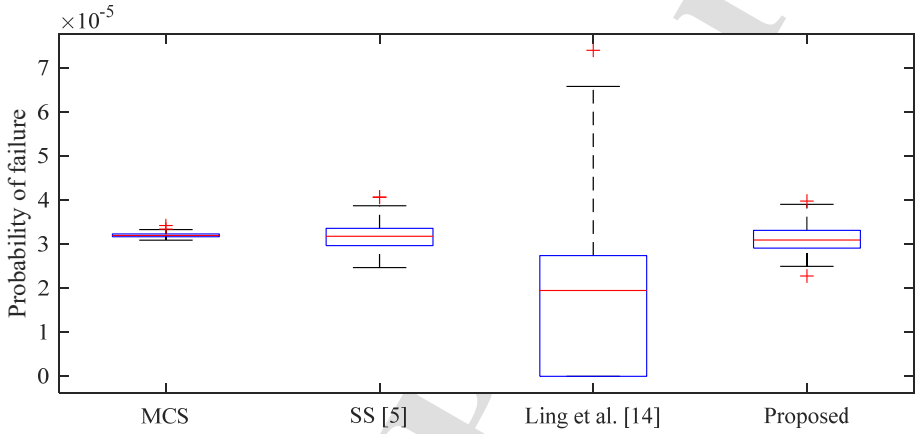

Fig. 8 Box plot of P_F for modified Rastrigin function example

5.4 *Response of a non-linear oscillator*

A SDOF oscillator (Fig. 9) with nonlinear restoring force and pulse load with random physical parameters is considered. The performance function of the response of the non-linear oscillator is given by [8,24,26,27]:

$$G(c_1,c_2,m,r,t_1,F_1)=3r-|z_{\max}|=3r-\left|\frac{2F_1}{m\omega_0^2}\sin\left(\frac{\omega_0 t_1}{2}\right)\right| \tag{21}$$

where r is the displacement at which one of the springs yields, $z_{\max}$ is the maximum displacement and $\omega_0=\sqrt{\frac{c_1+c_2}{m}}$. The statistics of the random variables are listed in Table 5 and the results are tabulated in Table 6.

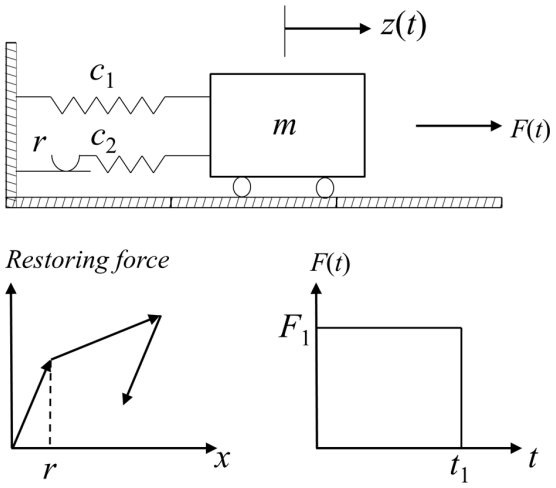

Fig. 9 Nonlinear SDOF oscillator

The *MSE* value from the proposed method is less than 10% (1.68×10^{-14} vs 1.94×10^{-13}) and N_{call} is less than two-thirds of Ling et al [14] (198 vs 324). When using the same simulation sample size 10^4, the proposed method has a smaller *MSE* than Ling et al [14] (8.57×10^{-14} vs 1.94×10^{-13}) using less than 40% of N_{call} (123 vs 324). Compared to SS [5], the proposed method only uses 0.07% N_{call} (198 vs 291,293) with 10% higher *MSE* (1.68×10^{-14} vs 1.51×10^{-14}), which partially results from the regression prediction error of the Kriging model. Same as the previous examples, the AK-MCS [8] and AK-SS [11] have limited capabilities to apply on this example.

If N starts with 10^3 for the proposed method, the *MSE* increases more than 40% (2.39×10^{-14} vs 1.68×10^{-14}) with mean N_{call} decreased by 3% (193 vs 198), implying a slight decrease in the computational effort with a large increase in *MSE*. Using 10^4 is a better choice.

