

# Reliability Analysis Based on Optimization Random Forest Model and MCMC

Fan Yang<sup>1,2,3,\*</sup> and Jianwei Ren<sup>1,2</sup>

<sup>1</sup>State Key Laboratory of Mechanics and Control of Mechanical Structures, Nanjing University of Aeronautics and Astronautics, Nanjing, 210016, China

<sup>2</sup>College of Aerospace Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, 210016, China

<sup>3</sup>State Key Laboratory for Strength and Vibration of Mechanical Structures, Xi'an Jiaotong University, Xi'an, 710049, China

\*Corresponding Author: Fan Yang. Email: fanyang@nuaa.edu.cn

Received: 22 October 2019; Accepted: 07 August 2020

**Abstract:** Based on the rapid simulation of Markov Chain on samples in failure region, a novel method of reliability analysis combining Monte Carlo Markov Chain (MCMC) with random forest algorithm was proposed. Firstly, a series of samples distributing around limit state function are generated by MCMC. Then, the samples were taken as training data to establish the random forest model. Afterwards, Monte Carlo simulation was used to evaluate the failure probability. Finally, examples demonstrate the proposed method possesses higher computational efficiency and accuracy.

**Keywords:** Markov chain; random forest; Monte Carlo simulation; nonlinear function; random uncertainty

## 1 Introduction

With increasing functionality and complexity of modern products, the structural reliability analysis becomes an emerging research focus in the field of reliability engineering. Failure probability can be expressed as

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

where  $\mathbf{x}$  denotes the random variable in reliability study,  $f(\mathbf{x})$  denotes the combined probability density function,  $g(\mathbf{x})$  is performance function, where the inequality  $g > 0$  represents the safe state and  $g < 0$  represents the failure state. When  $g$  is 0, this state is called limit state which divides the whole domain the safe region and the failure region. The integral denotes the total probability of joint probability density function in failure region. Because  $f(\mathbf{x})$  is usually unknown, it is difficult to calculate the failure probability directly from Eq. (1), especially for high-dimension problems or implicit problems. In recent decades, many methods about solving the integral function have been proposed, including sampling method based on Monte Carlo simulation and approximation method based on performance function, such as the first-order reliability methods (FORM) and the second-order reliability method (SORM) [1–5]. The first-order or second-order reliability



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

method makes a first-order function or second-order function to replace the corresponding limit state function, and the analytical solution requires to solve an optimization function of the shortest distance between limit state and origin in standard normal space [6]. However, the existing approximation methods are unstable if the performance function is nonlinear, or has multiple design points and failure modes [7,8].

Moreover, there are several approximate limit state functions combined surrogate model with Monte Carlo simulation to improve computational efficiency, such as polynomial response surface method (RSM) [9,10], polynomial chaos expansion (PCE) [11,12], Kriging model [13–16], support vector machine (SVM) [17–19], fault tree model [20] and artificial neural network algorithm (ANN) [21,22], etc. By means of surrogate model, the solution of integral function could be obtained with high computational accuracy and little computation cost. It has been become a research focus. Echard et al. [23] adopted the active learning strategy to establish Kriging model adaptively. Compared with conventional sampling methods, smaller samples were used to construct the Kriging model. Ghosh et al. [18] proposed a SVM model to substitute the performance function, and combined with Monte Carlo simulation to calculate the structural safety. However, the inherent parameters of the SVM model is difficult to determine, such as kernel parameters and the penalty factor. Cheng et al. [24] demonstrated a neural network based method combined with the uniform sampling method for predicting failure probability of structures. Oparaji et al. [25] applied the neural network to the reliability and sensitivity analysis of complex nonlinear systems, where the robustness of artificial neural network models was improved by Bayesian theory and model averaging technique. Although ANN has many advantages, including excellent learning capability, fitting for complex nonlinear function with chosen accuracy and so on, ANN also have several disadvantages, such as effect of subjective factors, large generalization errors and low convergence rate. With increasing model complexity, the nonlinear effect of system functions is enhanced.

As a kind of integrated learning algorithm based on decision tree for classification, random forest algorithm has been widely used for addressing issues like data classification and investigation [26]. The advantages of random forest regression model are that it has a better generalization ability and insensitive for abnormal data in samples. Besides, there are other advantages of random forest algorithm, including high accuracy, a fast convergence, less overfitting and so on. In terms of regression problems, the option of hyper-parameters for random forest algorithm has a significant effect on the accuracy of predication, and a further optimization is necessary.

Meanwhile, it requests good samples to construct the surrogate model. The conventional methods usually used the random sampling or design of experiments technique generate samples. For small failure probability event, it will take a lot of samples to establish a high precision surrogate model, because there should be at least one point in the failure domain. Therefore, it is requested an efficient algorithm which can generate samples in both failure region and the safe region.

Recently, the coupling algorithm of Markov Chain Monte Carlo (MCMC) is a powerful tool to simulate samples with random distribution [27,28]. MCMC simulation was adopted to generate samples in failure region. Regarding the generated samples as training data, this brings about an algorithm with higher accuracy [29].

In this study, a combined method called Markov Chain Random Forest (MCRF) algorithm, is adopted to adaptively generate samples. Based on the generated samples, the random forest model is established and improved to determine the best parameters of random forest.

Then, the reliability analysis is performed by Monte Carlo simulation with the random forest model. The work aims to use small number of calling performance function for high precision reliability analysis.

