Structural Reliability Analysis Using Genetic Algorithm and Gaussian Process Regression

Yanjie Xiao, Xun'an Zhang, Feng Yue*

School of Mechanics, Civil Engineering and Architecture, Northwestern Polytechnical University, Shaanxi, 710072, P.R. China.

*Corresponding author’s e-mail: yuefeng@mail.nwpu.edu.cn

Abstract: The implicit and computationally time-consuming performance function limits the application of classical reliability analysis methods in complex structures. To facilitate the reliability calculation of civil engineering structures, a reliability analysis method based on genetic algorithm (GA) and Gaussian process regression (GPR) is proposed in this paper. In this method, GPR is adopted to build the surrogate model of performance function, and GA is used for infill-sampling to improve the model accuracy at the limit state surface. Replacing the actual function with this model in Monte Carlo simulation (MCS), the approximate failure probability can be obtained. Four examples are analysed to validate the efficiency and accuracy of the proposed method. The results show that it can deal with the problems of static reliability and seismic reliability, and can be well combined with structural analysis software, which is convenient for engineering designers to use.

1. Introduction

In the practical use of civil engineering structures, uncertainties generally exist in the external load, component size and material property, etc. Therefore, it is necessary to consider the randomness in the design. However, the implicit performance functions of complex structures may cause analytical reliability methods fail to convergence. MCS method does not rely on the gradient and the reliability results are accurate, but it needs a large number of structural response evaluations. The high computational cost makes MCS unacceptable in practical engineering.

Surrogate models, such as GP, can effectively improve computational efficiency. In these approaches, the approximate function with low computational cost is employed to estimate the results of simulation analysis, which greatly improves the efficiency of reliability analysis. GPR is a model that uses GP for data regression analysis. Compared with other machine learning methods, GP can not only predict the output, but also analyze the uncertainty of the prediction. In recent years, the applications of GPR in civil engineering structural reliability analysis have been investigated[1]. Su et al. proposed a dynamic Gaussian process regression surrogate model based on MCS and calculated the static reliability of Yuzui Yangzze River Bridge[2]. Zhou et al. improved the applicability of AL-GPR-MCS in high-dimensional problems combined with active subspace and calculated the dynamic reliability of frame[3]. Zhong et al. applied GPR to the seismic risk assessment and optimization of cable-stayed bridges[4].

To facilitate the reliability calculation of civil engineering structures, a reliability method based on GPR and GA is proposed in this study. GA is used for infill-sampling, and a surrogate model with high accuracy at the limit state surface can be build. The number of calls to the performance function
in MCS can be greatly reduced with this model. The accuracy, efficiency and engineering applicability of this method are validated by four examples, in which Python is used for programming.

2. Surrogate model

2.1. Gaussian process regression

GPR regards the function $f(x)$ as a random process whose statistical characteristics are completely determined by the mean function $M(x)$ and the covariance function $k(x, x')$. For the input $x$, the observed value $y$ is

$$y(x) = f(x) + \varepsilon$$

where $\varepsilon$ is noise and $\varepsilon \sim \text{N}(0, \sigma_n^2)$. Assuming that there is a training set $T=\{(x^{(i)}, y^{(i)})|i=1,2,\ldots,m\}=(X, Y)$, the prediction $\hat{f}(x_*)$ and the predicted variance $\sigma_f^2(x_*)$ for a given point $x_*$ can be expressed as

$$\hat{f}(x_*)=K(x_*, X)\left(K(X, X)+\sigma_n^2I_m\right)^{-1}y$$

$$\sigma_f^2(x_*)=k(x_*, x_*)-K(x_*, X)\left(K(X, X)+\sigma_n^2I_m\right)^{-1}K(X, x_*)$$