Table 5 Statistics of random variables for non-linear oscillator example

Variable	Mean	Standard deviation	Distribution type
m	1	0.05	Normal
c_1	1	0.1	Normal
c_2	0.1	0.01	Normal
r	0.85	0.05	Normal
F_1	1	0.2	Normal
t_1	1	0.2	Normal

Table 6 Comparison of results for response of non-linear oscillator example

Method	N	$E(\hat{P}_F)$	$E(N_{call})$	V	B^2(%MSE)	MSE
MCS		1.908×10^{-6}	10^8	2.06×10^{-14}		
SS [5]	52,970	1.925×10^{-6}	291,293	1.48×10^{-14}	2.83×10^{-16} (1.9)	1.51×10^{-14}
Ling et al. [14]	10^4	1.861×10^{-6}	324	1.92×10^{-13}	2.18×10^{-15} (1.1)	1.94×10^{-13}
Proposed	10^4	1.885×10^{-6}	123	8.52×10^{-14}	5.22×10^{-16} (0.6)	8.57×10^{-14}
	52,970	1.927×10^{-6}	198	1.64×10^{-14}	3.74×10^{-16} (2.2)	1.68×10^{-14}

Fig. 10 shows the box plot of P_F predicted by the different methods. The proposed method has similar variability with SS [5] but without outliers, whereas the method in Ling et al. [14] has much larger prediction variability with outliers.

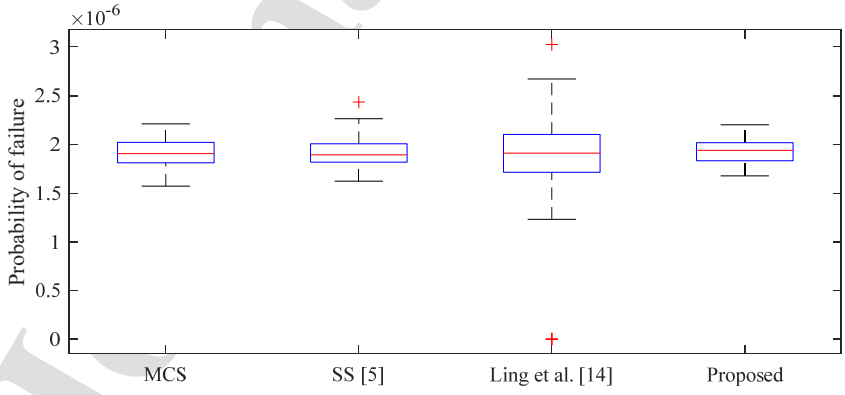

Fig. 10 Box plot of P_F for response of non-linear oscillator example

5.5 Ten-bar truss

Consider a ten-bar truss structure [28] illustrated in Fig. 11. The cross-sectional areas of horizontal, vertical, and diagonal members are A_1, A_2, and A_3 respectively, each with the same modulus of elasticity, denoted by E. The length of each horizontal and vertical member is L. The response variable of interest is the vertical displacement at node 2 (assuming the truss remain elastic) shown in Fig. 11, and the performance function is given by [28]

$$G = 0.076 - \frac{BPL}{A_1 A_3 E D_T}\left\{\begin{array}{l} 4\sqrt{2}A_1^3\left(24A_2^2 + A_3^2\right) + A_3^3\left(7A_1^2 + 26A_2^2\right) + \\ 4A_1A_2A_3\left\{\left(20A_1^2 + 76A_1A_2 + 10A_3^2\right) + \sqrt{2}A_3\left(25A_1 + 29A_2\right)\right\} \end{array}\right\} \quad (22)$$

where $D_T = 4A_2^2(8A_1^2 + A_3^2) + 4\sqrt{2}A_1A_2A_3(3A_1 + 4A_2) + A_1A_3^2(A_1 + 6A_2)$. It is assumed that A_1, A_2, A_3, B and E are independent random variables, in which B accounts for model uncertainty, the uncertainties due to manufacturing tolerances are modelled in A_1, A_2, and A_3, and the uncertainty in material property modelled in E. The distributions of the random variables are provided in Table 7. The static load $P = 2.5\times10^5$ N and the length $L = 9$ m are assumed to be deterministic.