## 2 Methodology

### 2.1 Monte Carlo Simulation

Here, the limit state of system can be expressed by  $g(\mathbf{x})$ , where  $\mathbf{x}$  is variables affecting structural performances. Meanwhile, the joint probability density of each variable is denoted as  $f_x(\mathbf{x})$ , thus the failure probability of the whole system can be represented as,

$$P_f = P(g(\mathbf{x}) \leq 0) = \int_F f_x(\mathbf{x}) d\mathbf{x} \quad (2)$$

where  $F$  denotes the failure domain, namely  $g(\mathbf{x}) \leq 0$ ;  $\int_F f_x(\mathbf{x}) d\mathbf{x}$  is the integration of variables for their joint probability density function on failure domain. Further, an indicator function of failure domain,  $I_F(\mathbf{x})$ , is adopted in this study. If the sample  $x_i$  does not belong to failure domain, there is  $I_F(x_i) = 0$ , else if  $x_i$  belongs to failure domain, there is  $I_F(x_i) = 1$ . So the equation of failure probability can be transformed into Eq. (3),

$$P_f = \int_{R^n} f_x(\mathbf{x}) I_F(\mathbf{x}) d\mathbf{x} \quad (3)$$

Eq. (3) could be solved by transforming it into the mathematical expectation of samples, and the mathematical expectation can be obtained via the mean of samples. According to Law of Large Numbers, the unbiased estimator  $\hat{P}_f$  and its variance  $\text{var}(\hat{P}_f)$  can be calculated by Eqs. (4) and (5), respectively.

$$\hat{P}_f = \frac{1}{N} \sum_{i=1}^N I_F(x_i) \quad (4)$$

$$\text{var}(\hat{P}_f) = \frac{1}{N-1} \left( \frac{1}{N} \sum_{i=1}^N (I_F(x_i))^2 - \hat{P}_f^2 \right) \quad (5)$$

where  $x_i$  is the  $i$ th sample.

### 2.2 Adaptive Sampling Based on Markov Chain

In this section, samples in important region around the limit state are generated by Markov Chain Monte Carlo simulation. Then the generated samples are regarded as training samples which are used as the design of experiments to training random forest model. The random forest model is adopted to approximate the original limit state function. Here, the method for samples generated by Markov Chain is introduced as following.

Step (1): Select the initial state for Markov Chain

Initial variable  $x^{(0)}$  needs to meet  $x^{(0)} \in F$ , namely  $x^{(0)}$  belongs to failure domain, where  $x^{(0)}$  can be determined from engineering or simple computation.

Step (2): Select appropriate stationary distribution function

The stationary distribution function is determined as conditional distribution density function to generate the failure region samples as,

$$f(\mathbf{x} | F) = I_F(\mathbf{x})f_x(\mathbf{x})/P_f \quad (6)$$

As the length of Markov Chain increases, the stationary distribution function ensure that the generated state points tend to fit into the selected target distribution function, namely generating samples in failure region.

Step (3): Select proposal distribution function

The proposal distribution function of Markov Chain directly affects the transforming method from current state to alternate state, which also has an influence on the distance between the candidate state and current state. Based on the basic principle of Markov Chain, proposal distribution adopting the current state  $\mathbf{x}^{(j)}$  as sampling center has a symmetric property. Here, the proposal distribution is a normal distribution, written as,

$$f^*(\boldsymbol{\varepsilon} | \mathbf{x}^{(j)}) = N(\mathbf{x}^{(j)}, V^2) \quad (7)$$

where  $\boldsymbol{\varepsilon}$  denotes the candidate state and  $\mathbf{x}^{(j)}$  represents the current state. The vector  $V$  is the standard variance vector of original joint probability density function, which affects the sampling region of Markov Chain. In order to enhance the convergence of Markov Chain, the gradual changing strategy is adopted. Setting the initial value of  $V$  to be  $V_0$  and taking coefficient  $\gamma$  ( $\gamma < 1$ ) make the standard variance gradually decreasing during the simulation. Those settings can enable Markov Chain simulation to seek its important zone of limit state more quickly, and enable more samples of simulation locating around the important domain and the limit state function.

Step (4): Determine the  $j$ th state  $\mathbf{x}^{(j)}$  of Markov Chain

Based on previous state  $\mathbf{x}^{(j-1)}$ , the  $j$ th state  $\mathbf{x}^{(j)}$  of Markov Chain can be determined by combining its proposal distribution with the Metropolis-Hastings criterion. Firstly, candidate state points  $\boldsymbol{\varepsilon}$  are generated by the proposal distribution  $f^*(\boldsymbol{\varepsilon} | \mathbf{x}^{(j-1)})$ , then the ratio between stationary distribution function  $f(\boldsymbol{\varepsilon} | F)$  of candidate state  $\boldsymbol{\varepsilon}$  and its previous distribution function  $f(\mathbf{x}^{(j-1)} | F)$  can be calculated by Eq. (8),

$$r = \frac{f(\boldsymbol{\varepsilon} | F)}{f(\mathbf{x}^{(j-1)} | F)} = \frac{I_F(\boldsymbol{\varepsilon})f_x(\boldsymbol{\varepsilon})/P_f}{I_F(\mathbf{x}^{(j-1)})f_x(\mathbf{x}^{(j-1)})/P_f} = \frac{I_F(\boldsymbol{\varepsilon})f_x(\boldsymbol{\varepsilon})}{I_F(\mathbf{x}^{(j-1)})f_x(\mathbf{x}^{(j-1)})} \quad (8)$$

