

Latin Hypercube Sampling with Evolutionary Algorithm for Static Security Risk Assessment

Junfang Li

The Research Centre of Power Grid Plan/Guangdong Power Grid Co., Ltd.
Dongfengdong Road, No. 757, Guangzhou, 510080, China
e-mail: jfliitwo@126.com

Abstract

Due to correlation coefficient matrix of initialized samples are not always positive definite, this paper presents the improved Latin Hypercube Sampling (LHS) methods with Evolutionary Algorithm (EA) to control correlation and handle power system probability analysis problem. To deal with the non-positive definite correlation matrix, an improved median Latin hypercube sampling with evolutionary algorithm (EA) called MLHS-EA into Monte Carlo simulationis proposed and investigated using IEEE 118-bus system with wind farms. This paper also discusses the misunderstandings about the non-positive definite correlation matrix and application of LHS in power system probabilistic analysis. With the proposed method in this paper, the correlation can be controlled more effectively than previous LHS methods. The accuracy of LHS for the static security assessment can also be improved further for solving the probabilistic analysis problem in power system. The effectiveness of the method is validated with the Matlab simulation results.

Keywords: Latin hypercube sampling, risk assessment, evolutionary algorithm, power system

Copyright © 2014 Institute of Advanced Engineering and Science. All rights reserved.

1. Introduction

To evaluate the impact of the load forecasting uncertainty on power systems' security, probabilistic analysis of the static security risk has attracted a large amount of interest from 1990s to recent years [1]-[3]. The probabilistic load flow (PLF) proposed by Borkowska in 1974 in [4] can be adopted to calculate the risk indices of line overload and bus over-voltage to evaluate the potential static security risk and weak points in power system. Monte Carlo simulation (MCS) [1]-[3], [5] is the most accurate, flexible and robust method when samples are enough. The theory of MCS with variance reduction technique is introduced in [3], [6], [7], better than MCS with simple random sampling (SRS).

For a deterministic function $Y=f(\mathbf{X})$, where $Y \in \mathbf{R}$ and $\mathbf{X} \in \mathbf{R}^k$ and $f(\cdot)$ is expensive to compute which may use a more complicated sampling scheme, McKay *et al.* described Latin hypercube sampling (LHS) first in 1979 [8]. LHS is a stratified sampling with a great merit in solving this problem with high efficiency. LHS has two steps: sampling and correlation control. The Median LHS (MLHS) to select the midpoint of each interval is the dominant approach. Correlation control to reduce undesired random correlation and to introduce prescribed correlation is important [9], because the correlation between the input variables has a profound impact on the uncertainty of the outputs.

Combined with correlation control methods, the LHS methods have many variants: LHS with Rank Gram-Schmidt (RGS) [9], LHS with random permutation (LHSRP) [10], LHS with Cholesky decomposition [11], [12], optimal LHS [13], LHS with Genetic Algorithm (GA) or columnwise-pairwise (CP) [14], LHS with simulated annealing [15] and single-switch-optimized sample ordering scheme (SSO) [16]. The performance in [14] indicates that CP methods are most efficient for small and medium size Latin hypercube, while an adopted GA performs better for large Latin hypercube. Crossover operator is important in GA. However, to a specified problem like PLF, it's difficult to use crossover operator. Based on the research on the quasi-Monte Carlo simulation, Latin hypercube Hammersley sequence sampling (LHHS) was developed [17]. LHHS generates the sample values with LHS and pairs these values with Hammersley sequence sampling (HSS) [17], [18]. LHHS needs enough prime numbers for different random variables. This fact limits its application when a problem has the high-

dimensional random variables. Theoretically, the consistency and unbiasedness about LHS with dependence are proved in [19]. The input distribution type for LHS methods is summarized [20].

The Cholesky decomposition needs that the correlation coefficient matrix ρ is a positive definite matrix. However, in real applications, ρ of input random variables can be non-positive definite when obtained by the samples in the first step.

In this paper, to improve the correlation control of LHS effectively when the correlation matrix is a non-positive definite matrix, correlation controlled by the Evolutionary Algorithm (EA) is presented. Many previous literatures in [21] rarely take into account the generator's reactive output constraint when calculating PLF. This paper considers the reactive power limit of the generators and the independence in different nodes when evaluating the steady-state security risk.

The remainder of the paper is organized as follows. Section II introduces the related background of LHS, presents the improved methods and analyzes the correlation matrix after sampling in the first step. Section III describes the simulation in the static risk assessment. Section IV applies the methods in IEEE30- and 118-bus system. Section V concludes the paper.