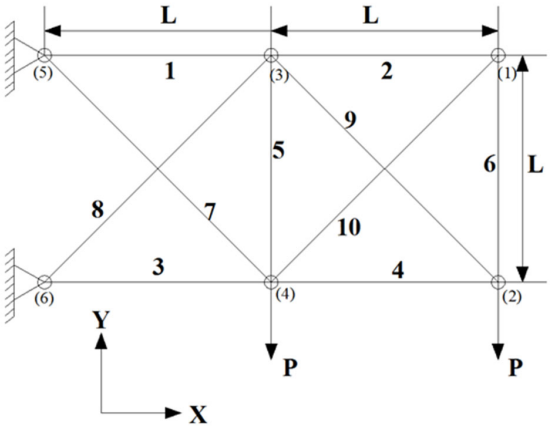

Fig. 11 Ten-bar truss structure

Table 7 Distributions of random variables for ten-bar truss example

Variable	Mean	c.o.v.	Distribution type
A_1	1×10^{-2} m^2	0.05	Normal
A_2	1.5×10^{-3} m^2	0.05	Normal
A_3	6×10^{-3} m^2	0.05	Normal
B	1.0	0.1	Normal
E	6.9×10^{4} MPa	0.05	Lognormal

The results are compared in Table 8 and Fig. 12. The *MSE* value from the proposed method is less than 10% (5.29×10^{-14} vs 5.59×10^{-13}) and N_{call} is less than 80% of Ling et al [14] (126 vs 160). When using the same simulation sample size 10^4, the proposed method has *MSE* of 2.59×10^{-13} which is 46% of 5.59×10^{-13} from the method of Ling et al [14] while using smaller N_{call} (74 vs 160). Compared to SS [5], the proposed method only uses 0.05% in terms of N_{call} (126 vs 289,377) with 30% higher *MSE* (5.29×10^{-14} vs 4.09×10^{-14}), which is partially due to the regression from the Kriging model. The box plot in Fig. 12 shows the proposed method and SS [5] have similar prediction variability, while the method in Ling et al [14] has larger variability with outliers. The AK-MCS [8] and AK-SS [11] have limited capabilities to apply on this example as the target failure probability is small.

If N starts with 10^3 for the proposed method, the *MSE* is increased by more than 50% (8.09×10^{-14} vs 5.29×10^{-14}) with one less mean N_{call} (125 vs 126), implying 10^4 is a better choice as an initial simulation sample size. With regards to the threshold for the stopping criterion in Eq. (17), decreasing the threshold from 0.1 to 0.05 renders 6% more N_{call} (134 vs 126), and 30% less *MSE* (3.59×10^{-14} vs 5.29×10^{-14}); on the other hand, increasing the threshold from 0.1 to 0.2 renders 2% less mean N_{call} (123 vs 126) and ~2.2 times *MSE* (1.19×10^{-13} vs 5.29×10^{-14}). For this example, 0.2 is not appropriate, and whether to use a threshold smaller than 0.1 is a matter of choice.

Table 8 Comparison of results for ten-bar truss example

Method	N	$E(\hat{P}_F)$	$E(N_{call})$	V	B^2(%MSE)	MSE
MCS		2.767×10^{-6}	10^8	2.25×10^{-14}		
SS [5]	52,650	2.762×10^{-6}	289,377	4.09×10^{-14}	2.24×10^{-17} (0.1)	4.09×10^{-14}
Ling et al. [14]	10^4	2.633×10^{-6}	160	5.41×10^{-13}	1.80×10^{-14} (3.2)	5.59×10^{-13}
Proposed	10^4	2.779×10^{-6}	74	2.59×10^{-13}	1.62×10^{-16} (0.1)	2.59×10^{-13}
	52,650	2.756×10^{-6}	126	5.28×10^{-14}	1.11×10^{-16} (0.2)	5.29×10^{-14}

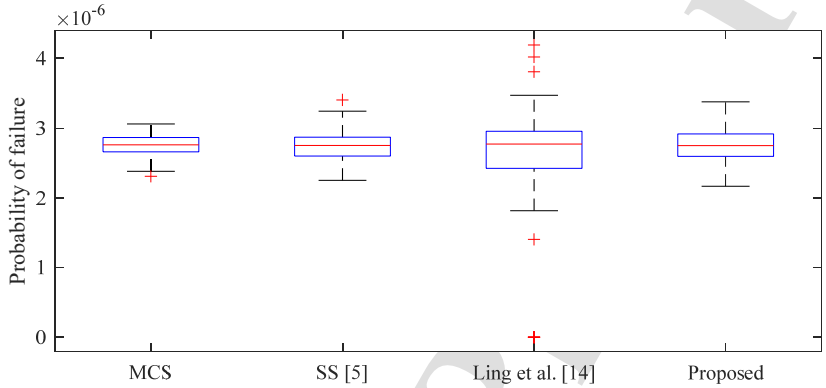

Fig. 12 Box plot of P_F for ten-bar truss example

5.6 *High-dimensional problem*

This example illustrates some issues that may be encountered with high-dimensional problems. The performance function of a problem with 50 dimensions modified from [11] is used, given by