According to Metropolis-Hastings criterion, candidate state  $\boldsymbol{\varepsilon}$  is regarded as the  $j$ th state  $\mathbf{x}^{(j)}$  of Markov Chain if  $r \geq 1$ , namely  $\mathbf{x}^{(j)} = \boldsymbol{\varepsilon}$ . If  $r < 1$ ,  $\mathbf{x}^{(j)}$  takes the value  $\boldsymbol{\varepsilon}$  of on the probability of  $r$ , or takes the value of  $\mathbf{x}^{(j-1)}$  on the probability of  $1 - r$ , written as,

$$\mathbf{x}^{(j)} = \begin{cases} \boldsymbol{\varepsilon} & r \geq 1 \\ \boldsymbol{\varepsilon} & \alpha < r < 1 \\ \mathbf{x}^{(j-1)} & r \leq \alpha \end{cases} \quad (9)$$

where  $\alpha$  is a random number subjected to uniform distribution in the interval  $[0, 1]$ .

For reliability analysis, each of the candidate state  $\epsilon$ , whether it is regarded as a Markov Chain state or not, is evaluated by calculating the performance function. Thus, all candidate states  $\epsilon$  are used for training the random forest model, which makes most of valid information during Markov process. The proposed Markov Chain is adaptive to generate samples around the limit state function, so the random forest model has a better approximate accuracy for calculating failure probabilities.

### 2.3 Random Forest Algorithm

Random forest algorithm is a kind of machine learning regression algorithm. Random forest algorithm is a coupling model assembled with  $b$ -regressive trees  $\theta$ , which can be labeled as  $\{h(x, \theta_k), k = 1, 2, \dots, b\}$ . By calculating the average of  $b$ -regressive trees, the prediction of random forest regressive model could be obtained. Once the samples are generated by the MCMC, the random forest model is constructed based on these samples. The establishing process of random forest algorithm for reliability analysis is described as follows.

Step (1): Based on Bootstrap method, abstracting  $b$  training sample groups repeatedly to construct  $b$  regressive trees from  $n$  samples generated by the MCMC. For  $b$  samples that are not selected during each time of abstraction for training samples assemble an out-of-bag (OOB) as a sample group for examination.

Step (2): Construct the regressive tree. From  $k$  variables among sub-branch points of each trees, select randomly  $m_{try}$  ( $m_{try} < k$ ) variables as candidate branch variables, then estimate their optimal branch variables based on the branch-weight criterion.

Step (3): Each regressive tree branches grows unceasingly, recursively from top to bottom. The coefficient  $n_{tree}$ -value of each tree, is regarded as the termination criterion of regressive growth.

Step (4): The produced  $b$  regressive trees construct regressive model of random forest. And the prediction of new model can be evaluated by the prediction accuracy of OOB, which is measured by the variance average of examination set of original data. Assuming the sample number of OOB to be  $m$ , then

$$MSE_{OOB} = \frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{m} \quad (10)$$

$$R_{RF}^2 = 1 - \frac{MSE_{OOB}}{\hat{\sigma}_y^2} \quad (11)$$

where  $y_i$  denotes the real value of dependent variable among OOB,  $\hat{y}_i$  denotes the corresponding predicted value obtained by random forest regressive model,  $\hat{\sigma}_y^2$  denotes the variance of predictions.

For random forest regressive model, the effect of independent variable on dependent variable can be evaluated by variable importance measure (VIM). Here, VIM is an assessment approach based on the mean square error reduction of random permutation and reduction of model accuracy to evaluate the influence of independent variables. The calculation process of mean square error reduction is given as following.

(1) Establish regressive tree for each training sample set, and use the established regressive tree to predict the OOB data. Then, the mean square error of  $b$ -OOB data, labeled as  $MSE_1, MSE_2, \dots, MSE_b$ , can be obtained.

(2) Due to the random selection of variables during constructing of regressive trees branch, variables  $X_i$  of OOB data sets are replaced randomly to assemble a new examination set of OOB data. Using the established random forest regressive model to predict the new examination set, the mean square variances of OOB data can be obtained after replacing randomly. The mean square errors can be assembled as the following matrix,

$$\begin{bmatrix} MSE_{11} & MSE_{12} & \dots & MSE_{1b} \\ MSE_{21} & MSE_{22} & \dots & MSE_{2b} \\ \vdots & \vdots & \ddots & \vdots \\ MSE_{k1} & MSE_{k2} & \dots & MSE_{kb} \end{bmatrix} \quad (12)$$

(3) Substring the corresponding mean error of  $i$ th row for the above matrix from  $MSE_{i1}$ ,  $MSE_{i2}$ , ...,  $MSE_{ib}$ , respectively. Then divide these mean values by their standard deviation, the VIM can be obtained, which is expressed as Eq. (13),

$$VIM_i(MSE) = \left( \frac{1}{b} \sum_{j=1}^b (MSE_j - MSE_{ij}) \right) / S_E, \quad (1 \leq i \leq k) \quad (13)$$

A large value means the independent variable is important for dependent variable. The key idea of the VIM is to apply random disturbance to each variable, and observe the change of accuracy on the established model. Based on the reduction of accuracy, the importance of variable can be evaluated. If the model's accuracy increases after a disturbance of variable decreases, it indicates that this specific variable possesses a higher importance.

