# Determining Probability of Failure of Structures Using Improved Active Learning Kriging Model with Clustering Algorithm

Pijus Rajak
*Research Scholar, Department of Civil Engineering, National Institute of Technology Durgapur, Durgapur-713209, India.*

Pronab Roy
*Assistant Professor, Department of Civil Engineering, National Institute of Technology Durgapur, Durgapur-713209, India.*

ABSTRACT: A major challenge of structural reliability analysis is estimating the failure probability based on a small number of function calls. This issue can be addressed using metamodeling, which approximates a computationally expensive model with a simpler metamodel, then classical reliability analysis methods can be combined with metamodeling. Kriging models based on active learning are widely used in engineering structural reliability analysis to reduce computational burden. The selection of sampling points has a significant impact on the accuracy and efficiency of metamodels. In most traditional methods of selection, which ignore the location information of Monte Carlo simulation (MCS), experimental samples are selected in unimportant regions. In this regard, unsupervised clustering builds reduced samples from MCS points. To further boost the efficiency of active learning, a sample selection strategy is proposed that finds training samples with high variance and is close to the limit state surface in this paper. Four numerical examples have been solved to evaluate the efficiency of the proposed method and the results are compared with those of MCS and the first-order reliability method (FORM).

## 1. INTRODUCTION

The primary objective of structural reliability analysis is to estimate the failure probability associated with some relevant limit states to assess the effect of uncertainties. Uncertainties can occur depending on the physical properties of the systems, such as material strength, manufacturing tolerances, operating conditions like applied loads, environmental conditions, and incomplete or lacking knowledge. To improve the Kriging model and achieve the desired approximate accuracy, an active learning method is used to sequentially choose training samples. Using a small number of training samples, it is possible to improve prediction accuracy by incorporating nearly informative training samples into the Kriging model using active learning method. The aim of this paper to propose an efficient method to determine structural probability failure using Kriging model with clustering technique.

## 2. METHODOLOGY

### *2.1. Create metamodel*

The first set of training samples represents the input samples of Kriging algorithm that can be obtained from unsupervised clustering (K-Means Clustering) and builds a reduced samples from MCS samples along with the three conditions given in Eq. (1). The elbow method for K-means clustering is used to determine the number $k$ when partitioning randomly generated samples using probability distribution into $k$ clusters. To begin, the algorithm creates a set of $k$ centroids, one for each cluster, using a random selection technique among the samples in the data. To build the groups, each point is assigned to the nearest centroid by Euclidean Distance, then the centroids are recalculated. It is repeated until the centroids are no longer modified by taking the average of the old cluster samples. Samples $X = \{x_1, x_2, x_3, \ldots, x_N\}$ is $V$ dimensional data divided into $k$ known clusters using the k-means algorithm, where $k < N$ then three samples $(P_1, P_2\ and\ P_3)$ in each cluster are selected as follows:

$$P_1 = argmin(D_{c,s}); \quad P_2 = argmin(\mu_{c,s}); \quad P_3 = argmax(D_{c,s}) \tag{1}$$

$$D_{c,s} = \sqrt{(x_{c1} - x_{s1})^2 + (x_{c2} - x_{s2})^2 + \ldots + (x_{cV} - x_{sV})^2} \tag{2}$$

where $D_{c,s} = \{D^1_{c,s}, D^2_{c,s}, D^3_{c,s}, \ldots, D^s_{c,s}\}$ refers to the $s$th Euclidean distance between centroid $(x_c)$ and samples $(x_s)$ within the $k$th cluster, where $s$ is the number of samples. The mean of $D_{c,s}$ is $\mu_{c,s}$ given by Eq. (3).

$$\mu_{c,s} = \frac{1}{s}\sum_{i=1}^{s} D^i_{c,s} \tag{3}$$

Total number of training samples $X_{\text{train}} = 3 \times k$. The response of the structure is determined using LSF. Then performance function is approximated by Kriging model.