$$G\left(\theta_1,\theta_2,\ldots,\theta_{n_d}\right)=\left(n_d+3\sigma\sqrt{n_d}\right)-\sum_{i=1}^{n_d}\theta_i-\sqrt{\theta_1}-\sqrt[3]{\theta_2} \tag{23}$$

where $n_d = 50$ and the n_d random variables each follows a lognormal distribution with mean value 1 and standard deviation $\sigma = 0.2$. The last two terms make the original performance function in [11] nonlinear. A linear regression model is used in the Kriging model and the initial training size $n_0 = 62$ is used (see section 4.3).

The results are compared in Table 9 and Fig. 13. The *MSE* value from the proposed method is less than 15% of the method of Ling et al. [14] (8.57×10^{-13} vs 5.90×10^{-12}) and N_{call} is around 40% of Ling et al [14] (138 vs 334). When using the same simulation sample size 10^4, the proposed method has *MSE* that is less than 60% of Ling et al [14] (8.57×10^{-13} vs 5.90×10^{-12}) using around 40% N_{call} (138 vs 334). Compared to SS [5], the proposed method only uses 0.07% N_{call} (138 vs 205,114) with around 30% higher *MSE* (8.57×10^{-13} vs 6.51×10^{-13}), which partially results from the regression prediction error of the Kriging model. For AK-MCS [8] and AK-SS [11], even though the target failure probability is higher than the previous examples, but with a much higher dimension, the training of the Kriging model fails without a new training point being selected. The box plot in Fig. 13 shows the proposed method and SS [5] have similar prediction variability, whereas the method in Ling et al. [14] has much larger prediction variability.

Table 9 Comparison of results for high dimensional problem example

Method	N	$E(\hat{P}_F)$	$E(N_{call})$	V	B^2*(%MSE)*	*MSE*
MCS		1.251×10^{-5}	10^8	1.37×10^{-13}		
SS [5]	44,590	1.245×10^{-5}	205,114	6.47×10^{-13}	3.46×10^{-15} (0.5)	6.51×10^{-13}
Ling et al. [14]	10^4	1.293×10^{-5}	334	5.72×10^{-12}	1.75×10^{-13} (3.0)	5.90×10^{-12}
Proposed	10^4	1.284×10^{-5}	95	3.33×10^{-12}	1.12×10^{-13} (3.3)	3.44×10^{-12}
	44,590	1.250×10^{-5}	138	8.57×10^{-13}	1.60×10^{-17} (0.0)	8.57×10^{-13}

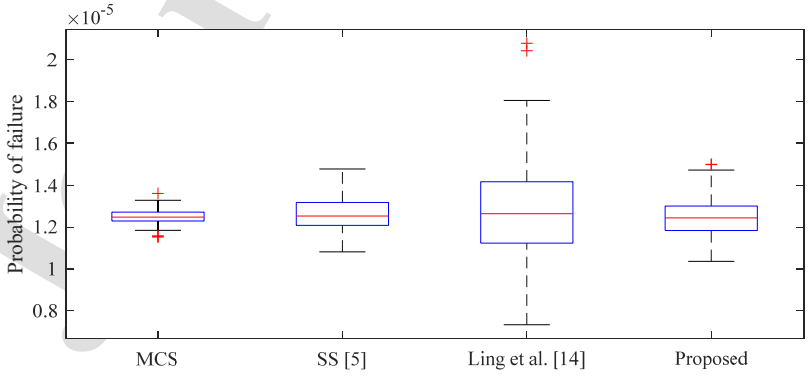

Fig. 13 Box plot of P_F for high-dimensional problem

5.7 Dynamic analysis of an offshore drilling riser system

The reliability of an offshore drilling riser system subjected to environmental wave loads is investigated, where time domain dynamic analyses are performed using the finite element software OrcaFlex [29]. Fig. 14 shows the model of the drilling riser system. The water depth considered is 1200m and the casing is 1000m deep. The top end of the riser is connected to the drillship, while the bottom end is fixed at the seabed. The main uncertainties arise from the wave height H and wave period T of the incoming waves described by Dean's stream regular wave function [29]. H is modelled by the Rayleigh distribution (with scale parameter 1.8 m), and T follows the three-parameter Weibull distribution (with location parameter 4 s, scale parameter 25 s, and shape parameter 6 s.). The mean of T is 27.193 s, the duration of the simulation time history for each run is set as 3 time of T.