2. The Proposed Method

2.1. Related Background of Latin Hypercube Sampling

The main idea of MLHS is described as follows. Let Y_k is the cumulative distribution function of X_k , namely $Y_k = F_k(X_k)$, $X_k \in \{X_1, \dots, X_K\}$. Then the domain of each variable is divided into equal probable separated intervals, and one sample value of X_k is selected from each interval directly from Y_k , namely $X_k = F_k^{-1}(Y_k)$. In MLHS method, due to adopting the midpoint value from each interval, the n th sample of X_k is chosen according to $x_{k,n} = F_k^{-1}((n-0.5)/N)$. Assume that the cumulative distribution curves of two random variables following normal distribution $N(10\text{MW}, (4\text{MW})^2)$ are divided into N equal sections, respectively, and $N=6$. The cumulative distribution function (CDF) and probabilistic distribution function (PDF) of the random variable are shown in Figure 1(a)-(b), respectively.

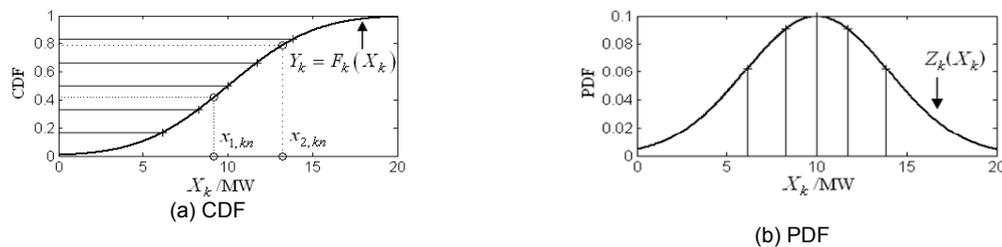


Figure 1. Sampling of MLHS, LHSRP method

In Figure 1(a), $x_{1,kn}$ and $x_{2,kn}$ represent the random point using MLHS and LHSRP, respectively. For three LHS methods, the random point in each interval is determined as follows.

$$\text{MLHS [12]: } x_{1,kn} = F_k^{-1}\left(\frac{(n-0.5)}{N}\right) \quad (1)$$

$$\text{LHSRP [22]: } x_{2,kn} = F_k^{-1}\left(\frac{r_{k,n}}{N} + \frac{(n-1)}{N}\right) \quad (2)$$

$$\text{LHHS [18]: } x_{H,kn} = F_k^{-1}\left(\frac{h_{k,n}}{N} + \frac{(n-1)}{N}\right) \quad (3)$$

The roles of parameter $r_{k,n}$ and $h_{k,n}$ are the random number generators in $[0, 1]$. The value of $r_{k,n}$ is a pseudo-random number. The value of parameter $h_{k,n}$ is from the Hammersley points. The details on the process to generate the Hammersley points can be found in [18]. For all LHS methods, the CDF of the input variable must be strictly increasing continuous function to ensure that the inverse function exists.

2.2. Description of EA for Correlation Control

The correlation control can be achieved by the permutation method. The main process of the EA is that random permutation for enough times and taking the best one when the objective correlation reaches an acceptable value. Inspired by the asexual propagation, the optimization permutation is called as "EA". The size of chromosomes represents the number of random variables. Each chromosome has the same size of genes, namely the sample size N . The fitness function evaluation is a minimization optimization problem of objective correlation function. The gene has a mutation behavior. In the context of the EA paradigm, mutation is seen as a change with a random element [23]. Thus, the update or change of the arrangement by a random permutation of N sampled values of X_i is called as "mutation". "Selection" means that the specific generation with a smaller objective value than all previous generations is selected as the best generation to survive, and the previous generations are eliminated.

Let X_1, \dots, X_K be a group of random variables and construct a K -dimensional random vector $\mathbf{X}=[X_1, \dots, X_K]^T$. Each random variable has N sampling values, and constructs a $K \times N$ sampling matrix \mathbf{M} . The covariance of random variables X_i and X_j is defined as:

$$c_{ij} = E([X_i - E(X_i)][X_j - E(X_j)]) \quad (4)$$

The correlation coefficient ρ_{ij} between X_i and X_j are defined by Pearson product-moment correlation coefficient:

$$\rho_{ij} = \frac{c_{ij}}{\sqrt{c_{ii} \cdot c_{jj}}} \quad (5)$$

The correlation matrix ρ with elements ρ_{ij} is symmetric. Before correlation control, two situations need to be distinguished. Situation 1: ρ of the independent variables can be denoted by an identity matrix. Situation 2: an objective correlation matrix ρ^* has most of elements $\rho_{ij} (i \neq j)$ in the interval $(0, 1)$ or $(-1, 0)$. Situation 1 is possible to happen in power system. The pre-determined objective correlation matrix ρ^* in Situation 2 in a bulk system is unrealistic. The reason for this is that it's very difficult to accurately give all realistic or approximately realistic non-diagonal elements in advance for a $K \times K$ matrix ρ^* and ensure ρ^* a positive definite matrix when K value is high. For example, in the IEEE 118-bus system, there are 99 active loads. If Situation 2 is considered, the matrix ρ^* has 99×99 elements. To accurately give more than 9000 non-diagonal elements in the interval $(0, 1)$ and ensure the positive definite matrix is unrealistic. Thus, the Situation 2 is unrealistic.