The statistical parameters are first defined by the training samples of the Kriging model, after which it can provide estimated responses for any unsampled points. A Kriging model assumes that the substitution function is the result of a Gaussian process denoted by $y(x)$. Kriging model consists of a global tendency and a local deviation, which is expressed in Eq. (4).

$$y(x) = G(x) + z(x) \tag{4}$$

$G(x)$, which is expressed as a linear combination of basic polynomial functions and the $z(x)$ represents the local deviation with mean zero and variance $\sigma^2$. The squared exponential function $R(a, b)$ is considered in the present study as given in Eq. (5) where $a_i$ and $b_i$ are the $i$th coordinate of $a$ and $b$.

$$R(a, b) = exp\left[-\sum_{i=1}^{m} \theta_i |a_i - b_i|^2\right] \tag{5}$$

The correlation between the samples is impacted by theta, an m-dimensional vector, where $m$ is the dimension of $x$. For more details about Kriging model read the reference (Zhang and Quek 2022; Jia and Wu 2022; You et al. 2022).

### *2.2. Updating metamodel in concern domain*

According to the empirical rule in statistics, if $X$ is an observation from a normally distributed random variable and $\mu$ and $\sigma$ are mean and standard deviation respectively, then the mathematical probability function can be written as Eq. (6).

$$Pr(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 95.45\% \tag{6}$$

Identify the concern domain (samples lies between $-2\sigma$ and $+2\sigma$), where the samples at higher risk of misclassification are close to the boundary state, or have a high Kriging variance, or both. In this regard, K-Means clustering, along with the three conditions given in Eq. (1), is used to represent the concern domain and to obtain a second set of training samples. The second set of training samples are added to the first set of training samples and the Kriging model is updated.

### *2.3. Sequentially updating metamodel*

For updating the Kriging model, a point has been selected to close to the LSF in the safe domain $(\bar{y}(x) > 0)$ and a point has been selected to close to the LSF in the unsafe domain $(\bar{y}(x) < 0)$ to avoid misclassification. Two samples ($S^1_{np}$ and $S^2_{np}$) are identified from total samples $(X)$ by using Eq. (7): one sample is selected from safe and another one has been selected from unsafe domain based on the magnitudes of the estimated $\bar{y}(x)$ by the previous updated Kriging model.

$S_{np}^{1} = argmin(\bar{y}(x) = 0)$ in the safe domain. (7)

$S_{np}^{2} = argmin(\bar{y}(x) = 0)$ in the unsafe domain.

The maximin criterion (Johnson et al. 1990) is applied in the safe and unsafe domain sequentially to prevent misclassifications with high variance in the concern domain. Roy and Chakraborty (2022) used the maximin criterion effectively to fill the reduced space. Two samples are selected using Eq. (8), one in the safe and another in the unsafe domain from the samples located in the concern domain $(\mathrm{CR_{pt}})$.

$$\max_{X} \min_{X_{train}, CR_{pt} \in X} d(X_{train}, CR_{pt}) \quad (8)$$

*2.4. Stopping criteria*

A crucial component of the active learning algorithm is the stopping criterion, which effectively ends the Kriging model update and regulates the estimation accuracy of the failure probability. Active learning method can be stopped by using an estimated relative failure probability (Xiang et al. 2020). The stopping criterion of Kriging model is considered based on the estimated relative failure probabilities two consecutives iterations. If the stopping value of three consecutive iterations is less than 0.01 then metamodel is considered as accurate enough and to stop the learning process. If it is not fulfilled, then the training samples will be increased and the metamodel is updated accordingly.

## 3. RESULTS AND DISCUSSION WITH NUMERICAL EXAMPLES

The failure probability $(P_f)$ results obtained using the proposed method is compared to those obtained using direct MCS $(P_{f,MCS})$. In this paper, relative error $(\epsilon_{P_f})$ has been calculated and compared with the results of MCS using Eq. (9).

$$\epsilon_{P_f} = \frac{|P_f - P_{f,MCS}|}{P_{f,MCS}} \times 100\ \% \quad (9)$$

**Example 1:** The schematic diagram of a two-span beam is shown in Figure 1. The length $(L)$ of each span is 6m. The maximum permissible deflection of a two-span beam is $\frac{L}{360}$.