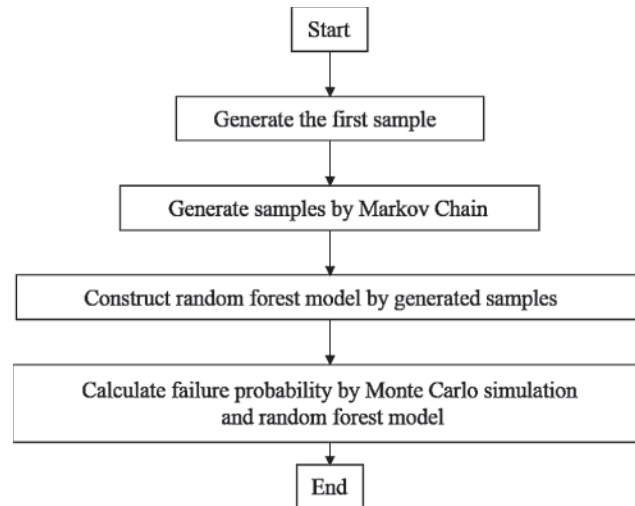
#### 2.4 Improved Random Forest Algorithm

The number of regressive trees  $n_{tree}$  and number of predictors sampled for splitting at each node  $m_{try}$  are important parameters for random forest algorithm. Partial observed values of independent variables are generated randomly by Bootstrap resampling technique during the construction process of random forest regressive tress. Partial explanatory variables selected randomly determine the branch number of regressive trees, which can create hundreds of regressive trees. The data is divided into  $n_{tree}$  training samples, and  $n_{tree}$  decision tree models are trained simultaneously. The splitting number  $m_{try}$  at each node can also affect the accuracy of results. Meanwhile the parameter  $m_{try}$  depends on  $n_{tree}$ , making the regressive model sensitive. Therefore, the sequential quadratic programming (SQP) is adopted to search the best parameter combination of  $n_{tree}$  and  $m_{try}$ . The objective function is the mean square error (MSE) of the random forest model.

#### 2.5 The Integral Process of the MCRF Method

The integral process of the Markov Chain, random forest modeling, and Monte Carlo simulation is shown in Fig. 1. Firstly, based on the engineering experience or numerical computation, an initial point is generated, which is located in failure region. Secondly, the adaptive sampling is performed by the Markov Chain simulation. During iteration, all the points at which the performance function is evaluated, are retained. Then these points obtained by the Markov Chain are used as the design of experiments, and the random forest model is constructed. The parameters

of the random forest are optimized by SQP method. Finally, the failure probability is calculated by the Monte Carlo simulation and random forest.



**Figure 1:** The overall process of the MCRF method

### 3 Model Application

In this section, cases including nonlinear function with cross terms, exponential-nonlinear function, and an engineering problem, sealing ring of aircraft, are employed to verify the effectiveness of the proposed method.

#### 3.1 Case I: Nonlinear Function with Cross Terms

Limit state function with cross terms, as a limit state nonlinear function with cross terms, is shown as Eq. (14),

$$g(\mathbf{x}) = 0.16(x_1 - 1)^3 + x_2 + 4 - 0.04 \cos(x_1 x_2) \quad (14)$$

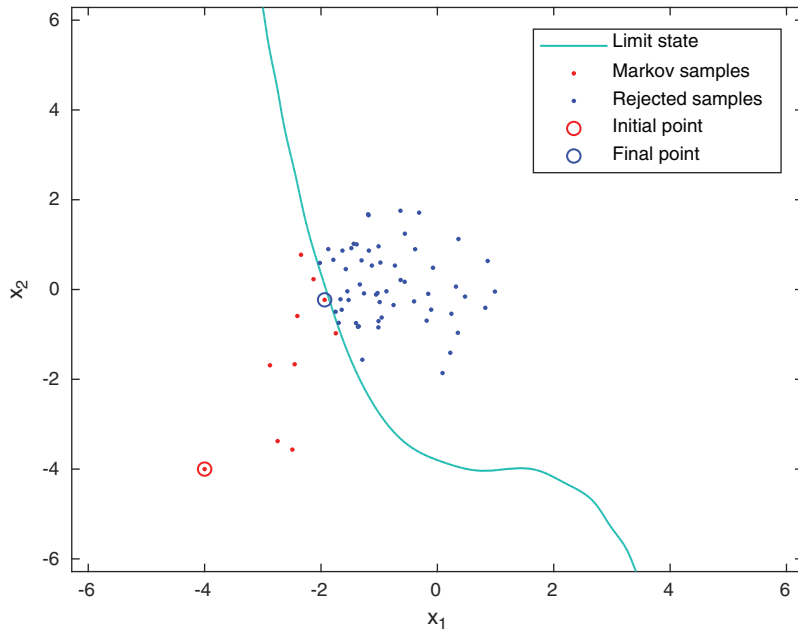
where  $x_1$ ,  $x_2$  denote independent random variable subjected to the standard normal distribution, respectively.

Initial selection point of Markov Chain is  $[-4, -4]$ . Samples are created after 67 Markov Chain simulations, which contain the accepted samples and rejected samples, as shown in Fig. 2.

