A comparative study of Monte Carlo Simulation and M5Tree method on reliability analysis of truss structure

Junhan Liao*

Department of Guangdong University of Technology, Guangzhou, China

*Corresponding author: 3120003161@mail2.gdut.edu.cn

Abstract. Accurate failure probability is essential in structural reliability analysis. Plenty of methods utilized to calculate the failure probability are proposed. Monte Carlo Simulation (MCS) is one of the typical methods for solving this kind of engineering problems. In recent years, M5Tree algorithm has begun to be applied in this domain and shown great potential. The paper expounds the principles and the truss-reliability analysis process of both MCS and M5Tree through two truss examples. While evaluating the reliability using MCS, the random variables from uncertain aspects are simulated to gain the failure probability directly. In the M5Tree+MCS method, input datasets are used to generate the performance function to obtain the probability. The results indicate that MCS has great operability and practicability in dealing with general engineering problems while its computational efficiency is unsatisfactory compared to M5Tree algorithm +MCS. Overfitting in M5Tree algorithm always happens and affects the analysis on testing data, a defect of M5Tree. The accuracy of failure probability and computational efficiency are the main criterions for judging.

Keywords: Monte Carlo Simulation, structural reliability analysis, M5Tree algorithm, Truss.

1. Introduction

Truss, the common structural element, is widely used in large-span structures such as warehouses, bridges, and stadiums in many industrial and civil buildings. Truss bars are mostly made of steel and bear axial tension or compression most of the time. Structure reliability analysis concerns about personal security, economic performance and other significant aspects, which signifies we cannot take it for granted.

Monte Carlo Simulation (MCS) and M5Tree methods can be applied to reliability analysis of truss structure. MCS is a popular approach in structural reliability analysis, while M5Tree is mainly used as a method for prediction in other domains. The evaluation of failure probability is the main effort in engineering problems. The limit state functions (LSF) can divide the design space of random variables into two area. The positive value of LSF corresponds to safety domain while the negative one corresponds to failure domain [1]. And the generated sample points by MCS in the space are also separated into two parts which provides a way to calculate the failure probability. M5Tree algorithm now is gradually adopted in predicting the failure probability. This method combined with MCS is more efficient, and the results are as highly accurate as the MCS's ones. M5Tree algorithm generates a model tree with substantial terminal (leaf) nodes. The regression fitting equation is the core of the M5Tree algorithm, which is acquired through the fitting process for sample points in their terminal nodes. Random variables of truss are put into the design area and evaluated by these performance functions to gain the failure probability.

In this study, two approaches are applied to analyze truss through two truss examples and compared to each other, aiming to find their discrepancy, merits and demerits.

2. Monte Carlo simulation

2.1. Methodology principle and process

MCS is widely used in different domains, especially structural reliability analysis. The random variables generated form material properties, geometric properties and the variation of loads and temperature cause the uncertainty of failure probability which is the main goal of structure analysis. The LSF used to gain the probability of failure can be defined as follows [2]:

$$Z=g(X)=r(X)-s(X) \tag{1}$$

Where r and s are mutually independent, in terms of random variable X. When g(X)>0, the structure can be considered safe. While r(X) equals s(X), which means g(X) equals 0, the structure is in a limited state. The structure would fail when the function g(X)<0. In general, the probability of failure can be defined as follows [3]:

$$P_f = \int_{g(X)\leq 0}^{n} \dots \int f_X(x_1, \dots, x_n) dx_1 \dots dx_n, \tag{2}$$

The failure probability can be calculated by integration. However, most of the LSF in the reliability prediction cannot be written in the form of explicit function due to complexity of structure, variation of load form and other reasons. To approximate the failure probability, especially in a structured system with random variables, MCS based on Statistical Sampling Theory is the appropriate approach to deal with it.