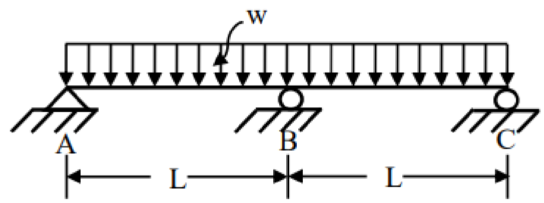


*Figure 1: Schematic diagram of two-span beam.*

LSF of the two-span beam for deflection is given in Eq. (10).

$$g(w, E, I) = \frac{L}{360} - \frac{wL^4}{185EI} \quad (10)$$

Where $w, E$ and $I$ represent the distributed load applied to the beam, the modulus of elasticity of the beam material and the moment of inertia of the beam cross section respectively. Probabilistic characteristics of random variables are given in Table 1. For two-span beam, mean and standard deviation of $E$ and $I$ have been considered as per the literature (Li et al. 2006).

*Table 1: Probabilistic characteristics of random variables of Example 1.*

| Random variable (unit) | Mean | Standard deviation | Distribution type |
|---|---|---|---|
| w (kN/m) | 12 | 0.5 | Normal |
| $E$ $(\mathrm{kN/m^2})$ | $2 \times 10^7$ | $0.5 \times 10^7$ | Normal |
| $I$ $(\mathrm{m^4})$ | $8 \times 10^{-4}$ | $1.5 \times 10^{-4}$ | Normal |

For two-span beam, first iteration is to generate $10^5$ samples randomly using probability distribution. A K-means clustering is used to divide the samples into six clusters, and 18 training samples are selected using Eq. (1) to train the Kriging model. In the second iteration, using the elbow method obtained 18 samples from six clusters, which were then combined with the first iteration training samples to update the Kriging model. Then, in each iteration, four training samples are added sequentially using Eq. (7) and Eq. (8) until the stopping criterion is met. The results show that the failure probability has been converged after six iterations ($N_{call} = 18 + 18 + 4 \times 4 = 52$). The reliability index obtained at the 5th iteration is 2.638 using FORM. The failure probabilities obtained from direct MCS, FORM, and the proposed method are given in Table 2, and relative errors have been determined based on the direct MCS result. To determine failure probability failure using the FORM method, the generalized reliability index and failure probability relationship ($P_f = \Phi(-\beta)$) are considered as a reference (Marelli and Sudret 2018).

*Table 2: Obtained failure probabilities of Example 1.*

| Method | $N_{call}$ | $P_f$ | $\epsilon_{P_f}$ (%) |
|---|---|---|---|
| MCS | $10^6$ | 0.00559 | -- |
| FORM | -- | 0.00417 | 25.40 |
| Proposed method | 52 | 0.00588 | 5.19 |

**Example 2:** A 23-bar truss structure is shown in Figure 2, which has been investigated in (Su et al. 2014; Hong et al. 2022). Probabilistic characteristics of random variables are provided in Table 3. $A_1$ is the cross-sectional area, $E_1$ is the modulus of elasticity of horizontal bars, and $A_2$ is the cross-sectional area, $E_2$ is the modulus of elasticity of diagonal bars. Six external loads are applied to the truss ($P_1$ to $P_6$). LSF is given in Eq. (11), where $V(x)$ denotes vertical displacement at the mid-span.