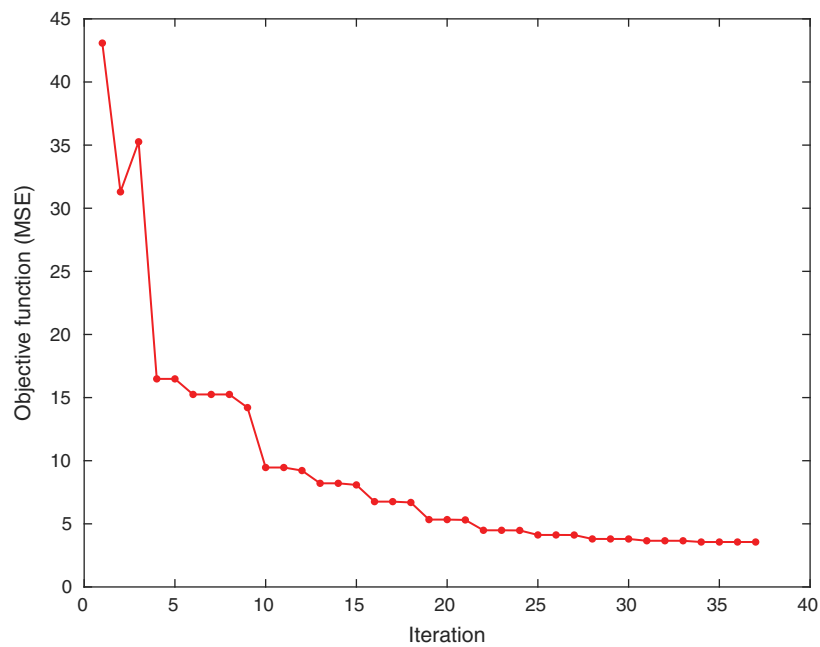
Using the generated samples to construct the random forest model. The random forest model is optimized by SQP. The parameter  $m_{try}$  is set as 39 and the  $n_{tree}$  is set as 500. The iteration of objective function is shown in Fig. 3. As can be seen, the RF model converged in about 37 iterations.

The prediction values and training values of random forest model are shown in Fig. 4. From Fig. 4, these prediction values agree well with that of expectation, indicating the trained random forest model has a higher accuracy to replace their corresponding original performance function for reliability analysis. Based on these prediction values, it is possible to apply random forest model to calculate output values and to obtain their failure probability through Monte Carlo Simulation. Aiming to validate the effectiveness of proposed method, the failure probability is

calculated by crude Monte Carlo Simulation. Then, the first order reliability method (FORM) method, SVM, ANN and subset simulation [30] are listed in Tab. 1, respectively.

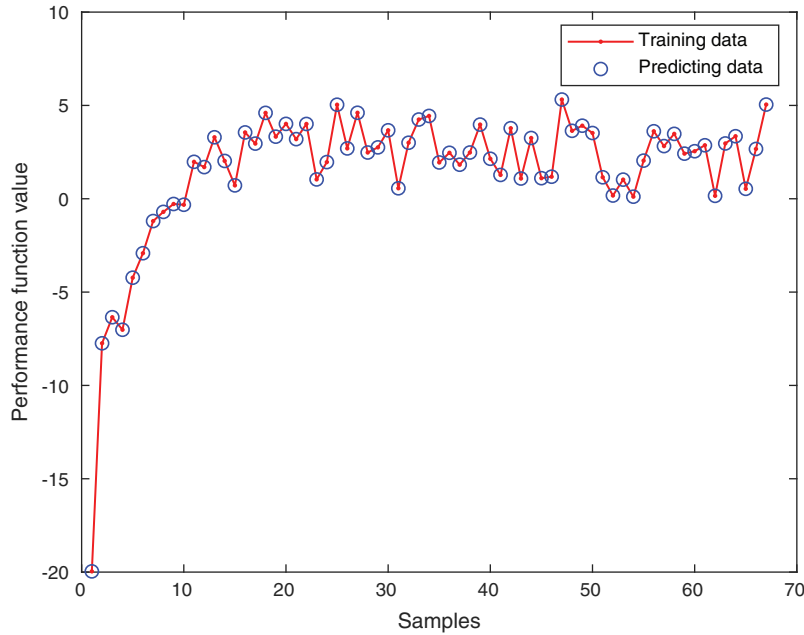


**Figure 2:** Samples generated by MCMC for Case I



**Figure 3:** Convergence history of objective function





**Figure 4:** Comparison between prediction values and training values of random forest model

**Table 1:** Reliability results for Case I

Method	$P_f$	$N_{\text{call}}$	$\beta$	Relative error (%)
MCS	0.0328	$3 \times 10^5$	1.8411	/
Proposed RF	0.0331	67	1.8371	0.22
FORM	0.0316	73	1.8578	0.91
SVM	0.0322	67	1.8494	0.45
ANN	0.0318	67	1.8550	0.75
Subset	0.0337	3800	1.8290	0.66

where,  $\beta$  is reliability index, calculated by  $\beta = -\Phi^{-1}(P_f)$ .

The parameters used in the methods are explained as follows. For SVM, the RBF kernel function is applied, the parameters of the kernel are obtained by the grid searching method, which is 256 (as penalty factor) and 0.0625 (as radius of the kernel function), respectively. For ANN, the architecture is 25-10-5, namely, 25 neurons in the input layers, 10 neurons in the hidden layers and 5 neurons in the output layers. The learning rate is 0.1, and the max iteration number is 200. The activity function is sigmoid function. The standard subset simulation code is adopted here [30], the number of samples in each simulation level is 2000, the conditional probability is 0.1, and the maximum number of simulation levels is 10.

