Stochastic Power Flow Calculation of High-Density Correlation Input Power Grid Based on Hybrid Gaussian Model

Wang Qing\textsuperscript{1,a}, Hongxin Li\textsuperscript{1,b}, Wang Wei\textsuperscript{1,c} and Xue Rong\textsuperscript{1,d}

\textsuperscript{1}Electric Power Research Institute, Shenzhen Power Supply Co., Ltd, Guangdong, Shenzhen, 518000, China

\textsuperscript{a}hanhuichen0@163.com; \textsuperscript{b}csgwang@126.com; \textsuperscript{c}626557882@qq.com; \textsuperscript{d}cquhvxr@163.com

Abstract. Traditional stochastic power flow calculation assumes that the input variables are uncorrelated and directly used to calculate high-density correlation input variables has certain errors. To this end, this paper proposes a stochastic power flow algorithm based on mixed Gaussian models that takes into account input correlation. First, a mixed Gaussian model is established based on load data, then uniform sampling and Naf transformation are used to process random variables, and the high-density correlation in the original space is converted into independent random variables, and the piecewise linear Monte Carlo method is used to carry out random power flow. Calculation can greatly reduce the calculation time and reduce the truncation error as much as possible. Finally, combined with the actual calculation examples of IEEE-30, the results show that the algorithm can simplify the actual calculation process, improve the calculation efficiency, and has certain practical value in engineering.

1. Introduction

With the rapid development of random new energy (wind power, solar energy and other clean energy), the grid in the future will exhibit high-dimensional randomness, which will bring great challenges to the operation control and optimization of power systems [1]. Therefore, it is of great significance to accelerate the establishment of grid assessment methods adapted to the new environment [2].

In 1974, Probabilistic Load Flow (PLF) [3] was proposed by Borkowska to solve random factors in power systems. After decades of development, three major results have gradually formed, namely the Monte Carlo method (Monte Carlo Simulation (MCS) [4], Point Estimate Method (PEM) [5], analytical method [6]. The Monte Carlo method is based on sampling the input probability function and then performing multiple deterministic power flow calculations. The accuracy is high, but the amount of calculation is large. The point estimation method determines the estimated values and weights of parameters based on the probability distribution function of the input quantity. Its calculation is simple, but the accuracy is poor. The analytical method is to linearize a nonlinear equation and then perform convolution calculations. Not only is the calculation speed fast and the accuracy is high, but it takes a large amount of calculation to consider the correlation of the input vector. Literatures, such as [7], use Latin hypercube sampling technology to significantly improve the sampling efficiency. The literature [8] used the bimodal distribution and the double Weibull distribution to fit the load and wind speed models respectively, which is suitable for some special occasions, but lacks generality. Literature [9]...
established the K-means model of load, and used the clustering method to process the load data of different periods. The model implicitly included the correlation between node loads, but the test required a lot of time.

This paper proposes a high-density correlation input power grid stochastic power flow calculation based on a hybrid Gaussian model. First, a mixed Gaussian model is established based on the load data, and then uniform sampling and NaF transform are used to process the random variables. The high-density correlation of the original space is converted into independent random variables, and the piecewise linear Monte Carlo method is used for random power flow calculations can minimize truncation errors. Finally, combined with the improved IEEE-30 actual examples to verify, the results show that the algorithm can simplify the actual calculation process, improve the calculation efficiency, and has a certain engineering practical value.

2. Gaussian mixture model of input variables

2.1. Gaussian mixture model
The randomness and diversity of the load make it difficult to fit the probability density distribution function of the input variable with a single standard normal distribution, and the Gaussian mixture model has been widely used as a current research hotspot to fit the probability distribution of any load. The Gaussian mixture model is a superposition of several Gaussian normal distribution functions. Its probability density function is as follows:

\[
f_{\text{GMM}}(x) = \sum_{i=1}^{n} w_i f_G(x; \mu_i, \sigma_i^2)
\]

In the formula, \( n \) is the number of sub-components, \( f_G(x; \mu_i, \sigma_i^2) \) is the Gaussian distribution of the \( i \)-th sub-component, \( \mu_i \) and \( \sigma_i^2 \) are the mean and variance of the \( i \)-th sub-component weight, respectively, and the \( i \)-th sub-component weight, which also represents the probability of being selected, and satisfies \( \sum_{i=1}^{n} w_i = 1 \). The total number of Gaussian components \( n \) in the above formula directly affects the accuracy of the Gaussian mixture model. The larger the value of \( n \), the more accurate the model is, but the higher the computational complexity, it is necessary to reasonably estimate the parameters to ensure that the precise conditions are met. Small amount.