where $K(X, X)=K=[k_{ij}]_{n \times n}$ is the covariance matrix, and $k_{ij}=k(x^{(i)}, x^{(j)})$ is the covariance between $x^{(i)}$ and $x^{(j)}$; $K(X, x_*)=K=[K(x_*, x_*)]^{m \times 1}$ is the $m \times 1$ covariance matrix between $x_*$ and $X$; $I_m$ is the $m$-dimensional identity matrix. The distribution of $f(x_*)$ is as follows:

$$f(x_*) \sim \text{N}(\hat{f}(x_*), \sigma_f^2(x_*))$$

2.2. Basic reliability methods

The reliability of a structure is the probability that the structure can complete the specified function under certain conditions. MCS is a widely applicable reliability analysis method, which is often used to test the results of other methods. The reliability calculated by MCS can be expressed as

$$f \approx \hat{p} = \frac{N_f}{N}$$

where $p_f$ is the failure probability of structure, $\hat{p}$ is the estimated value of failure probability, $N$ is the number of sample points randomly generated according to the statistical parameters of variables, and $N_f$ is the number of sample points falling into the failure domain $F=\{x|G(x)\leq 0\}$. If a GP surrogate model of $G(x)$ is established based on a small number of training sample points, the predicted value $\hat{G}(x)$ and standard deviation $\sigma_G(x)$ of any point can be obtained. With $G(x)$ replaced by $\hat{G}(x)$, the failure probability of problems can be estimated as

$$\hat{p} = \frac{1}{N} \sum_{i=1}^{N} I[\hat{G}(x_i)] = \frac{\hat{N}_f}{N}$$

where $\hat{N}_f$ is the approximation of $N_f$. If $\hat{G}(x) > 0$, $I[\hat{G}(x)]=0$, otherwise $I[\hat{G}(x)]=1$.

3. Proposed structural reliability analysis method

The basic idea of this proposed method is to build the initial GPR model based on a few training points of the performance function, and gradually improve the accuracy of the surrogate model at the limit state surface by genetic algorithm until it meets the engineering needs. Then the failure probability can be calculated by Eq. (6). This method is described in detail below. In order to make the training samples reflect the necessary information, the sampling range of training points $[x_{lb}, x_{ub}]$ should be the region where most Monte Carlo samples can fall into. And the initial sample should be selected by the experimental design method (OLHS is used in this paper) so that the sample points can uniformly distribute in the variable space. In addition, training points still need to be added at key positions for the initial model is generally not accurate enough. It can be seen from Eq. (6) that the accuracy of the reliability results depends on whether the surrogate model can accurately predict the
sign of \(G(x)\). Therefore, the accuracy of the model at the limit state surface is particularly important. According to the characteristics of GPR model, if the prediction error of a certain point which may be near the limit state surface is large, adding the point to the training samples can improve the model accuracy at the surface. Such point can be found by solving the optimization problem expressed as

\[
\max_x \sigma_G(x) \quad \text{s.t.} \quad \hat{G}(x) = 0, \quad x_a \leq x \leq x_b
\]

However, the equality constraint in the optimization problem is not easy to deal with, so it is transformed into the following optimization problem described as

\[
\min_x \frac{\hat{G}(x)}{\sigma_G(x)} \quad \text{s.t.} \quad x_a \leq x \leq x_b
\]

Using genetic algorithm to solve the problem, the optimal solution \(x^*\) can be obtained. If the accuracy of the model at \(x^*\) does not reach the stop condition of infill-sampling, this point is added to the training set. In this way, the model accuracy is gradually improved until the stop condition is met. When the model error of \(x^*\) is less than a certain limit value \(\zeta\), it can be considered that the model does not need to be further improved. Therefore, it is considered as the infill-sampling stop condition that \(\sigma_G(x^*) \leq \zeta/2\) and \(|\hat{G}(x^*) - G(x^*)| \leq \zeta\). \(\zeta\) needs to be a small enough number to ensure the accuracy of reliability result. It can be taken as about 1/20 ~ 1/50 of the standard deviation of the function values of the initial training samples, and can also be adjusted according to the engineering needs.