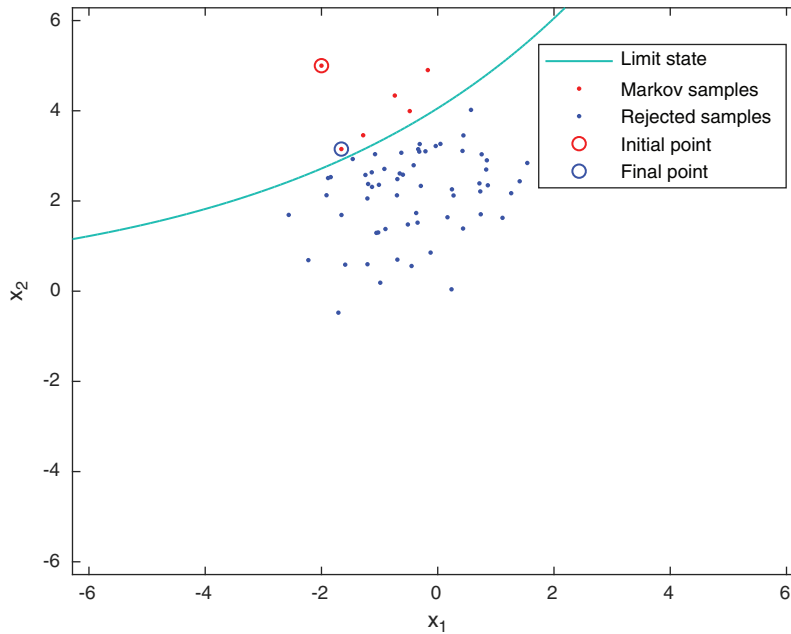
Here, the failure probability is obtained by  $3 \times 10^5$  times of Monte Carlo simulations, which is considered as a benchmark solution. Compared with Monte Carlo simulation, results by the proposed method have a higher accuracy. Using the proposed model, less times of performance functions are conducted. Moreover, it needs not to calculate the design point for reliability analysis by the proposed RF method, which improves the stability of the algorithm.

### 3.2 Case II: Exponent Nonlinear Function

In this section, an exponent nonlinear function is validated [31], which is expressed by Eq. (15) as,

$$g(\mathbf{x}) = \exp(0.2x_1 + 1.4) - x_2 \quad (15)$$

where all basic random variables are independent with each other and obey standard normal distribution. By randomly selecting initial state points of Markov Chain, some samples are generated during the process of Markov Chain, and are shown in Fig. 5.



**Figure 5:** Samples generated by MCMC for Case II

**Table 2:** Reliability results for Case II

Method	$P_f (\times 10^{-4})$	$N_{\text{call}}$	$\beta$	Relative error $\beta$ (%)
MCS	3.59	$10^7$	3.3826	/
Proposed RF	3.58	66	3.3824	5.913e-3
FORM	4.04	77	3.3500	0.96
SVM	2.90	66	3.4408	1.72
ANN	4.70	66	3.3079	2.21
Subset	3.62	7400	3.3803	0.068

From Fig. 5, 66 samples, including 60 rejected samples and 6 accepted samples, were produced in the process of Markov Chain. Their calculations of failure probability are list in Tab. 2. It shows, the failure probability reaches  $3.59 \times 10^{-4}$  after  $10^7$  Monte Carlo Simulations, which can be regarded as the benchmark solution of this problem. The calculation obtained by the proposed model is very similar to that by FORM method. But compared with FORM method,

the proposed method requires less callings of performance function, meaning that the proposed method has a higher computational efficiency than FORM method. Compared with SVM and ANN, the proposed method has much higher precision and can be used in probability analysis. The result by standard subset simulation method is also better than the FORM, SVM, and ANN, however it takes a large amount of calling performance function. The subset simulation combined with surrogate model will be effective to reduce the number of performance calling.

### 3.3 Case III: Sealing Ring

The reliability analysis of sealing is carried out as the last example [32]. The sealing equipment is the aircraft hydraulic system, including an O-ring, a piston rod, a seal groove etc. The simulation model is shown in Fig. 6, where the finite element model is performed in ANSYS. The diameter  $D$ , elastic modulus  $E$ , the amount of compression  $l$  and working oil pressure  $P$  are taken as random parameters, whose distributions are listed in Tab. 3.

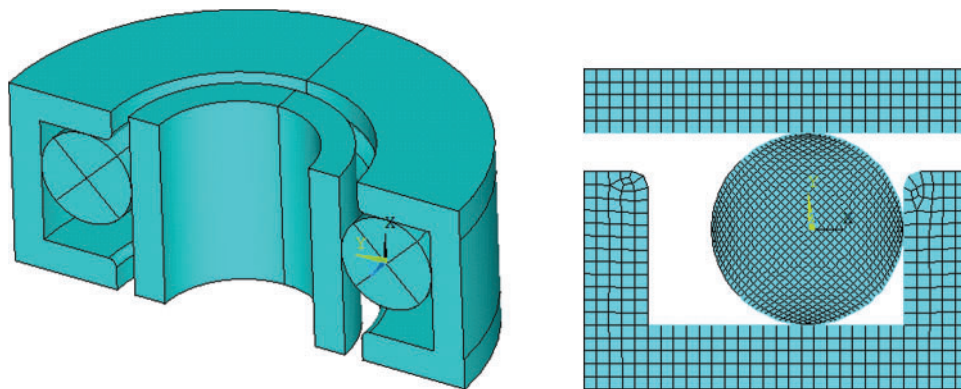


Figure 6: Sealing equipment model

Table 3: Random parameters and distribution

Variables		Distribution type	Means	Variances
Elastic modulus $E$ (MPa)	$x_1$	Normal	14.4	1.189
Diameter of ring $d$ (mm)	$x_2$	Normal	4	0.3304
Amount of compression $l$ (mm)	$x_3$	Normal	0.4	0.03304
Oil pressure $P$ (MPa)	$x_4$	Normal	5	0.413

The failure model of sealing presumes that the maximum contact pressure  $\sigma$  to be greater than the working oil pressure  $P$  between the ring and cylinder. Therefore, the performance function of the sealing can be written as,

$$Z = g_i = \sigma(x_i) - P \quad (i = 1, 2, 3, 4) \tag{16}$$

where  $x_i$  is the random vector for reliability analysis,  $x = (E, d, l, P)$ ;  $\sigma(x_i)$  is the contact pressure, and  $p(x_i)$  is the working oil pressure, where  $P = x_4$ .