2.2. Parameter estimation of mixed Gaussian models
In general, when determining the parameters \( w_i, \mu_i, \sigma_i^2 \) of Gaussian mixture models, the maximum likelihood method is often used for parameter estimation. The maximum likelihood function is as follows:

\[
L = \sum_{i=1}^{N_L} \ln \left( \sum_{i=1}^{n} w_i f_G(x_i; \mu_i, \sigma_i^2) \right)
\]

In the formula, \( x_i \) is the \( i \)-th load data, and \( N_L \) is the total load points.

In order to find the most suitable set of parameter values, so that the above formula obtains the maximum value, and then the value of the measurement data \( x_i \) is unknown, and it cannot be directly derived, so the expectation-maximum algorithm (EM) is introduced to solve the process, as follows:

Estimate the probability \( p \) of the data point \( x_i \) generated by each Gaussian component \( k \), that is, the expected value of each component. The calculation formula is as follows:

\[
P(i,k) = \frac{w_k f_G(x_i; \mu_k, \sigma_k^2)}{\sum_{i=1}^{n} w_i f_G(x_i; \mu_i, \sigma_i^2)}
\]
Then obtain the optimal model parameters according to the maximum likelihood method of equations (4) to (6).

\[ w_k = \frac{1}{N_L} \sum_{i=1}^{N_L} P(i, k) \]  

(4)  

\[ u_k = \frac{\sum_{i=1}^{N_L} P(i, k) x_i}{\sum_{i=1}^{N_L} P(i, k)} \]  

(5)  

\[ \sigma_k^2 = \frac{\sum_{i=1}^{N_L} P(i, k) (x_i - u_k)^2}{\sum_{i=1}^{N_L} P(i, k) x_i} \]  

(6)  

Repeat the iteration until the convergence accuracy of the likelihood function (2) can meet the requirements. In order to improve the convergence rate, the K-means clustering method can be used to obtain the approximate result as the initial value of the parameter, and then the expected maximum method can be used to iteratively solve.

3. Correlation input samples

The uniform sampling technique reflects the overall distribution information through less sample data. It has a small calculation scale and the collected samples are evenly distributed. It is a statistical sampling.

Set the correlation coefficient matrix \( \rho_x \) of the m-dimensional random input variable \( X = [x_1, x_2, x_3, \ldots, x_m]^T \), according to the probability density function \( f_i(x_i) \) and cumulative distribution function \( F_i(x_i) \) of the random variable \( x_i \), a standard normal distribution vector \( Z \) can be generated, and the correlation coefficient matrix \( \rho_Z \) is as shown in Equation (9).

\[
\begin{align*}
F_i(x_i) &= \Phi(z_i) \\
\phi_i &= F_i^{-1}(x_i)
\end{align*}
\]

(9)  

Where \( \Phi(\cdot) \) is the cumulative distribution function of the standard normal quantity. The correlation coefficient matrix \( \rho_x \) of X and the off-diagonal elements of the normal distribution vector \( \rho_Z \) are denoted by \( \rho_{x_{ij}} \) and \( \rho_{z_{ij}} \), which satisfy the empirical formula [10], as shown in formula (10).

\[ \rho_{z_{ij}} = 2 \sin \left( \frac{\pi}{6} \rho_{x_{ij}} \right) \]  

(10)  

When \( \rho_{x_{ij}} \) known, you can quickly find \( \rho_{z_{ij}} \). L can be obtained by the formula \( \rho_Z = LL^T \), where L is the defined singular value decomposition \( \rho_Z \). Then take equation (10) to get \( Z \), finally get the independent standard normal distribution \( Y \) by \( Y = LZ^* \).

\[
Z^* = LY = \begin{bmatrix} x_1 & \ldots & l_{i_1} & \ldots & y_1 \\ \vdots & \ldots & \vdots & \ldots & \vdots \\ l_{i_m} & \ldots & l_{i_m} & \ldots & y_m \end{bmatrix}
\]  

(11)  

Finally, by using formula (12), the input random variable sample \( X \) with correlation can be obtained. The above process is the Nataf transform algorithm.

\[ X = F^{-1} \left( \Phi \left( Z^* \right) \right) \]  

(12)
4. Piecewise linear Monte Carlo method

Let $X$ be the node voltage phase angle and amplitude, $Z$ be the reactive and active power injection vector of the line, and $Y$ be the reactive and active power injection vector of the node. Expand the linear power flow equation into a Taylor series at the reference point, and keep ignoring the higher-order terms more than 2 times to obtain the following formula:

$$
\begin{align*}
X &= X_0 + \Delta X = X_0 + S \Delta Y \\
Z &= Z_0 + \Delta Z = Z_0 + T \Delta Y
\end{align*}
$$

In the formula, the subscript is 0 for running at the reference point, and $\Delta Z$ and $\Delta Y$ represent the fluctuation of the line and node voltage, $S$ and $T$ represent the sensitivity matrix, $S = J^T$, where $J$ is the Jacobian matrix, $T = GJ^T$, $G = \partial Z / \partial X$ when $X = X_0$.