Numerous values of random variables can be simulated through MCS. The sampling results are used to calculate the failure probability by the equation as follows [2]:

$$P_f = \frac{1}{N}\sum_{i=1}^{N} I[g(X)] \tag{3}$$

Where I[g(X)] is the failure indicator. It equals one when $g(X) \leq 0$. Otherwise, it equals zero. N is the sample size which represents the times of MCS. To gain the precise failure probability, the times of MCS is usually massive, up to 1 million. MCS can be conducted by most software, such as MATLAB, R, and Python. The distribution type of random variables and their mean and standard deviation are needed when doing MCS. Large amounts of functions of distribution type are recorded in this software as follows (Table 1).

Table 1. The functions of distribution type in R and MATLAB

Distribution Type	Functions in R	Functions in MATLAB
Binominal	binorm	bino
Poisson	pois	Poiss
Geometric	geom	Geo
Hypergeometric	hyper	hyge
Uniform	unif	Unif
Exponential	exp	exp
Normal	norm	norm
Weibull	weibull	wbl
Gamma	gamma	gam
Beta	beta	beta

Generally, the procedure of MCS can be concluded as follows: Start with setting up the distribution of random variables and decision function based on the structural constitution in the programming software. And then, determine the simulation times of MCS and begin sampling. The failure probability can be calculated using Eq.(3) according to statistical sampling results.

2.2. Case study of MCS

The top of the warehouse's span equals 18 meters and is a truss system constructed of 31 steel components. The roof connects the truss system with pin joints. The snow load applies on the roof while the wind applies on the left column. Two concentrated load P applies respectively on the outrigger of the columns. The loads applied to this structure are shown in Figure 1. The 3D view of the roof is demonstrated in Figure 2a. And the cross section is demonstrated in Figure 2b. The parameters of the components, such as cross section, and size, are shown in Table 2. Figure 3 illustrates the cross-section of the member bar.

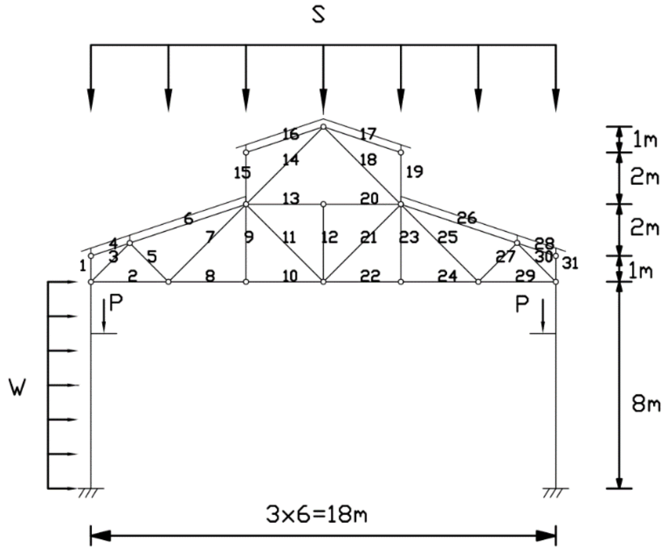

Fig. 1 The size and loads

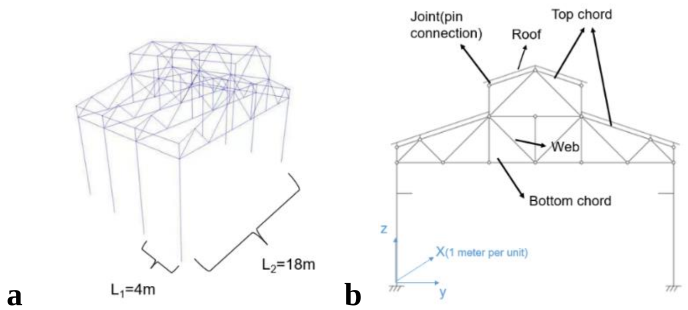

Fig. 2 (a) 3D view of warehouse; (b) Cross-section of warehouse

Table 2. The statistical parameter of member bar

Material	H*B	t1	t2	r	Section area(cm^2)
A283GRC	175*175	7.5	11	13	51.43

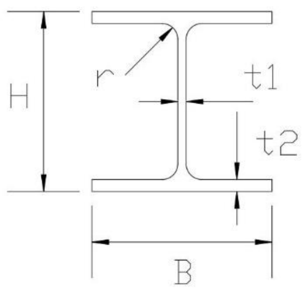

Fig. 3 The cross section of the bar

The dead loads include concentrated load *P*, assumed as 1500.8kN and the weight of the roof, assumed as $0.5kN/m^2$. The distribution type of loads and resistance and their mean and standard deviation are as follows in Table 3.