In summary, the procedure of the proposed reliability method proposed is as follows:

1. Take a small number of training sample points (at least 3 points) by experimental design method, calculate the function value of each point, and select the error limit \(\zeta\). If the variables are normally distributed, the sampling range can be \([\mu_x - 4\sigma_x, \mu_x + 4\sigma_x]\)
2. Construct the GPR model with the training samples
3. Solve the optimization problem in Eq. (8) to get the optimal solution \(x^*\)
4. Calculate the function value of \(x^*\). If \(\sigma_G(x^*) \leq \zeta/2\) and \(|\hat{G}(x^*) - G(x^*)| \leq \zeta\), infill-sampling can be stopped. Otherwise, add \(x^*\) to the training set and return to Step 2
5. Calculate the approximate failure probability by Eq. (6)

4. Test Examples

To verify the effectiveness of the proposed method, four examples are selected. Since the calculation time in practical engineering is mainly spent on the response evaluations, the number of calls to performance function \(N_{\text{call}}\) is used to evaluate the calculation efficiency in this paper.

4.1. Example 1: a truss structure

Figure 1 shows a truss under static force. It is required that the vertical displacement of the midpoint \(D_m\) should not exceed 0.1m, and the performance function can be described as

Figure 1. Truss structure.  Figure 2. Change of failure probability (example 1).
\[ G(E_1, E_2, A_1, A_2, P_1, ..., P_6) = 0.1 - D_m \]  
(9)

where \( E_1 \) and \( E_2 \) are the elastic modulus of the materials, \( A_1 \) and \( A_2 \) are the cross-sectional area of the members, and \( P_1, ..., P_6 \) are the vertical loads. The statistical parameters of these 10 independent variables are listed in Table 1. \( D_m \) is calculated by the software ADINA.

### Table 1. Random variables of example 1.

| Variable \( E_1, E_2/Pa \) \( A_1/m^2 \) \( A_2/m^2 \) \( P_1, ..., P_6/N \) | Distribution type | Mean \( 2.1 \times 10^{11} \) \( 2.0 \times 10^{-3} \) \( 1.0 \times 10^{-3} \) \( 5.0 \times 10^4 \) \( 5.0 \times 10^3 \) | Standard deviation \( 2.1 \times 10^{10} \) \( 2.0 \times 10^{-4} \) \( 1.0 \times 10^{-4} \) \( 5.0 \times 10^3 \) |
|-----------------|-----------------|-----------------|-----------------|-----------------|

MCS, FORM-Kriging, AL-GPR-MCS and the proposed method are respectively used to calculate the failure probability of this example, and the results are listed in Table 2. With the result of MCS \( \hat{P}_f^{MC} \) being the reference value, \( \epsilon_f = |\hat{P}_f - \hat{P}_f^{MC}| / \hat{P}_f^{MC} \) indicates relative the error of \( \hat{P}_f \). The failure probability values corresponding to the number of training points during infill-sampling of the proposed method are plotted in Figure 2. It can be seen that the result of the proposed method is accurate and the calculation efficiency is high.

### Table 2. Reliability results of example 1.

| Method | \( N_{call} \) | \( \beta \) | \( \hat{P}_f /10^{-2} \) | \( \epsilon_f/\% \) |
|--------|----------------|-------------|-----------------|-----------------|
| MCS    | \( 2 \times 10^4 \) | 1.834       | 3.33            | -               |
| FORM-Kriging | 100           | 1.798       | 3.608           | 8.34            |
| AL-GPR-MCS | 88            | 1.834       | 3.33            | 0               |
| Proposed method | 65(3+62)     | 1.834       | 3.33            | 0               |