The first-order vessel motions are simulated using the Response Amplitude Operator (RAO) while the waves loads are simulated using OrcaFlex. One potential failure mechanism of a drilling riser system is at the upper flex joint. Failure is defined by the exceedance of the allowable limit of the absolute maximum upper flex joint angle $|\varepsilon_{\max}|$. The LSF is given by

$$G(H,T) = 2.52 - |\varepsilon_{\max}| \tag{24}$$

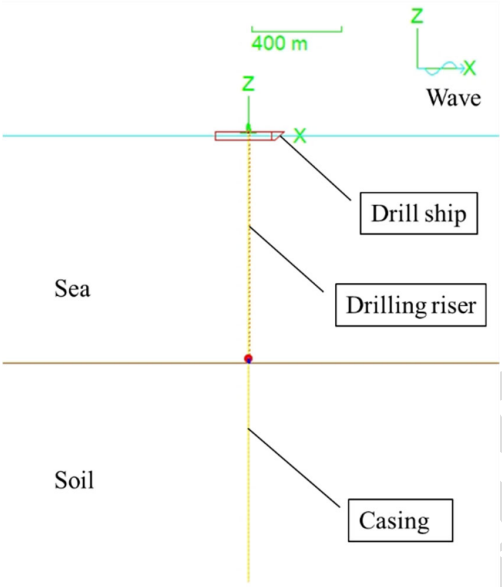

Fig. 14 Drilling riser modelled in OrcaFlex

One key issue is the excessive computational time taken to evaluate Eq. (24) as it requires the extraction of extreme values from nonlinear time domain analyses. One single MCS run takes ~2 weeks with parallel computing using 24 threads. As such, only a single run with a sample size of 10^6 is used to estimate P_F, and the prediction variance 1.37×10^{-10} is approximated based on Eq. (3). The mean time for each run using the proposed method is ~4 h (much less than ~2 weeks). The initial training size for the proposed method is 14 (section 4.3 discusses this choice). The method of Ling et al. [14] is compared, while SS [5] is not run 100 times for comparison due to its large number of performance function evaluations required. From one of the 100 simulations, the last iteration of the proposed method is plotted in Fig. 15. The absolute maximum upper flex joint angle is governed by the surge motion of the vessel, where the maximum surge resonant response occurs when the wave period T is around 15s. It could be seen that the training points are spread across the whole domain while more points are close to the limit state function.