$$g(x) = 0.11 - V(x) \tag{11}$$

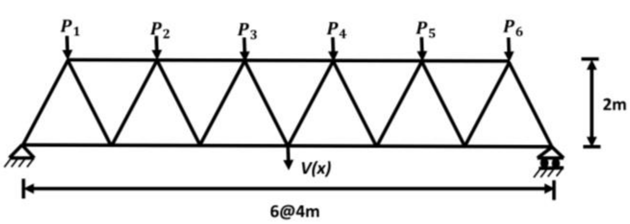


*Figure 2: 23-bar truss structure.*

*Table 3: Probabilistic characteristics of random variables of Example 2.*

| Random variable (unit) | Mean | Standard deviation | Distribution type |
|---|---|---|---|
| $E_1$ (N/m$^2$) | $2.1 \times 10^{11}$ | $2.1 \times 10^{10}$ | Lognormal |
| $E_2$ (N/m$^2$) | $2.1 \times 10^{11}$ | $2.1 \times 10^{10}$ | Lognormal |
| $A_1$(m$^2$) | $2.0 \times 10^{-3}$ | $2.0 \times 10^{-4}$ | Lognormal |
| $A_2$(m$^2$) | $1.0 \times 10^{-3}$ | $1.0 \times 10^{-4}$ | Lognormal |
| $P_1(N)$ | $5.0 \times 10^4$ | $7.5 \times 10^3$ | Gumbel |
| $P_2$(N) | $5.0 \times 10^4$ | $7.5 \times 10^3$ | Gumbel |
| $P_3$(N) | $5.0 \times 10^4$ | $7.5 \times 10^3$ | Gumbel |
| $P_4$(N) | $5.0 \times 10^4$ | $7.5 \times 10^3$ | Gumbel |
| $P_5$(N) | $5.0 \times 10^4$ | $7.5 \times 10^3$ | Gumbel |
| $P_6$(N) | $5.0 \times 10^4$ | $7.5 \times 10^3$ | Gumbel |

Using the method of joints and unit load method, Eq. (11) can be written as Eq. (12).

$$g(E_1, E_2, A_1, A_2, P_1, P_2, P_3, P_4, P_5, P_6) = 0.11 - \frac{(864P_1 + 2400P_2 + 3360P_3 + 3360P_4 + 2400P_5 + 864P_6)}{24A_1E_1} - \frac{2\sqrt{2}(24P_1 + 72P_2 + 120P_3 + 120P_4 + 72P_5 + 24P_6)}{24A_2E_2} \tag{12}$$

From $10^5$ randomly generated samples, 18 samples are obtained in the first iteration by six clustering. Nine samples are added in the second iteration to represent the concern domain. Failure probability has been converged after 17 iterations, and a total of 87 function calls ($N_{call} = 18 + 9 + 4 \times 15 = 87$) are needed to achieve the convergence. The reliability index is 2.574, and it converges in the fourth iteration when using FORM. The failure probabilities obtained from direct MCS, FORM, and the proposed method are given in Table 4, and relative errors have been determined based on direct MCS result.

*Table 4: Obtained failure probabilities of Example 2.*

| Method | $N_{call}$ | $P_f$ | $\epsilon_{P_f}$ (%) |
|---|---|---|---|
| MCS | $10^6$ | 0.00883 | -- |
| FORM | -- | 0.00502 | 43.15 |
| Proposed method | 87 | 0.00885 | 0.23 |

**Example 3:** A benchmark problem of a nonlinear oscillator for the dynamic response (Pan and Dias 2017) is considered for investigation in this paper and schematic diagram is shown in Figure 3.

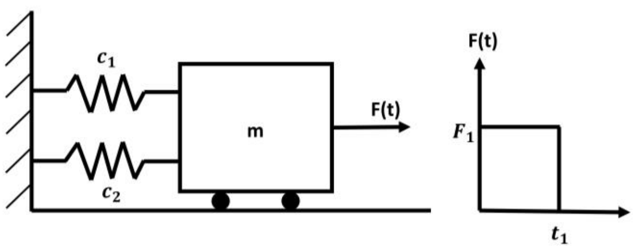


*Figure 3: Nonlinear oscillator.*

LSF is given in Eq. (13).

$$g(m, c_1, c_2, r, F_1, t_1) = 3r - S_{max} \tag{13}$$

$$S_{max} = \begin{cases} \dfrac{2F_1}{m\omega_0^2}; & t_1 \geq \dfrac{\pi}{\omega_0} \\ \dfrac{2F_1}{m\omega_0^2}\sin\left(\dfrac{\omega_0 t_1}{2}\right); & t_1 < \dfrac{\pi}{\omega_0} \end{cases}$$

Where $\omega_0 = \sqrt{\frac{c_1+c_2}{m}}$ , and $S_{max}$ is maximum displacement. $r$ is displacement where one of the springs yields. $m$ stands for the mass and inertial characteristics of the structure, $c_1$ and $c_2$ are its elastic restoring force and potential energy storage respectively, and $F(t)$ is the time-dependent external forces acting on the structure. Probabilistic characteristics of random variables are given in Table 5.