Table 3. Distribution type together with mean and standard deviation

Load/Resistance	Distribution type	mean	Standard deviation
Snow load	Gumbel	0.4112	0.0800
Wind load	Gumbel	0.5060	0.0835
Bar resistance	Log-normal	6.8475	0.0098

The unit method is used to calculate each component's internal forces under the variable loads. The loads applied to the structure are assumed as one kN/m^2, respectively. In this situation, the internal forces of each component can be regarded as coefficient k_i. The actual value of the internal force of each component is calculated using Eq.(4). The calculation results are listed in Table 4:

$$F_N = k_1P + k_2Sq_1 + k_3Sq_2, \tag{4}$$

Table 4. The coefficients of internal force

Bar	k1	k2	k3	Bar	k1	k2	k3
1	-0.01	-1.22	0.09	17	0	-0.03	-0.83
2	-0.02	-0.34	7.98	18	0	0.09	-2.24
3	0	-0.13	-10.57	19	0	-0.02	-1.37
4	-0.05	0.22	-1.09	20	-0.02	-0.03	-9.42
5	-0.02	0.18	2.53	21	0	-0.10	-0.36
6	-0.04	0.06	-11.19	22	-0.05	0.04	11.52
7	0.02	-0.19	-2.34	23	0	0	-0.02
8	-0.05	-0.10	11.76	24	-0.05	0.04	11.76
9	0	0	-0.02	25	0.02	0.22	-2.34
10	-0.05	-0.10	11.52	26	-0.04	-0.17	-11.19
11	0	0.09	-0.36	27	-0.02	-0.16	2.53
12	0	0	0.47	28	-0.05	-0.31	-1.09
13	-0.02	-0.03	-9.42	29	-0.02	0.33	7.98
14	0	-0.09	-2.24	30	0	0.10	-10.57
15	0	0.01	-1.37	31	-0.01	-0.12	-1.22
16	0	0.03	-0.83				

If F_N, the axial force of each bar, is bigger than the resistance of the components, the components would fail. Negative forces represent axial pressure otherwise are axial tension. Therefore, this condition can be used as a decision function in MCS. The Uniform load q_1 is the weight of roof equal to 0.5 kN/m^2 while the other one q_2 is snow load. S represents the projected area of the roof which equals 18m^2. Compare the values of internal force generated by random sampling according to the distribution of loads with the values of resistance also attained by random sampling. Circulate the program 1 million times and the failure probability of each bar can be obtained (Table 5).

Table 5. The failure probability of each bar

Bar	Failure probability
6	1.9×10-5
26	3.3×10-5
others	0

The top chards 6 and 26 are more likely to fail than any other bars. If any of bar 6 and bar 26 fail in the truss system, the whole system turns out to be geometrically unstable, which means the structure would collapse. Given that it can be regarded as a chain system made up of bar 6 and 26 and several other main parts, the failure probability of the truss is figured out, which equals 5.2×10^{-5}.

3. M5 model tree

3.1. Methodology principle and process

M5tree algorithm is a first-order linear model tree algorithm with high accuracy built after the proposed regression tree [4]. The algorithm splits the parameter space into different rectangles whose margins are parallel to each other based on the principle of sample attribute differentiation. Each space will continue to split into subspaces until the standard deviation of the class values is smaller than a certain threshold or the number of the samples in their area is less than a certain value. There will be a local regression model in different subspaces [4].