4.2. Example 2: frame under stochastic seismic excitation

This example is a seismic reliability problem of a three-story frame (Figure 3). The floor mass is \( m_1=3.57, m_2=3.36 \) and \( m_3=3.14(\times 10^5 kg) \), and the interstory stiffness is \( k_1, k_2 \) and \( k_3 \) respectively. The structural damping is Rayleigh damping, and the damping ratio is 0.05. Here, the performance function is constructed according to whether the maximum response of the structure in time history analysis exceeds the limit value[3]. The performance function of the frame can be expressed as

\[ G(k_1, k_2, k_3, \eta) = D_p - D(k_1, k_2, k_3, \eta) \]  
(10)

where \( D \) is the maximum interstory displacement; \( D_p \) denotes the allowable value of the interstory displacement and is taken as 28mm. The dynamic response of the frame is calculated by Newmark-\( \beta \) method, where \( \gamma = 0.5 \) and\( \beta = 0.25 \). EL-Centro wave is used in the time history analysis and its peak value is adjusted to \( \eta \). The statistical parameters of random variables are listed in Table 3.

### Table 3. Random variables of example 2.

| Variable \( k_1, k_2, k_3/N \cdot m^{-1} \) \( \eta/m \cdot s^{-2} \) | Distribution type | Mean \( 8 \times 10^9 \) \( 4 \times 10^6 \) \( 1.4 \) | Standard deviation \( 4 \times 10^6 \) \( 0.28 \) |
The reliability results are listed in Table 4. The result of the proposed method is accurate for this seismic dynamic reliability problem. The failure probability values corresponding to the number of training points in the infill-sampling process are plotted in Figure 4. It can be seen that the proposed method avoids the unnecessary calculation due to the excessive pursuit of accuracy.

| Method         | $N_{\text{call}}$ | $\beta$   | $\hat{P}_f/10^{-2}$ | $\varepsilon_f/\%$ |
|----------------|-------------------|-----------|---------------------|-------------------|
| MCS            | $4 \times 10^4$   | 2.082     | 1.865               | -                 |
| FORM-Kriging   | 100               | 2.102     | 1.779               | 4.60              |
| AL-GPR-MCS     | 257               | 2.082     | 1.865               | 0                 |
| Proposed method| 71(3+68)          | 2.081     | 1.8725              | 0.40              |

4.3. Example 3: series system with four branches

This problem has multiple most probable points (MPP). Its performance function is

$$G(x_1, x_2) = \min \left\{ \frac{3 + (x_1 - x_2)^2}{10} - \frac{(x_1 + x_2)}{\sqrt{2}}, \frac{3 + (x_1 - x_2)^2}{10} + \frac{(x_1 + x_2)}{\sqrt{2}} \right\}$$

(11)

where $x_1$ and $x_2$ are independent and follow the standard normal distribution. The results obtained by different methods are listed in Table 5. The error limit $\zeta$ is taken as 0.04. It can be seen from the table that the results of the proposed method are still accurate and the $N_{\text{call}}$ values are small. This example demonstrates that the proposed method is still efficient and accurate in the multi-MPP problem.

| Method         | $N_{\text{call}}$ | $\beta$   | $\hat{P}_f/10^{-3}$ | $\varepsilon_f/\%$ |
|----------------|-------------------|-----------|---------------------|-------------------|
| MCS            | $5 \times 10^5$   | 2.842     | 2.244               | -                 |
| FORM           | 9                 | 3         | 1.350               | 39.84             |
| FORM-Kriging   | 100               | 3.155     | 0.804               | 64.19             |
| AL-GPR-MCS     | 64                | 2.842     | 2.242               | 0.09              |
| Proposed method| 44(3+41)          | 2.841     | 2.246               | 0.09              |
|                | 40(11+29)         | 2.841     | 2.252               | 0.36              |

4.4. Example 4: seismic reliability analysis of MSCS

Mega-sub controlled structure (MSCS) is a new structure formed by combining mega frame structure (MFS) with tuned substructure control principle. Part of the connection between the main structure and the substructure is removed in this structure, and damping devices are installed between them[5]. The response of the structure under earthquake is reduced by utilizing the relative motion of the main and sub-structures, thus improving the seismic performance.