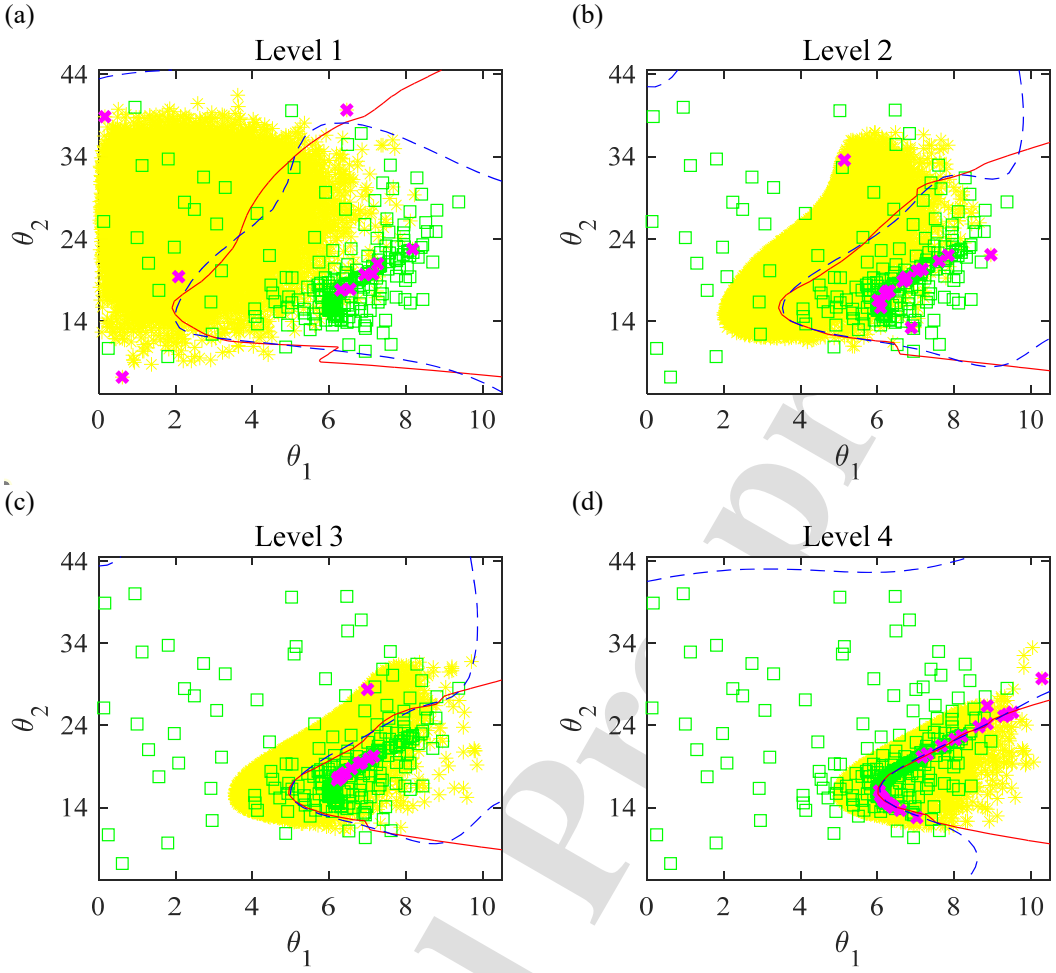

Fig. 15 Results of subset simulation levels for N = 40,000 (legend as in Fig. 2, θ_1 is H and θ_2 is T)

The results are compared in Table 10. The *MSE* value from the proposed method is less than 20% of the method of Ling et al. [14] (1.02×10^{-10} vs 4.96×10^{-10}) and N_{call} is around 55% of Ling et al [14] (356 vs 647). When using the same simulation sample size 10^4, the proposed method has ~2.5 times *MSE* than Ling et al [14] (1.24×10^{-9} vs 4.96×10^{-10}) using around 30% N_{call} (356 vs 647).

Even with a large initial simulation sample size $N = 10^7$ for AK-MCS [8] with learning function U, the stopping criterion is met ($\min(U(\boldsymbol{\Theta})) = 5.88$) with no new training point being selected, and when the enriched simulation sample size reaches 4×10^7, the Kriging model training fails. AK-SS [11] has the same limitation. The results from AK-MCS [8] with learning function *EFF* (with initial $N = 10^6$) are listed in Table 10, and it has the lowest number N_{call} with the smallest *MSE*. It shows its

better efficiency over other methods for this relatively high failure probability problem ($P_F > 10^{-4}$) as only a global Kriging model is trained instead of multiple Kriging models on different simulation levels. But it has limitations on low failure probability problems as are discussed in previous examples.

Table 10 Results for offshore drilling riser example

Method	N	$\widehat{P}_F/E(\widehat{P}_F)$	$E(N_{call})$	V	B^2(%MSE)	MSE
MCS		1.370×10^{-4}	10^6	1.37×10^{-10}*		
AK-MCS [8]	3.3×10^6	1.346×10^{-4}	79	1.70×10^{-11}	5.94×10^{-12} (26)	2.29×10^{-11}
Ling et al. [14]	10^4	1.305×10^{-4}	647	4.54×10^{-10}	4.19×10^{-11} (8.5)	4.96×10^{-10}
Proposed	10^4	1.236×10^{-4}	205	1.06×10^{-9}	1.79×10^{-10} (14)	1.24×10^{-9}
	3.9×10^4	1.345×10^{-4}	356	9.58×10^{-11}	6.47×10^{-12} (6.3)	1.02×10^{-10}

*Calculated based on Eq. (3).