The results of failure probability calculated by literature and the proposed method are listed in Tab. 4. As can be seen, the literature uses 350 samples to establish the neural network and

obtain the failure probability, while here only 200 samples are used to construct the random forest model and get the failure probability. The relative error between the proposed method and literature is less than 5%, indicating the proposed method is effective.

**Table 4:** Reliability analysis results of sealing

Method	$P_f (\times 10^{-4})$	$N_{\text{call}}$	$\beta$	Relative error (%)
ANN	2.66	350	3.4641	/
Proposed RF	2.65	268	3.4651	0.029

#### 4 Conclusion

In this paper, a novel reliability analysis method combined with Markov Chain, random forest model and Monte Carlo simulation was proposed. Firstly, a series of samples, including rejected samples and accepted samples, are generated by Markov Chain. Then, based on the samples by Markov Chain, random forest model is established to substitute for the real performance function. Next, according to the distribution of random variables, the reliability analysis was performed by Monte Carlo simulation based on the random forest model. After validation by cases, the proposed random forest model has shown a higher accuracy and efficiency. Compared with Monte Carlo simulations, the proposed model could significantly reduce the calling times of performance functions, while holding a higher calculation accuracy. Compared with the first order reliability method, the proposed method does not require to calculate the gradient of limit state function, which reduces the explicit requirement of limit state function. Compared with SVM, ANN and subset simulation, the proposed method holds a higher precision in terms of failure probability.

The proposed method in this study makes a full use of the adaptability of Markov Chain and the excellent nonlinear fitting ability of random forest model, which reduces the dependence of conventional approximation method on experimental design. The high accuracy prediction of sampling simulation could be realized by means of random forest model, ensuring the calculation accuracy of reliability analysis and enhancing the calculation effectiveness. Application examples of the proposed model shows it is effectiveness for engineering. The proposed method is applicable in low dimension problems as demonstrated in this paper, and it will be a promising method for high dimension problems. In further study, it should be developed for high dimensional problems. Moreover, the parallel computing should be also adopted to further improve the computing efficiency.

**Funding Statement:** This study was supported by the Fundamental Research Funds for the Central Universities (Grant No. NS2020005), Natural Science Foundation of Jiangsu Province (Grant No. BK20190424), Natural Science Foundation of Shaanxi Province (Grant No. 2019JQ-470).

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

#### References

1. Hohenbichler, M., Gollwitzer, S., Kruse, W., Rackwitz, R. (1987). New light on first- and second-order reliability methods. *Structural Safety*, 4(4), 267–284. DOI 10.1016/0167-4730(87)90002-6.