![Figure 5. Finite element model of MSCS.](image)

![Figure 6. Acceleration time history.](image)
Figure 5 shows the finite element model of MSCS. The storey height is 4m. I-beams with a cross section of 4m×4m×0.04m×0.04m are adopted for the mega beams. The mega column is made of square steel with a wall thickness of 0.05m, and the distance of the column center is 26m. The section of the substructure column is 0.6m×0.6m×0.02m square steel, and the section of the substructure beam is 0.5m×0.2m×0.02m×0.02m I-steel. The main structure consists of three mega storeys, and viscous dampers connected with substructure are installed on the second and third mega storeys. The nonlinear force of the damper is $f_d = c_d \nu$, where $\nu$ is the deformation rate of the damper, the damping coefficient $c_d$ is 1.5×10^6N/(m·s⁻¹), and the damping index $e$ is 0.4.

The performance function is constructed in the same way as in example 2, which is described as

$$G(\eta, E, b_1, b_2, b_3) = \phi_b - \varphi_m(\eta, E, b_1, b_2, b_3)$$

where $b_1, b_2, b_3$ are the column widths of the first, second, and third mega storey, $E$ is the elastic modulus of steel; $\varphi_m$ is the maximum interstory drift angle obtained by nonlinear dynamic time history analysis; the deformation limit $\varphi_b$ is taken as 1/250. The seismic waves used in the analysis are shown in Figure 6. The statistical parameters of random variables are listed in Table 6.

### Table 6. Random variables of MSCS.

| Variable  | Distribution type | Mean | Standard deviation |
|-----------|-------------------|------|--------------------|
| $\eta/\text{m/s}^2$ | Extremum type II | 2.33 | 2.91 |
| $E/\text{GPa}$ | Normal | 206 | 20.6 |
| $b_1, b_2, b_3/\text{m}$ | Normal | 4.4 | 0.22 |

The reliability results are listed in Table 7. The reliability result obtained by this method is very close to that obtained by AL-GPR-MCS, but the $N_{\text{call}}$ value is less than 1/3 of the latter. It can be seen that the proposed method is suitable for seismic reliability analysis of engineering structures with complex implicit performance function.

### Table 7. Reliability results of example 4.

| Method            | $N_{\text{call}}$ | $\beta$ | $\hat{P}_f/10^{-2}$ |
|-------------------|-------------------|---------|---------------------|
| AL-GPR-MCS        | 230               | 0.993   | 16.08               |
| Proposed method   | 64                | 0.989   | 16.12               |

5. Conclusions

A structural reliability analysis method based on GPR and GA is proposed in this paper. It uses genetic algorithm for infill-sampling to make the GPR model accurate at the limit state surface. The number of calls to the performance function in MCS can be significantly reduced with this model, and the result is close to that of MCS. Four examples are analyzed to validate the efficiency and accuracy of the proposed method, and the results show that it can deal with static reliability, seismic reliability and multi-MPP problems. This method does not need to transform random variables into standard normal space, and can be well combined with finite element software, which is convenient for engineering designers to use.

References

[1] Jia, B., Yu, X., Yan, Q. (2017). A new sampling strategy for Kriging-based response surface method and its application in structural reliability. Adv. Struct. Eng., 20: 564-581.
[2] Su, G., Peng, L., Hu, L. (2017). A Gaussian process-based dynamic surrogate model for complex engineering structural reliability analysis. Struct. Saf., 68: 97-109.
[3] Zhou, T., Peng, Y. (2020). Structural reliability analysis via dimension reduction, adaptive sampling, and Monte Carlo simulation. Struct. Multidiscip. O., 62: 2629-2651.
[4] Zhong, J., Wan, H. P., Yuan, W., He, M., Ren, W. X. (2020). Risk-informed sensitivity analysis and optimization of seismic mitigation strategy using Gaussian process surrogate model. Soil Dyn. Earthq. Eng., 138: 106284.
[5] Qin, X., Sheldon, C. (2008). Study on semi-active control of mega-sub controlled structure by MR damper subject to random wind loads. Earthq. Eng. Eng. Vib., 7: 285.