In a bulk system, the forecasting uncertainties of some load nodes are considered as approximate independent. The loads with a correlation coefficient equaling one can be denoted by linear correlation. The correlation matrix ρ is an identity matrix.

The optimization problem in the correlation control has a fitness function. To solve the situation 1, because completely independent variables strictly satisfying the correlation coefficient element $\rho_{ij} = 0 (i \neq j)$ are not easy to achieve, but approximate to zero. updating the sample permutation to get a minimum value of correlation objective function, Thus, the root mean square correlation among X_1, \dots, X_K is adopted as the objective function, that is,

$$\min \rho_s = \sqrt{2 \sum_{j=2}^K \sum_{i=1}^{j-1} \rho_{ij}^2 / [(K-1)K]} \quad (6)$$

The objective function ρ_s represents the root mean square value of off-diagonal lower triangular or upper triangular elements of ρ . If ρ_s is close to zero, the correlation between variables is small. The direction of mutation is to get a new arrangement with a minimum objective value ρ_s . The objective function ρ_s is calculated using the updated sampling matrix \mathbf{M} .

2.3. Steps of LHS Method with EA

The above three LHS methods with EA can be denoted by MLHS-EA, LHSRP-EA and LHHS-EA. Assume there are K random variables. For the above-mentioned Situation 1, the steps of LHS with EA are as follows.

Step 1: Generate an initialized $K \times N$ sampling matrix \mathbf{M}_0 for one of three LHS methods according to the Eq. (1)-(3). It can be deduced that the elements of \mathbf{M}_0 are sorted in an ascending order from the Eq. (1). Take MLHS-EA for example:

$$\mathbf{M}_0 = \begin{bmatrix} x_{1,11} & x_{1,12} & \cdots & x_{1,1N} \\ x_{1,21} & x_{1,22} & \cdots & x_{1,2N} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1,K1} & x_{1,K2} & \cdots & x_{1,KN} \end{bmatrix}_{K \times N} \quad (7)$$

Step 2: Calculate the initialized correlation matrix ρ of the matrix \mathbf{M}_0 . For example, using MLHS-EA, IEEE 14-, 30-, 57- and 118-bus systems are tested after the first step. The matrix ρ is found to be a non-positive definite matrix with all elements equalling one, namely:

$$\rho = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}_{K \times K} \quad (8)$$

If using LHSRP-EA, IEEE 14-, 30-, 57- and 118-bus systems are found that the off-diagonal elements of ρ are all in the interval (0.9, 1.0), and very close to one.

Step 3: Control correlation using EA. Set the maximum generation is G , and initialize the objective function value $\rho_{smin}=1$. The matrix \mathbf{M}_g in the g th generation ($1 \leq g \leq G$) can be seen as a cell. Each row of the cell is a chromosome with N genes. Thus, the cell \mathbf{M}_g has K chromosomes and total $K \times N$ genes. After each evolution, the sequence of genes in each chromosome will be sorted randomly again. For example, the \mathbf{M}_g evolves to be the following form after g th iterations.

$$\mathbf{M}_g = \begin{bmatrix} x_{1,1N} & x_{1,11} & x_{1,13} & \cdots & x_{1,12} \\ x_{1,22} & x_{1,21} & x_{1,2N} & \cdots & x_{1,23} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{1,KN} & x_{1,K2} & x_{1,K3} & \cdots & x_{1,K1} \end{bmatrix}_{K \times N} \quad (9)$$

Then calculate the ρ_s in the g th generation. If $\rho_s \leq \rho_{smin}$, set $\rho_{smin} = \rho_s$.

Step 4: Continue iterations when $g < G$ and stop when $g = G$. The matrix \mathbf{M}_G is the final sampling matrix approximately satisfying the independence between the random variables.

3. Research Method in the Static Security Risk Assessment

Risk, defined by J. D. McCalley and V. Vittale *et al.* in [1], is the mean impact of an event. In this paper, we adopt two risk indices and two severity indices, a little different from the "risk" defined in [1], because the risk indices based on the full probability events may cover the detailed risk information of a single event. The risk assessment in this paper aims at evaluating risk and the most severity degree considering load and wind forecast uncertainties.