6. Conclusions

This paper presents an improved method coupling SS with an active learning Kriging model for low failure probability prediction. An active learning technique with new learning and stopping criteria to select training points based on the prediction uncertainty associated with each point is employed to efficiently update the Kriging model for approximating the performance function at each simulation level. The Kriging model facilitates the estimation the intermediate threshold as well as the conditional failure probability associated with the target threshold to be determined with low prediction error and computational cost. This large sample size (N) is estimated based on a specified coefficient of variation of the predicted failure probability, whereas the number of costly performance function evaluations (N_{call}), is determined by the proposed active learning technique. The training points at a simulation level are passed to the subsequent level without no increase in the computation cost.

The proposed method, the method of Ling et al. [14] and the SS [5] are compared in terms of efficiency (number of expensive computation calls, N_{call}) and prediction error (*MSE*). For low failure probability prediction with AK-MCS [8] and AK-SS [11], a large simulation sample size is required, which makes the Kriging model training computationally expensive or even infeasible; and the active learning procedures do not guarantee a new training point is selected, which could lead to the failure of the algorithms. The limitations are demonstrated in the series system with four branches example (see section 5.2) and discussed in the other examples, justifying the use of different Kriging models at different simulation levels with a new active learning technique in the proposed method. It is shown that the original SS method produced the lowest *MSE* but with much higher N_{call}, and the proposed method requires lower N_{call} and has lower *MSE* than that of Ling et al. [14]. For the proposed method, the effect of simulation sample size N and the prediction error introduced by the Kriging model are also discussed. Using an initial simulation sample size N of 10^3 has more than 40% higher *MSE* than 10^4 (as is discussed in section 5.4 and 5.5) with similar N_{call}, which shows 10^4 is a better choice as N. With the use of active learning technique, the initial training set need not be large as more points will be added subsequently. The active learning threshold of 0.1 seems appropriate when compared with results from thresholds of 0.01 and 0.2 as demonstrated in the ten-bar truss example (see section 5.5).

The proposed method hinges on the quality of the surrogate model that can be derived from Kriging and the number of training points recommended by the active learning technique. In section 5.6, the proposed method is shown to work for a problem with dimensions of 50. Although the issue of high dimensionality in the original SS method has already been resolved to some extent by researchers, the capability of the Kriging model to approximate high dimensional performance function (such as a SDOF linear oscillator subjected to white noise excitation with thousands of random variables [5]) without using a sizeable training set remains a challenge and needs further research. Another limitation relates to an interest in getting the whole CDF curve covering the entire domain based on a single run of the proposed method, where the number of thresholds used is small and hence may not have adequate points to approximate the CDF. The first intermediate threshold corresponds to a probability of 0.1, with the next one with a probability of 0.01, and so on. Hence, the

tail of the CDF near the failure domain may be approximated fairly well but not the entire CDF. To get the CDF from a single run, other reliability methods (such as density estimation based methods) should be adopted.

Acknowledgements

The financial support provided by NUS to carry out this research under Grant No. C-302-000-024-001 is gratefully acknowledged.

References

[1] O. Ditlevsen, P. Bjerager, Methods of structural systems reliability, Struct. Saf. (1986). https://doi.org/10.1016/0167-4730(86)90004-4.

[2] S.T. Tokdar, R.E. Kass, Importance sampling: a review, Wiley Interdiscip. Rev. Comput. Stat. 2 (2010) 54–60. https://doi.org/10.1002/wics.56.

[3] S. Rahman, H. Xu, A univariate dimension-reduction method for multi-dimensional integration in stochastic mechanics, Probabilistic Eng. Mech. (2004). https://doi.org/10.1016/j.probengmech.2004.04.003.

[4] J. Nie, B.R. Ellingwood, Directional methods for structural reliability analysis, Struct. Saf. 22 (2000) 233–249. https://doi.org/10.1016/S0167-4730(00)00014-X.

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

[6] N. Pedroni, E. Zio, An adaptive metamodel-based subset importance sampling approach for the assessment of the functional failure probability of a thermal-hydraulic passive system, Appl. Math. Model. 48 (2017) 269–288. https://doi.org/10.1016/j.apm.2017.04.003.

[7] Q. Pan, D. Dias, An efficient reliability method combining adaptive Support Vector Machine and Monte Carlo Simulation, Struct. Saf. (2017). https://doi.org/10.1016/j.strusafe.2017.04.006.

[8] B. Echard, N. Gayton, M. Lemaire, AK-MCS: An active learning reliability method combining Kriging and Monte Carlo Simulation, Struct. Saf. 33 (2011) 145–154. https://doi.org/10.1016/j.strusafe.2011.01.002.