2. Hu, Z., Du, X. (2015). First order reliability method for time-variant problems using series expansions. *Structural and Multidisciplinary Optimization*, 51(1), 1–21. DOI 10.1007/s00158-014-1132-9.
3. Keshtegar, B., Meng, Z. (2017). A hybrid relaxed first-order reliability method for efficient structural reliability analysis. *Structural Safety*, 66, 84–93. DOI 10.1016/j.strusafe.2017.02.005.
4. Meng, Z., Pu, Y., Zhou, H. (2018). Adaptive stability transformation method of chaos control for first order reliability method. *Engineering with Computers*, 34(4), 671–683. DOI 10.1007/s00366-017-0566-2.
5. Huang, X., Li, Y., Zhang, Y., Zhang, X. (2018). A new direct second-order reliability analysis method. *Applied Mathematical Modelling*, 55, 68–80. DOI 10.1016/j.apm.2017.10.026.
6. Keshtegar, B., Chakraborty, S. (2018). A hybrid self-adaptive conjugate first order reliability method for robust structural reliability analysis. *Applied Mathematical Modelling*, 53, 319–332. DOI 10.1016/j.apm.2017.09.017.
7. Chen, Z., Wu, Z., Li, X., Chen, G., Gao, L. et al. (2019). A multiple-design-point approach for reliability-based design optimization. *Engineering Optimization*, 51(5), 875–895. DOI 10.1080/0305215X.2018.1500561.
8. Hadidi, A., Azar, B. F., Rafiee, A. (2017). Efficient response surface method for high-dimensional structural reliability analysis. *Structural Safety*, 68, 15–27. DOI 10.1016/j.strusafe.2017.03.006.
9. Bezerra, M. A., Santelli, R. E., Oliveira, E. P., Villar, L. S., Escalera, L. A. (2008). Response surface methodology (RSM) as a tool for optimization in analytical chemistry. *Talanta*, 76(5), 965–977. DOI 10.1016/j.talanta.2008.05.019.
10. Guimarães, H., Matos, J. C., Henriques, A. A. (2018). An innovative adaptive sparse response surface method for structural reliability analysis. *Structural Safety*, 73, 12–28. DOI 10.1016/j.strusafe.2018.02.001.
11. Zhou, Y., Lu, Z., Cheng, K., Ling, C. (2019). An efficient and robust adaptive sampling method for polynomial chaos expansion in sparse Bayesian learning framework. *Computer Methods in Applied Mechanics and Engineering*, 352, 654–674. DOI 10.1016/j.cma.2019.04.046.
12. Zhou, Y., Lu, Z. (2019). Active polynomial chaos expansion for reliability-based design optimization. *AIAA Journal*, 57(12), 5431–5446. DOI 10.2514/1.J058020.
13. Yang, X., Liu, Y., Gao, Y., Zhang, Y., Gao, Z. (2015). An active learning kriging model for hybrid reliability analysis with both random and interval variables. *Structural and Multidisciplinary Optimization*, 51(5), 1003–1016. DOI 10.1007/s00158-014-1189-5.
14. Xiao, N. C., Zuo, M. J., Zhou, C. (2018). A new adaptive sequential sampling method to construct surrogate models for efficient reliability analysis. *Reliability Engineering & System Safety*, 169, 330–338. DOI 10.1016/j.ress.2017.09.008.
15. Yang, X., Mi, C., Deng, D., Liu, Y. (2019). A system reliability analysis method combining active learning Kriging model with adaptive size of candidate points. *Structural and Multidisciplinary Optimization*, 60(1), 137–150. DOI 10.1007/s00158-019-02205-x.
16. Yun, W., Lu, Z., He, P., Dai, Y., Jiang, X. (2019). An efficient method for estimating the parameter global reliability sensitivity analysis by innovative single-loop process and embedded Kriging model. *Mechanical Systems and Signal Processing*, 133, 106288. DOI 10.1016/j.ymsp.2019.106288.
17. Alibrandi, U., Alani, A. M., Ricciardi, G. (2015). A new sampling strategy for SVM-based response surface for structural reliability analysis. *Probabilistic Engineering Mechanics*, 41, 1–12. DOI 10.1016/j.probenmech.2015.04.001.
18. Ghosh, S., Roy, A., Chakraborty, S. (2018). Support vector regression based metamodeling for seismic reliability analysis of structures. *Applied Mathematical Modelling*, 64, 584–602. DOI 10.1016/j.apm.2018.07.054.
19. Roy, A., Manna, R., Chakraborty, S. (2019). Support vector regression based metamodeling for structural reliability analysis. *Probabilistic Engineering Mechanics*, 55, 78–89. DOI 10.1016/j.probenmech.2018.11.001.
20. Zhang, F., Tan, S., Zhang, L., Wang, Y., Gao, Y. (2019). Fault tree interval analysis of complex systems based on universal grey operation. *Complexity*, 2019(8), 1–8.

21. Lehký, D., Slowik, O., Novák, D. (2018). Reliability-based design: artificial neural networks and double-loop reliability-based optimization approaches. *Advances in Engineering Software*, 117, 123–135. DOI 10.1016/j.advengsoft.2017.06.013.
22. Li, H., He, Y., Nie, X. (2018). Structural reliability calculation method based on the dual neural network and direct integration method. *Neural Computing and Applications*, 29(7), 425–433. DOI 10.1007/s00521-016-2554-7.
23. Echard, B., Gayton, N., Lemaire, M. (2011). AK-MCS: an active learning reliability method combining Kriging and Monte Carlo simulation. *Structural Safety*, 33(2), 145–154. DOI 10.1016/j.strusafe.2011.01.002.
24. Cheng, J., Li, Q. S., Xiao, R. C. (2008). A new artificial neural network-based response surface method for structural reliability analysis. *Probabilistic Engineering Mechanics*, 23(1), 51–63. DOI 10.1016/j.proengmech.2007.10.003.
25. Oparaji, U., Sheu, R. J., Bankhead, M., Austin, J., Patelli, E. (2017). Robust artificial neural network for reliability and sensitivity analyses of complex non-linear systems. *Neural Networks*, 96, 80–90. DOI 10.1016/j.neunet.2017.09.003.
26. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. DOI 10.1023/A:1010933404324.
27. Zhang, F., Xu, X., Cheng, L., Tan, S., Wang, W. et al. (2020). Mechanism reliability and sensitivity analysis method using truncated and correlated normal variables. *Safety Science*, 125, 104615. DOI 10.1016/j.ssci.2020.104615.
28. Lam, H. F., Yang, J. H., Au, S. K. (2018). Markov chain Monte Carlo-based Bayesian method for structural model updating and damage detection. *Structural Control and Health Monitoring*, 25(4), e2140. DOI 10.1002/stc.2140.
29. Wang, Z., Broccardo, M., Song, J. (2019). Hamiltonian Monte Carlo methods for subset simulation in reliability analysis. *Structural Safety*, 76, 51–67. DOI 10.1016/j.strusafe.2018.05.005.
30. Li, H. S., Cao, Z. J. (2016). Matlab codes of subset simulation for reliability analysis and structural optimization. *Structural and Multidisciplinary Optimization*, 54(2), 391–410. DOI 10.1007/s00158-016-1414-5.
31. Zhou, C., Lu, Z., Yuan, X. (2013). Use of relevance vector machine in structural reliability analysis. *Journal of Aircraft*, 50(6), 1726–1733. DOI 10.2514/1.C031950.
32. Zhang, F., Deng, W. W., Zhang, Q., Zhang, L. L., Luo, K. L. (2017). Sealing ring reliability assessment for hydraulic systems based on BP neural networks. *Electrical Engineering and Automation: Proceedings of the International Conference on Electrical Engineering and Automation (EEA2016)*, pp. 702–710. Hong Kong, China.