3.1. Static Security Risk Assessment Based on PLF

To solve some unavoidable risk problem even though the power system operation mode has been optimized, the risk assessment on overload and voltage beyond limit is

necessary, because it is possible that some small contingencies may result in a catastrophic outcome: cascading overload, voltage instability and so on. The uncertainty of the input controlled variable is modeled as a specific distribution. After load flow calculation, the output variables are random variables following different distribution characteristics:

$$\begin{cases} \mathbf{G} = g(\mathbf{U}, \boldsymbol{\theta}) \\ \mathbf{H} = h(\mathbf{U}, \boldsymbol{\theta}) \\ \text{s.t. } Q_{G\min} \leq Q_G \leq Q_{G\max} \end{cases} \quad (10)$$

\mathbf{G} is formed by active power injection \mathbf{P}_Ω and reactive power injection \mathbf{Q}_Ω generated by the generators and load demands at all buses, namely $\mathbf{G} = [\mathbf{P}_\Omega, \mathbf{Q}_\Omega]$. Besides, the output vector \mathbf{H} contains active load flow vector \mathbf{P}_R and reactive load flow vector \mathbf{Q}_R , namely $\mathbf{H} = [\mathbf{P}_R, \mathbf{Q}_R]$.

3.2. Risk Indices of Static Security Assessment

When the cost-consequences are difficult to measure, a severity measurement can be used instead. Based on the solutions by PLF calculation, two risk indices for evaluating the risk of over-voltage and overload of transmission lines, and two severity indices are defined as:

1) Risk of over-voltage ($Risk_U$)

It is defined as the over-voltage probability when the security region is $[U_{\min}, U_{\max}]$:

$$Risk_U = 1 - \Pr\{U_{\min} \leq U \leq U_{\max}\} \quad (11)$$

2) Risk of overload ($Risk_{PF}$)

It is defined as the overload probability in lines when the security region is $[P_{f\min}, P_{f\max}]$.

$$Risk_{LF} = 1 - \Pr\{P_{f\min} \leq P_f \leq P_{f\max}\} \quad (12)$$

If $Risk_U \leq 0.1$ and $Risk_{LF} \leq 0.1$, the system maintains static security. Otherwise, the system has potential crisis.

3) Most severity metric of over-voltage (Sev_U)

It denotes the most severity degree of each node's voltage magnitude deviating from the safety range $[U_{\min}, U_{\max}]$ in all N samples.

$$Sev_U = \max\left\{\frac{|U_i - (U_{\max} + U_{\min})/2|}{(U_{\max} - U_{\min})/2}\right\}, i = 1, 2, \dots, N \quad (13)$$

4) Most severity metric of overload (Sev_{LF})

This index denotes the most severity degree of the active load flow in each line $i-j$ deviating from the safety range $[P_{f\min}, P_{f\max}]$ in all N samples.

$$Sev_{LF} = \max\left\{\frac{|P_{ij} - (P_{ij}^{\min} + P_{ij}^{\max})/2|}{(P_{ij}^{\max} - P_{ij}^{\min})/2}\right\} \quad (14)$$

3.3. Process of PLF

The process of solving PLF problems is as follows.

Step 1: Initialize the value of N , K , the mean value μ and standard deviation σ of probability distribution of input random variables, including active power, reactive power of load and wind power.

Step 2: Generate $K \times N$ sampling matrix \mathbf{M}_0 according to the Eq. (1)-(3). Control correlation using EA and get the matrix \mathbf{M}_G .

Step 3: Based on the sampling matrix \mathbf{M}_G , Newton-Raphson load flow program runs for N times. Finally, the statistic values of μ and σ of solutions are obtained.

To evaluate the performance, the relative error [12] is used.

$$\varepsilon_z^s = \left| \frac{X_z^s - X_{zbase}^s}{X_{zbase}^s} \right| \times 100\% \quad (15)$$

The reference value X_{zbase}^s is obtained by enough samples using SRS. To evaluate whole system, the average relative errors of the expected value of the output random variables are calculated. The relative change rate of ρ_s , namely $\eta(\%)$, is used to evaluate the performance of controlling correlation.

$$\eta = \frac{\rho_{s1} - \rho_{s2}}{\rho_{s2}} \times 100\% \quad (16)$$

Where, ρ_{s1} and ρ_{s2} are equal to ρ_s obtained by algorithm with EA and without EA, respectively. The sampling size of SRS needs to be no less than the value satisfying the stopping criteria. Two common stopping criteria are adopted. Criterion A: The average percentage changing rate of the results under two continuous sampling sizes is less than 0.01%. The results, when the sampling sizes are $N-1$ and N ($N > 1$, positive integer), are y_{N-1} and y_N , respectively, the percentage changing rate r is calculated as Eq. (17). Criterion B: Each output random variable' sample variance is within 0.01% of the expected value in above 90% probability.

$$r = \left| (y_N - y_{N-1}) \times 100\% / y_{N-1} \right| \quad (17)$$

Referring to the previous articles [12], [16], for IEEE 118-bus system, 30000 times of MCS with SRS for the PLF is enough.