[9] B. Echard, N. Gayton, M. Lemaire, N. Relun, A combined importance sampling and Kriging reliability method for small failure probabilities with time-demanding numerical models, Reliab. Eng. Syst. Saf. 111 (2013) 232–240. https://doi.org/10.1016/j.ress.2012.10.008.

[10] N. Lelièvre, P. Beaurepaire, C. Mattrand, N. Gayton, AK-MCSi: A Kriging-based method to deal with small failure probabilities and time-consuming models, Struct. Saf. 73 (2018) 1–11. https://doi.org/10.1016/j.strusafe.2018.01.002.

[11] X. Huang, J. Chen, H. Zhu, Assessing small failure probabilities by AK-SS: An active learning method combining Kriging and subset simulation, Struct. Saf. 59 (2016) 86–95. https://doi.org/10.1016/j.strusafe.2015.12.003.

[12] V. Papadopoulos, D.G. Giovanis, N.D. Lagaros, M. Papadrakakis, Accelerated subset simulation with neural networks for reliability analysis, Comput. Methods Appl. Mech. Eng. 223–224 (2012) 70–80. https://doi.org/10.1016/j.cma.2012.02.013.

[13] F. Cui, M. Ghosn, Implementation of machine learning techniques into the subset simulation method, Struct. Saf. 79 (2019) 12–25. https://doi.org/10.1016/j.strusafe.2019.02.002.

[14] C. Ling, Z. Lu, K. Feng, X. Zhang, A coupled subset simulation and active learning Kriging reliability analysis method for rare failure events, Struct. Multidiscip. Optim. 60 (2019) 2325–2341. https://doi.org/10.1007/s00158-019-02326-3.

[15] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, Equation of state calculations by fast computing machines, J. Chem. Phys. 21 (1953) 1087–1092. https://doi.org/10.1063/1.1699114.

[16] J. Sacks, W.J. Welch, T.J. Mitchell, H.P. Wynn, Design and analysis of computer experiments,

Stat. Sci. (1989). https://doi.org/10.1214/ss/1177012413.

[17] S.N. Lophaven, H.B. Nielsen, J. Søndergaard, DACE, a MATLAB Kriging toolbox, version 2.0. Tech. Rep. IMM-TR-2002-12; Technical University of Denmark; 2002a.

[18] S.N. Lophaven, H.B. Nielsen, J. Søndergaard. Aspects of the MATLAB toolbox DACE. Tech. Rep. IMM-REP-2002-13; Technical University of Denmark; 2002b.

[19] B. Settles, Active learning kiterature survey, Mach. Learn. (2010). https://doi.org/10.1.1.167.4245.

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

[21] D. Wackerly, W. Mendenhall, R.L. Scheaffer, Mathematical statistics with applications, Thomson, 2008.

[22] W. Jian, S. Zhili, Y. Qiang, L. Rui, Two accuracy measures of the Kriging model for structural reliability analysis, Reliab. Eng. Syst. Saf. (2017). https://doi.org/10.1016/j.ress.2017.06.028.

[23] Z. Wang, A. Shafieezadeh, REAK: Reliability analysis through error rate-based adaptive Kriging, Reliab. Eng. Syst. Saf. 182 (2019) 33–45. https://doi.org/10.1016/j.ress.2018.10.004.

[24] L. Schueremans, D. Van Gemert, Benefit of splines and neural networks in simulation based structural reliability analysis, Struct. Saf. 27 (2005) 246–261. https://doi.org/10.1016/j.strusafe.2004.11.001.

[25] H. Mühlenbein, M. Schomisch, J. Born, The parallel genetic algorithm as function optimizer, Parallel Computing. 17 (1991) 619–632.

[26] M.R. Rajashekhar, B.R. Ellingwood, A new look at the response surface approach for reliability analysis, Struct. Saf. (1993). https://doi.org/10.1016/0167-4730(93)90003-J.

[27] C.G. Bucher, U. Bourgund, A fast and efficient response surface approach for structural reliability problems, Struct. Saf. (1990). https://doi.org/10.1016/0167-4730(90)90012-E.

[28] S.K. Choi, R.A. Canfield, R. V. Grandhi, Reliability-based structural design, 2007. https://doi.org/10.1007/978-1-84628-445-8.

[29] Orcina Ltd, OrcaFlex documentation. https://www.orcina.com/webhelp/OrcaFlex/.

Figure Captions

Table Captions