*Table 5: Probabilistic characteristics of random variables of Example 3.*

| Random variable | Mean | Standard deviation | Distribution type |
|---|---|---|---|
| $m$ | 1.0 | 0.05 | Normal |
| $c_1$ | 1.0 | 0.1 | Normal |
| $c_2$ | 0.1 | 0.01 | Normal |
| $r$ | 0.5 | 0.05 | Normal |
| $F_1$ | 1.0 | 0.2 | Normal |
| $t_1$ | 1.0 | 0.2 | Normal |

In the first iteration, the Kriging algorithm uses 18 training samples obtained from $10^5$ random samples using probability distribution, and the LSF is used to determine the response of these input samples. Subsequently, 12 training samples are selected for the second iteration to represent the concern domain. The failure probability converges after seven iterations, and 50 function calls $(18 + 12 + 4 \times 5 = 50)$ are required. Based on FORM, the reliability index obtained at the 5th iteration is 1.897. The failure probabilities obtained from direct MCS, FORM, and the proposed method are tabulated in Table 6.

*Table 6: Obtained failure probabilities of Example 3.*

| Method | $N_{call}$ | $P_f$ | $\epsilon_{P_f}$ (%) |
|---|---|---|---|
| MCS | $10^6$ | 0.0282 | -- |
| FORM | -- | 0.0289 | 2.48 |
| Proposed method | 50 | 0.0279 | 1.06 |

**Example 4:** A two degree of freedom dynamic system (Keshtegar 2016; Zhu et al. 2020) is given in Eq. (14).

$$g = F_s - K_s \times P(E[x_s^2])^{1/2} \tag{14}$$

$F_s =$ force capacity, $K_s =$ the stiffness of second spring, P = the peak factor that is considered as 3; $E[x_s^2] =$ the mean-square relative displacement response, which is given in Eq. (15).

$$E[x_s^2] = \frac{\pi S_0}{4\xi_s\omega_s}\left[\frac{\xi_a\xi_s}{\xi_p\xi_s(4\xi_a^2+\theta^2)+\gamma\xi_a^2} \times \frac{(\xi_p\omega_p^3+\xi_s\omega_s^3)\omega_p}{4\xi_a\omega_a^4}\right] \tag{15}$$

Where, mass ratio $\gamma = \frac{M_s}{M_p}$; average frequency $\omega_a = \frac{\omega_p+\omega_s}{2}$; average damping ratio $\xi_a = \frac{\xi_p+\xi_s}{2}$; tuning parameter $\theta = \frac{\omega_p-\omega_s}{\omega_a}$; $S_0 =$ intensity of the white noise; $K_P = \omega_p^2 M_p$; $K_s = \omega_s^2 M_s$. Table 7 presents the probabilistic characteristics of random variables.

*Table 7: Probabilistic characteristics of random variables of Example 4.*

| Random variable | mean | Standard deviation | Distribution type |
|---|---|---|---|
| $M_p$ | 1 | 0.1 | Lognormal |
| $M_s$ | 0.01 | 0.001 | Lognormal |
| $K_p$ | 1 | 0.2 | Lognormal |
| $K_s$ | 0.01 | 0.002 | Lognormal |
| $\xi_p$ | 0.05 | 0.02 | Lognormal |
| $\xi_s$ | 0.02 | 0.01 | Lognormal |
| $S_0$ | 100 | 10 | Lognormal |
| $F_s$ | 15 | 1.5 | Lognormal |

In this problem, the Kriging model is trained using 60 samples that represent the entire domain in the first iteration. The second iteration uses 21 training samples that are situated in the concern domain to update the Kriging model. 257 function calls $(60 + 21 + 44 \times 4 = 257)$ are required to converge the failure probability after 46 iterations. The failure probabilities obtained from direct MCS, and the proposed method are tabulated in Table 8. According to Zhu et al. (2020), using FORM, the reliability index for this problem is $\beta = 2.016$ Probability of failure is 0.02188) when the number of function calls is 436.