4. Results and Discussion

The validity of the proposed method in the static risk assessment is demonstrated on IEEE 118-bus system. The program is developed with MATPOWER 4.0 on Dual Core 2.71GHz PC with 1.75G of RAM. The wind power and the load are seen as independent random variables. The forecasted wind power values from different wind farms are independent. The wind power uncertainty is modelled as Beta distribution [24], the mean value of which equals the average forecasted power and the standard deviation is estimated as 30% of the mean value. The nodal active and reactive power distributions of loads follow normal distribution. The mean value is the same as the original data provided by MATPOWER 4.0. And the standard deviation of loads is equal to 10% of the mean value, namely $\sigma = 10\% \mu$.

4.1. IEEE 30-bus System

For testing the efficiency of controlling correlation, take IEEE 30-bus system for example. The numerical data are from MATPOWER 4.0. The curves of ρ_s and η of the four methods are shown in Figure 2 (a)-(b). When $N=50$, the results are acceptable. The computation time for LHSRP, LHHS, MLHS-EA and LHIS-EA when $N=50$ is 2.37s, 2.31s, 1.94s, 1.97s, respectively. From Figure 2, the correlation can be controlled with EA better than without EA.

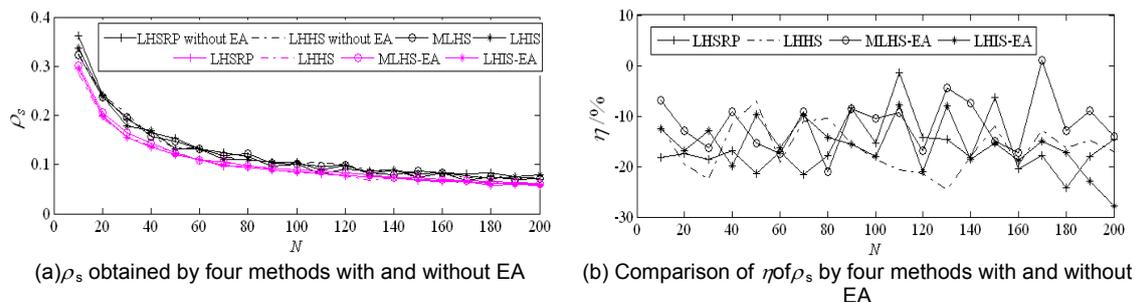


Figure 2. Comparison of effectiveness of correlation control by EA

4.2. Modified IEEE 118-bus System with Wind Farms

IEEE 118-bus system has 186 lines, shown in Figure 3. There are 189 independent input random variables including 99 active loads and 90 reactive loads. The test system has been modified to include three wind farms having all Doubly Fed Induction Generators in nodes 10, 65 and 89, respectively. Wind power replaces part of the conventional generation in the three nodes, e.g., 120 MW wind power in node 10, 180 MW wind power in node 65, and 162 MW wind power in node 89. After wind power integrated into 118-bus system, the capacities of the thermal units in the nodes 10, 65 and 89 are 330MW, 211MW and 445MW, respectively. $[U_{\min}, U_{\max}] = [0.95\text{pu}, 1.05\text{pu}]$, and $[P_{\min}, P_{\max}] = [-500\text{MW}, 500\text{MW}]$.

We assume that the total rated power of each wind farm is 300MW. The parameters of Beta distribution for each wind farm calculated according to [24] are given in Table 1. The Beta distributions are shown in Figure 4. In EA, $G=1000$. In MLHS, the sample size N is 300.

The error of the expected value of voltage magnitude in each node is shown in Figure 5. The indices $Risk_U$ and Sev_U of the first three nodes in a descending order are given in Table 2. The indices $Risk_{LF}$ and Sev_{LF} of the first three lines in a descending order are given in Table 3. To analyze the influence of load and wind power uncertainties on the load flow, the situation of $\sigma=10\%\mu$ is compared with $\sigma=1\%\mu$. The indices $Risk_U$ and Sev_U of the first three nodes in a descending order are given in Table 4. The CDF curves of voltage magnitude in node 53 and node 118 are shown in Figure 6.