The division criterion is the principle of sample attribute differentiation which is characterized as follows [4]:

$$\text{SDR}= sd(T)\text{-}\sum_i \frac{|T_i|}{|T|} sd(T_i) \quad (5)$$

Where SDR is standard deviation reduction; T represents the set of training examples that reaches the node; T_i means the ith subset divided from collection T; sd(T_i) is the standard deviation of collection T_i. The subspace is the node. The splitting procedure amounts to the growth of the model tree and classifies all the training examples. Then the original model tree is generated.

To improve the model's efficiency, prune the branches for the overgrown model tree, which means incorporating some nodes and replacing them with terminal (leaf) nodes, is necessary. While examining nodes, utilizing the linear regression method to figure out the system of nonlinear multivariable equations is the first step. The criterion to judge whether the nodes should be preserved or replaced by terminal nodes is based on the reduction of predicted error. The reduction of predicted error can be calculated using the following equation [5]:

$$E_R = |N|R_{MSE} - |N_l|R_{MSEl} - |N_r|R_{MSEr} \quad (6)$$

Where RMSE signifies the root mean square error that is predicted by the local fitting equation in a node (including all the subordinate subtree and leaf); R_{MSEl} represents the root mean squared error of the left terminal node while R_{MSEr} represents the root mean squared error of the right terminal node; The subtree would be preserved when the value of ER is positive. Otherwise, it would be replaced by a terminal node.

After the pruning process, the original model tree has turned out to be neat and efficient.

However, there probably remains some discontinuity between two linear models of adjacent terminal nodes. It may result in a nonlinear part of the two models so that the accuracy of the prediction would be influenced. To cope with this issue, the smoothing process is to merge each multivariate linear fitting equation of child nodes with the multivariate linear fitting equation of their parent nodes into a new linear equation. The equation shown as follows works for it.

$$f_{new} = \frac{nf_{child}+kf_{parent}}{n+k} \ , \quad (7)$$

Where f_{parent} means the fitting equation of the parent node;f_{child} signifies the fitting equation of child node; f_{new} is the new linear equation that merged by the first two equations; n is the number of data that reaches the terminal node; the constant k is usually determined as 15. It is only when the d-value of R_{MSE} between f_{parent} and f_{child} is smaller than a certain threshold, the smoothing process should be conducted [5].

The M5tree algorithm applied in structural reliability analysis combined with MCS can be generally summarized in follows steps [3]:

Construct the M5tree model. The input data works as a training dataset and is generated randomly from the space of each random variable. And the space is divided into several subspaces, which means the M5tree is calibrated. Meanwhile, the performance function is also generated in each terminal node.

Generate Monte Carlo samples. The sample points of random variables worked as a testing dataset are generated based on their distribution, means and standard deviation.

Calculate the failure probability. The M5tree model would evaluate the sample points according to the performance functions. The failed sample points are labeled as 1, and others are 0. The approximation of failure probability is figured out using Eq.(3).

Compute the coefficient of variation. The coefficient of variation is an indicator to determine whether the sample size is adequate for estimating the failure probability. The generated samples are regarded as being enough only when the value of the below equation is less than 5%.

$$CoV^{M5Tree} = \sqrt{\frac{1-P_f^{M5Tree}}{(NS-1)P_f^{M5Tree}}}. \quad (8)$$

3.2. Case study of M5tree model

In this case, a uniform load is applied on a roof truss of which compression components are made of steel reinforced concrete while the tension members are made of steel. The perpendicular deflection of the vertex point C (Δc) can be computed using the below equation:

$$g = 0.03 - \left(\frac{ql^2}{2}\right)\left(\frac{3.81}{AcEc} + \frac{1.13}{AsEs}\right) \quad . \quad (9)$$

The schematic view of the roof truss is shown in Figure 4. The statistical parameters of random variables are listed in Table.6.