*Table 8: Obtained failure probabilities of Example 4.*

| Methods | $N_{call}$ | $P_f$ | $\in_{P_f}$ |
|---|---|---|---|
| MCS | $10^6$ | 0.00416 | -- |
| Proposed Method | 257 | 0.00415 | 0.24 |

## 4. CONCLUSIONS

The active learning method is used in this study to estimate failure probability accurately with less computational effort. The suggested procedure is divided into these steps: (i) selecting first set of training samples and build a metamodel, (ii) the metamodel is updated adding the second set of training samples which are selected from the concerned domain. (iii) then, four training samples are chosen sequentially in each iteration until the stopping criterion has been met, (iv) the updated metamodel which represents the failure boundary is used to calculate failure probability. To demonstrate the accuracy and efficiency of the active learning method, four numerical examples are presented, and results of the proposed method have been compared with these of MCS and FORM. The proposed method requires lower number of function calls $(N_{call})$ which is quite efficient for complex engineering problems. In each problem, it has been shown to be efficient for estimating failure probability. According to the numerical results, the proposed method predicts failure probability more accurately than the FORM in terms of relative error $(\in_{P_f})$. The high dimensional and low failure probability problem will be focused in the future research work.

## 5. REFERENCES

Hong, L., Li, H. and Fu, J. (2022). “A novel surrogate-model based active learning method for structural reliability analysis.” *Computer Methods in Applied Mechanics and Engineering*, 394, 114835.

Jia, D.W. and Wu, Z.Y. (2022). “An importance sampling reliability method combining Kriging and Gaussian Mixture Model through ring subregion strategy for multiple failure modes.” *Structural and Multidisciplinary Optimization*, 65(2), 1-24.

Johnson, M.E., Moore, L.M. and Ylvisaker, D. (1990). “Minimax and maximin distance designs.” *Journal of Statistical Planning and Inference*, 26(2), 131-148.

Keshtegar, B. (2016). “Chaotic conjugate stability transformation method for structural reliability analysis." *Computer Methods in Applied Mechanics and Engineering*, 310, 866-885.

Li, H.S., Lu, Z.Z. and Yue, Z.F. (2006). "Support vector machine for structural reliability analysis." *Applied Mathematics and Mechanics*, 27(10), 1295-1303.

Marelli, S. and Sudret, B. (2018). "An active-learning algorithm that combines sparse polynomial chaos expansions and bootstrap for structural reliability analysis." *Structural Safety*, 75, 67-74.

Pan, Q. and Dias, D. (2017). "An efficient reliability method combining adaptive support vector machine and Monte Carlo simulation." *Structural Safety*, 67, 85-95.

Roy, A. and Chakraborty, S. (2022). "Reliability analysis of structures by a three-stage sequential sampling based adaptive support vector regression model." *Reliability Engineering & System Safety*, 219, 108260.

Su, G., Yu, B., Xiao, Y. and Yan, L. (2014). "Gaussian process machine-learning method for structural reliability analysis." *Advances in Structural Engineering*, 17(9), 1257-1270.

Xiang, Z., Chen, J., Bao, Y. and Li, H. (2020). "An active learning method combining deep neural network and weighted sampling for structural reliability analysis." *Mechanical Systems and Signal Processing*, 140, 106684.

You, X., Zhang, M., Tang, D. and Niu, Z. (2022). "An active learning method combining adaptive kriging and weighted penalty for structural reliability analysis." *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 236(1), 160-172.

Zhang, X. and Quek, S.T. (2022). "Efficient subset simulation with active learning Kriging model for low failure probability prediction." *Probabilistic Engineering Mechanics*, 68, 103256.

Zhu, S.P., Keshtegar, B., Chakraborty, S. and Trung, N.T. (2020). "Novel probabilistic model for searching most probable point in structural reliability analysis." *Computer Methods in Applied Mechanics and Engineering*, 366, 113027.