Table 1. The Statistical Parameters for Wind Farms in the Node 10, 65 and 89

	Node 10	Node 65	Node 89
Mean (p.u.)	0.4	0.6	0.54
Std. (p.u.)	0.12	0.18	0.162
Beta distribution	$\beta(6.27, 9.40)$	$\beta(3.84, 2.56)$	$\beta(4.57, 3.89)$
Rated power (MW)	300	300	300

Table 2. Sorting table of $Risk_U$ and Sev_U when loads satisfy $\sigma=10\%\mu$

No.	Node	$Risk_U$	No.	Node	Sev_U
1	53	0.9933	1	53	1.1848
2	118	0.7100	2	118	1.0716
3	2	0	3	21	0.9660

Table 3. Sorting table of $Risk_{LF}$ and Sev_{LF} when loads satisfy $\sigma=10\%\mu$

No.	Line	$Risk_{LF}$	No.	Line	Sev_{LF}
1	9-10	0.0667	1	9-10	1.0966
2	8-9	0.0500	2	8-9	1.0827
3	1-2	0	3	8-5	0.7516

Table 4. Sorting table of $Risk_U$ and Sev_U when loads satisfy $\sigma=1\%\mu$

No.	Node	$Risk_U$	No.	Node	Sev_U
1	53	1	1	53	1.0907
2	118	1	2	118	1.0193
3	2	0	3	9	0.9215