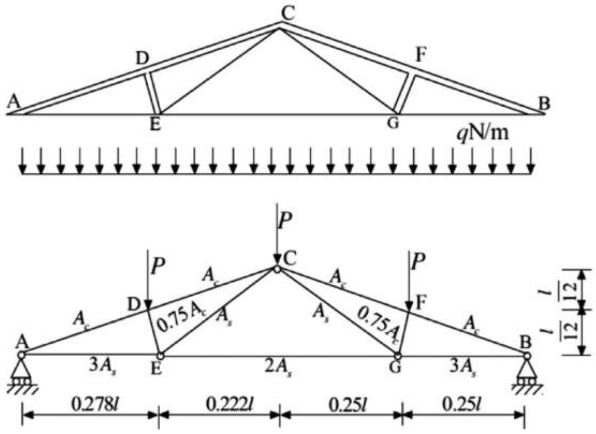

Fig. 4 Schematic view of roof truss example [3]

Table 6. The statistical parameters of basic random variables [3]

Random variables	q(N/m)	l(m)	As(m^2)	Ac(m^2)	Es(Pa)	Ec(Pa)
Mean	20.000	12	9.82×10^{-4}	0.04	1×10^{11}	2×10^{10}
Standard deviation	1400	0.12	5.9852×10^{-5}	0.0048	6×10^{9}	1.2×10^{9}

Table 7. The results of MCS and M5Tree+MCS [3]

Method	N_{call}	P_f	β	$Abs.error(\beta)\%$
MCS	9.5×10^{5}	0.00949	2.3460	-
	100	0.00883	2.3727	1.112
	200	0.00930	2.3536	0.32
M5Tree+MCS	300	0.00957	2.3426	-0.15
	400	0.00951	2.3452	-0.04
	500	0.00953	2.3345	-0.07

As and Ac signify the cross-section of steel bars and reinforced concrete. Likewise, Es and Ec mean their elastic modulus. *q* is the uniform load. *l* is the span of the roof truss. The reliability analysis results shown in Table 7 are obtained through MCS and M5Tree combined with MCS. N_{call} represents the number of calls, which means the time evaluating the LSF to get accurate results. To compare the results of these two methods, the reliability index β of MCS is set as a nominal value for determining the $Abs.error(\beta)\%$ based on the below equation [3]:

$$Abs.error(\beta)\% = \frac{\beta-\beta_{MCS}}{\beta} \times 100 \quad . \quad (10)$$

It is manifest that the MCS method needs almost one million calls to gain an accurate failure probability, while the N_{call} of M5Tree+MCS is less than a thousand. When the figure of M5Tree comes to 500, the failure probability is almost the same as the one conducted by MCS with merely -0.07% error.

4. Discussion

4.1. Pros and cons of MCS

MCS is widely used in most domains, especially in reliability analysis, because of its operability and practicability. It can efficiently tackle engineering problems involving randomicity and uncertainty according to the known distributions, means and standard deviations of random variables. The simulation can be conducted million times in a short period.

Some defects can also be seen on MCS. Given that the random variables generated by MCS are Pseudo-random Numbers, which are not truly random, the MCS results may have a slight discrepancy compared with real situations. Another drawback is the calculated quantity. Million times of simulation is conducted to approximate the accurate results, which is computationally inefficient.

4.2. Pros and cons of the M5tree

M5Tree algorithm has the advantage in tackling Muti-dimensional problems. The model tree can grow well even if the training data is not much. Due to the linear regression equations on terminal nodes, the prediction of results obtained through less than a thousand calls is entirely accurate and uses fewer sample points than MCS.

Though some advantages are seen in M5Tree algorithm, the defects cannot be ignored. The way M5Tree algorithm generates a model is pretty easy by using a little training data. That also would explain why it often generates a complex model tree and the over-fitting problem emerges. To put it another way, the results figured by M5Tree are excessively accurate, so the tree cannot analyze the testing data reasonably. The model tree might be unstable. Attributed to slight data variance, the generated model tree is completely different from one another.