For loads with high density, correlation, and randomness, the range of change is often large. The use of a single-point linearization model is usually accompanied by a large truncation error, which affects the calculation accuracy. Therefore, a piecewise linear power flow equation is introduced on this basis. Let $P_{\text{tot}}$ be the total active power of the system, the calculation formula is as follows:

$$
P_{\text{tot}} = \sum_{i=1}^{N_L} P_i - \sum_{j=1}^{N_G} P_j
$$

In the above formula, $N_L$ is the load at the PQ node, $P_i$ is the active power at the load node $i$, $N_G$ is the PV node load, and $P_j$ is the active power output by the generator $j$.

Due to the randomness of the high-density correlation load, $P_{\text{tot}}$ also has randomness, but its probability density distribution function can be determined by the above formula. Figure 1 roughly gives the probability density distribution function of $P_{\text{tot}}$, which is divided into $T$ regions at equal intervals, which are $R_1$-$R_T$, generally 6-8, and $\mu$ is expected.

Select the corresponding reference point to run in each area, and linearize the points of the power flow equation to:

$$
\begin{align*}
X_1 \approx X_{10} + S_1 \left( W_1 - \bar{W}_1 \right) \\
X_2 \approx X_{20} + S_2 \left( W_2 - \bar{W}_2 \right) \\
\vdots \\
X_T \approx X_{T0} + S_T \left( W_T - \bar{W}_T \right)
\end{align*}
$$

Figure 1. Probability density distribution function of $P_{\text{tot}}$. 

In the above formula, $N_L$ is the load at the PQ node, $P_i$ is the active power at the load node $i$, $N_G$ is the PV node load, and $P_j$ is the active power output by the generator $j$. 

Due to the randomness of the high-density correlation load, $P_{\text{tot}}$ also has randomness, but its probability density distribution function can be determined by the above formula. Figure 1 roughly gives the probability density distribution function of $P_{\text{tot}}$, which is divided into $T$ regions at equal intervals, which are $R_1$-$R_T$, generally 6-8, and $\mu$ is expected.

Select the corresponding reference point to run in each area, and linearize the points of the power flow equation to:
Among \( \overrightarrow{W} = f(X_i) \), \( S_i \) is a sensitivity matrix corresponding to the region \( R_i \), \( \overrightarrow{W} \) is a reference power vector corresponding to \( R_i \), and \( X_i \) is a state vector of the node \( i \).

In summary, the flow of the stochastic power flow algorithm for high-density correlation input power grids based on a hybrid Gaussian model proposed in this paper is shown in figure 2.

![Figure 2. Algorithm flowchart](image)

### 5. Case Analysis

In order to verify the reliability of this algorithm, this paper uses IEEE30 node system as an example. The input probability distribution characteristics of each node can be found in the literature [11]. The model parameters obtained by using the Gaussian mixture model are shown in Table 1. N or 2 or 3 can meet the convergence accuracy requirements.

| Node active power | \( w_i \) | \( w_i \) | \( w_i \) | \( \mu_i \) | \( \mu_i \) | \( \mu_i \) | \( \sigma_i^2 \) | \( \sigma_i^2 \) | \( \sigma_i^2 \) |
|-------------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| P5                | 0.6     | 0.4     | -0.069  | -0.077  | 0.0036  | 0.0021  |
| P13               | 0.5     | 0.2     | 0.3     | -0.025  | -0.02   | -0.018  | 0.00063 | 0.0009  | 0.0004  |
| P20               | 0.4     | 0.2     | 0.3     | -0.17   | -0.18   | -0.16   | 0.00026 | 0.00048 |

Based on the 100-time piecewise linear Monte Carlo method as the calculation basis, the average error index in the statistical characteristic parameters of the output is used as the evaluation index. The calculation formulas are shown in equations (16) and (17). The comparison results are shown in Table 2.

\[
\varepsilon_i^\gamma = \left| \frac{X^\gamma_i - \overline{X^\gamma_i}}{X^\gamma_{im}} \right| \times 100\% \tag{16}
\]

\[
\bar{\varepsilon_i} = \left| \frac{\varepsilon_i^\gamma_1 + \varepsilon_i^\gamma_2 + \ldots + \varepsilon_i^\gamma_n}{n} \right| \times 100\% \tag{17}
\]