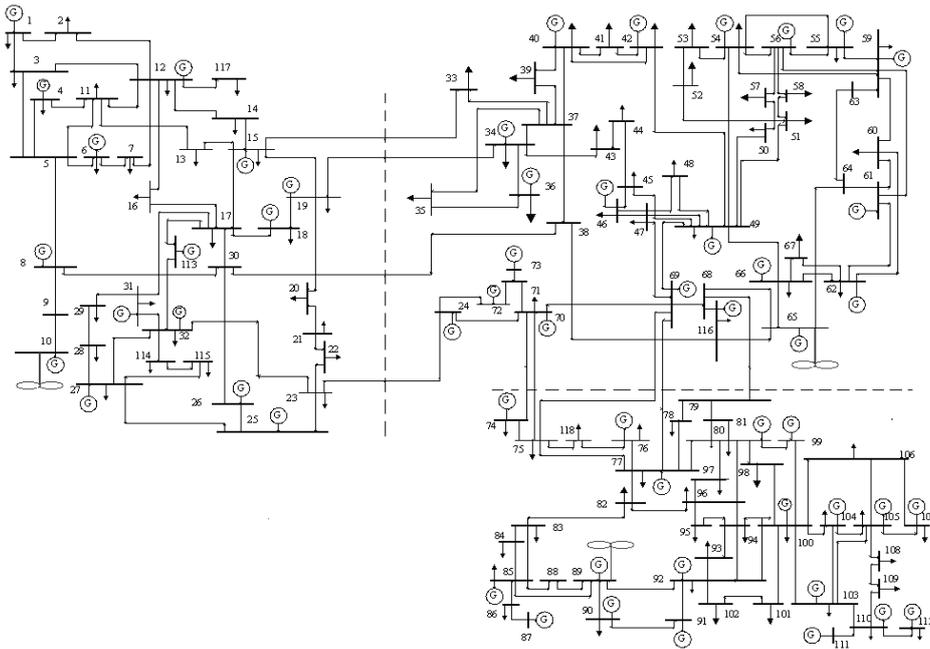


Figure 3. Wind farms in node 89, 65 and 10 in IEEE 118-bus system

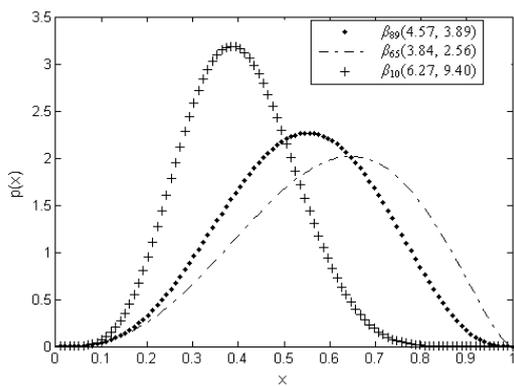


Figure 4. Beta distribution

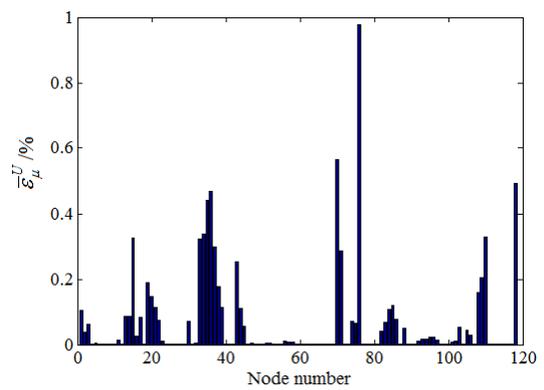


Figure 5. Error of μ of voltage magnitude in 118 nodes

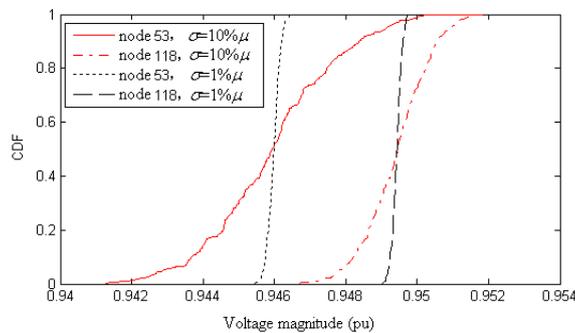


Figure 6. CDF of voltage magnitudes of node 53 and 118 in wind integrated power system

4.3. Results Analysis

The computation time using MLHS-EA is 8.4 seconds. The expected values of voltage magnitude and active power flow obtained by MLHS-EA for the PLF have error about 0.067% and 2.027%, respectively. The reason for only giving the indices of the first three nodes in Table 2 and Table 3 is that the $Risk_U$ of other nodes after the first three nodes are all zero. From Table 2 and Table 3, the nodes 53 and 118 have very high possibilities of over-voltage and severity degree. The lines 9-10 and 8-9 have small possibilities of overload but high severity degree. In Figure 5, the maximum error of the expected values of voltage magnitudes in all 118 nodes is less than 1%. As shown in Figure 6, Table 2 and Table 4, the nodes 53 and 118 have high crisis. If the fault events are considered, the system has great crisis. Thus, a prevention control to keep voltage security is necessary. The results show that LHS with EA is effective to solve the PLF problem when the correlation matrix obtained after the sampling initialization is non-positive definite. For a large-scale power system with most independent random variables, the improved methods are effective.

4.4. Discussion

This section aims at discussing the misunderstandings about the application of LHS in power system probabilistic analysis. These misunderstandings and questions are as follows:

(1) Why the correlation matrix ρ is non-positive definite? Why not generate a positive definite matrix ρ that could be done by the Cholesky decomposition?

The answer is that the matrix ρ is the correlation matrix of the matrix M_0 formed by the sampling rule (i.e. Eq. (1), (2) and (3)) in the first step of LHS. The sample points are all naturally sorted in an increasing form. The correlation matrix is obtained by calculation, not given in advance. The calculated matrix ρ after the first step using MLHS is a non-positive matrix that can't be done by Cholesky decomposition.

The random variables are completely correlated in the first step. Because the precondition of the PLF problem is that the random variables are independent, it needs to control correlation to make the sample matrix M satisfy the precondition. That is the reason why use EA, not Cholesky decomposition to control correlation in the second step.

(2) There exist rare nodes with high errors of expected values of voltage magnitudes, or rare lines with high errors of expected values of active power flow when other nodes or lines have all statistical results with very small errors. Does this show no benefit of using LHS-EA?

The answer is that rare nodes or lines with high errors are realistic, especially for a random variable with a very low actual value. For example, if the actual statistical result $\mu \pm \sigma$ of the voltage magnitude in a node is 0.9871 ± 0.0005 pu, and the result obtained using LHS is 0.9881 ± 0.0004 pu, it can be found that using Eq. (15), the relative error of μ is 0.1%, and the relative error of σ is 20% (i.e. $|0.0004 - 0.0005| / 0.0005$). Because the relative error rate (i.e. σ/μ) is quite small, i.e. $\sigma/\mu = 5 \times 10^{-4}$, the relative error of σ has no influence on judging the performance of LHS-EA. IEEE 30-, 57-bus system are also tested and found that the phenomena is very common. Thus, it should pay great attention to the authenticity of the results. The phenomena with high error for rare nodes or lines can't deny the validity of LHS with EA.

5. Conclusion

This paper presents the improved LHS methods with Evolutionary Algorithm to control correlation and handle power system static security risk assessment problem. To deal with the non-positive definite correlation matrix, an improved median Latin hypercube sampling with EA called MLHS-EA into Monte Carlo simulation is proposed and investigated using modified IEEE 118-bus system with wind farms in this paper. With the method proposed in this paper, the correlation is effectively controlled and the accuracy of the LHS for the static security assessment can be improved further than previous LHS methods. The methods can be used for solving the probabilistic analysis problem in power system.