4.3. Optimization methods

In terms of computational efficiency in MCS, Latin Hypercube Sampling working based on MCS is an approach to improve it. In most of the situations, the value of random variables sampling by MCS are concentrated in a certain design space while the margin of the space is not sampled. Latin Hypercube Sampling divides the design space of random variable X into several subspace with equal length and samples each space to ensure the margin space can be sampled. Therefore, the results of MCS with Latin Hypercube Sampling are more representative. And the sample size of MCS with Latin Hypercube Sampling minimizes to a quarter of the origin [6,7,8].

Overfitting is one of the trickiest problems in M5Tree. Pruning the tree can solve this issue by preventing the model tree to develop into a complicated structure. There are two kinds of popular pruning methods called Reduced-Error Pruning (REP) and Pessimistic-Error Pruning (PEP). Pay attention to the side effect of pruning which it might also has an influence on the accuracy of the results [9,10].

5. Conclusion

The paper expounds the principles and the truss-reliability analysis process of both MCS and M5Tree through two truss examples.

Monte Carlo Simulation, a traditional method of random variables sampling is widely applied in structural reliability analysis. M5Tree algorithm also shows its potential in this domain. The MCS calculates the failure probability based on large sampling of random variables of all uncertain factors such as material properties, geometric properties, etc. The failure probability equals the numbers of failure issues that evaluated by the limit state function divided by the sample size. The procedure of M5Tree+MCS is to evaluate the random variables generated by MCS according to the Tree model which is generated form training data, and then obtain the failure probability based on the numbers of failure probability and sample size which is the same as MCS. It is noticeable that the efficiency of MCS is far below the M5Tree+MCS based on the indicator N_{call}. However, the $Abs.error(\beta)\%$ of M5Tree+MCS is close to zero which means M5Tree+MCS used less samples than MCS and gained almost the same results as MCS. MCS can combine with Latin Hypercube Sampling while sampling the random variables to improve the efficiency. Some defects on M5Tree cannot be neglected such as overfitting. It can be tackled by pruning methods like REP, PEP.

References

[1] Mohamed Seghier, Behrooz Keshtegar, José Correia, Abílio De Jesus, G. Lesiuk, Structural Reliability Analysis of Corroded Pipeline made in X60 Steel Based on M5 Model Tree Algorithm and Monte Carlo Simulation, Procedia Structural Integrity, 2018, 13: 1670-1675.

[2] Binghui Wu, Jianjun Xi, Na Gong, Structural Reliability Prediction Based on Monte Carlo Method, Machinery Manufacturing & Automation, 2015,44(06):68-70.

[3] Behrooz Keshtegar, Ozgur Kisi, M5 model tree and Monte Carlo simulation for efficient structural reliability analysis, Applied Mathematical Modelling, 2017, 48: 899-910.

[4] L. Neissi, M. Golabi, J.M. Gorman. Spatial interpolation of sodium absorption ratio: A study combining a decision tree model and GIS. Ecological Indicators 2020, 117: 106611.

[5] Yayun Gu, Linxian Hu. The Application of M5'Model Tree on Optimal Load Distribution in Thermal Power Plants. Energy Saving Technology, 2013,31(05):426-429.

[6] Chenlu Zhang, Jinquan Zhao, Weihua Luo, Jun Zhao, An optimal power flow calculation of probability based on Latin Hypercube Sampling, Science Technology and Engineering, 2014,14(14):49-53+59.

[7] Junwen Bao, Xinyu Hu, Mingyuan Xing, Shaofei Zhao, Comparing LHS with traditional rejection sampling methods in analysis of reliability, Journal of North China Institute of Science and Technology, 2021,18(03):81-84.

[8] Zhang J, Gao L, Xiao M. A new hybrid reliability-based design optimization method under random and interval uncertainties. International Journal for Numerical Methods in Engineering. 2020, 121(19):4435-57.

[9] Nuo Qu, Concept elaboration and algorithm optimization of decision tree, China New, 2018,20(21):157-158.

[10] Pham BT, Nguyen MD, Bui KT, Prakash I, Chapi K, Bui DT. A novel artificial intelligence approach based on Multi-layer Perceptron Neural Network and Biogeography-based Optimization for predicting coefficient of consolidation of soil. Catena. 2019, 173:302-11.