Where \( \varepsilon_i^\gamma \) is the relative error evaluation index, and \( X^\gamma_i \) and \( X^\gamma_{im} \) respectively represent the comparison of the output of this method and Monte Carlo method. S represents the mean and standard deviation of the statistic.
Table 2. Calculation result of average relative error index

| Output variable | Voltage amplitude | Voltage phase angle | Active voltage | Voltage reactive power |
|-----------------|-------------------|---------------------|----------------|-----------------------|
| $E_{u}$         | 0.022             | 0.334               | 0.1357         | 0.908                 |
| $E_{v}$         | 0.558             | 0.276               | 0.43           | 0.895                 |

It can be seen from the above table that the maximum error is 0.908%, which satisfies the allowable range of errors. It proves that the singular value uniform sampling and processing of the input method can obtain accurate mathematical statistics results.

At the same time, the paper also extracts the probability distribution curve of the voltage amplitude of the node 13 and the voltage phase angle of the node 14 and compares it with the Monte Carlo method. The results are shown in Figure 3. The dotted line represents the method in this paper.

The results show that the accuracy of the method in this paper is approximately the same as that of the Monte Carlo method, but the running time of the Monte Carlo algorithm is 63.186s, while the algorithm in this paper only requires 2.66s, which greatly improves the calculation efficiency.

![Figure 3. Nodal voltage amplitude and phase angle probability distribution curve](image)

6. Conclusion

Since the traditional stochastic load flow calculation assumes that the input variables are uncorrelated, it has a certain error when used directly to calculate high-density correlation input variables. To this end, this paper proposes a random power flow algorithm that takes into account input correlation based on a hybrid Gaussian model. This method has the following characteristics: the input variables can be accurately fitted by the mixed Gaussian function; the input correlation can be transformed into the input non-correlation using the uniform sampling technique and Naf transformation; Calculating time can also reduce the truncation error of the linearized power flow equation. Finally, it is verified with IEEE-30 actual examples, which proves the superiority of the algorithm and has certain engineering practical value.

Reference

This project is funded by the Shenzhen Power Supply Bureau Co., Ltd. Science and Technology Project: "Research and Application of Key Technologies for Voltage Quality Control of Active Distribution Networks in Urban High-Density Areas" (Project Number: 090000KK52170133).
References

[1] Chen Yan, Wen Jinyu, Cheng Shijie. Probabilistic power flow calculation method considering the correlation of input variables [J]. Chinese Journal of Electrical Engineering, 2011, 31 (22): 80-87.

[2] Leite da Silva, A.M, Arienti, V.L. Probabilistic load flow by a multilinear simulation algorithm[J]. IEE Proceedings C Generation, Transmission and Distribution, 137(4):276.

[3] Feng Si. Reconstruction of distribution networks with wind power and photovoltaics taking into account uncertainty and correlations [D].

[4] Ding Ming, Li Shenghu, Huang Kai. Probabilistic power flow calculation based on Monte Carlo simulation [J]. Power System Technology, 2001 (11): 11-15 + 23.

[5] Hamed Ahmadi, Hassan Ghasemi. Probabilistic optimal power flow incorporating wind power using point estimate methods[C]// Environment and Electrical Engineering (EEEIC), 2011 10th International Conference on. IEEE, 2011.

[6] Chen X, Tung Y K. Investigation of polynomial normal transform[J]. Structural Safety, 2003, 25(4):p.423-445.

[7] Régis Lebrun, Dutfoy A. Do Rosenblatt and Nataf isoprobabilistic transformations really differ?[J]. Probabilistic Engineering Mechanics, 2009, 24(4):577-584.

[8] Huang Yu, Xu Qingshan, Tong Haihong. Semi-invariant stochastic power flow calculation based on Latin hypercube sampling technology [J]. Electric Power Automation Equipment, 2016, 036 (011): 112-119.

[9] Xu Qingshan, Huang Yu, Liu Jiankun, et al. Monte Carlo stochastic power flow method using hybrid Gaussian model and edge transform technology [J]. Automation of Electric Power Systems, 2016, 40 (16): 23-30.

[10] H. Yu, C.Y. Chung, K.P. Wong. Probabilistic Load Flow Evaluation With Hybrid Latin Hypercube Sampling and Cholesky Decomposition[J]. IEEE Transactions on Power Systems, 2009, 24(2):661-667.

[11] Billinton, R, Jonnavithula, S. A test system for teaching overall power system reliability assessment[J]. Power Systems IEEE Transactions on, 11(4):0-1676.