References

- [1] JD McCalley, V Vittal and N Abi-Samra. *An overview of risk based security assessment*. IEEE Power Engineering Society Summer Meeting. July 18-22, Edmonton, Canada. 1999; 1: 173-178.
- [2] F Xiao and JD McCalley. Power system risk assessment and control in a multiobjective framework. *IEEE Trans. on Power Syst.* 2009; 24(1): 78–85.
- [3] DS Kirschen, D Jayaweera, DP Nedic, et al. A probabilistic indicator of system stress. *IEEE Trans. on Power Syst.*, 2004; 19(3): 1650–1657.
- [4] B Borkowska. Probabilistic load flow. *IEEE Trans. Power App. Syst.* 1974; PAS-93(3): 752–759.
- [5] W Li. *Risk Assessment of Power Systems: Models, Methods, and Applications*. New York: IEEE Press/Wiley. 2004: 79–86.
- [6] A Breipohl, FN Lee, J Huang et al. Sample size reduction in stochastic production simulation. *IEEE Trans. Power Syst.* 1990; 5(3): 984–992.
- [7] Z Bie and X Wang. Studies on variance reduction technique of Monte Carlo simulation in composite system reliability evaluation. *Electric Power Systems Research*. 2002; 63(1): 59–64.
- [8] MD McKay, RJ Beckman, and WJ Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*. 1979; 21(2): 239–245.
- [9] AB Owen. Controlling correlations in Latin hypercube samples. *Journal of the American Statistical Association*. 1994; 89(428): 1517–1522.
- [10] P Jirutitijaroen and C Singh. Comparison of simulation methods for power system reliability indexes and their distributions. *IEEE Trans. Power Syst.* 2008; 23(2): 486–493.
- [11] RL Iman and WJ Conover. A distribution-free approach to inducing rank correlation among input variables. *Commun. Statist. –Sim. Comput.* 1982; 11(3): 311–334.
- [12] H Yu, CY Chung, KP Wong et al. Probabilistic load flow evaluation with hybrid Latin hypercube sampling and Cholesky decomposition. *IEEE Trans. Power Syst.* 2009; 24(2): 661–667.
- [13] Jeong-Soo Park. Optimal latin-hypercube designs for computer experiments. *Journal of Statistical Planning and Inference*. 1994; 39(1): 95–111.
- [14] M Liefvendahl and R Stocki. A study on algorithms for optimization of Latin hypercubes. *Journal of Statistical Planning and Inference*. 2006; 136(9): 3231–3247.
- [15] M Vořechovský and D Novák. Correlation control in small-sample Monte Carlo type simulations I: A simulated annealing approach. *Probabilistic Engineering Mechanics*. 2009; 24(3): 452–462.
- [16] DE Huntington and CS Lyrintzis. Improvements to and limitations of Latin hypercube sampling. *Probabilistic Engineering Mechanics*. 1998; 13(4): 245–253.
- [17] JR Kalgnanam and UM Diwekar. An efficient sampling technique for off-line quality control. *Technometrics*. 1997; 39(3): 308–319.
- [18] Urmila M Diwekar. A novel sampling approach to combinatorial optimization under uncertainty. *Computational Optimization and Applications*. 2003; 24(2-3): 335–371.
- [19] Packham, Natalie and Schmidt, Wolfgang M. *Latin Hypercube Sampling with Dependence and Applications in Finance*. Germany: Frankfurt School of Finance & Management, 2008: 1-23.
- [20] LP Swiler and GD Wyss. *A user's guide to Sandia's Latin hypercube sampling software: LHS Unix Library/standalone version*. Sandia National Laboratories, Albuquerque, NM. Technical Report SAND 2004–2439.
- [21] P Chen, Z Chen, and B Bak-Jensen. *Probabilistic load flow: a review*. The 3rd International Conference on Electric Utility Deregulation and Restructuring and Power Technologies. Nanjing, China. 2008; 1: 1586-1591.
- [22] R Wang, Urmila Diwekar, Catherine E Grégoire Padró. Efficient sampling techniques for uncertainties in risk analysis. *Environmental Progress*. 2004; 23(2): 141–157.
- [23] S Das and PN Suganthan. Differential evolution: A survey of the state-of-the-art. *IEEE Trans. on Evolutionary Computation*. 2011; 15(1): 4–30.
- [24] Usaola J. Probabilistic load flow in systems with wind generation. *Generation, Transmission & Distribution, IET*. 2009; 3(12): 1031–1041.