Design Studies and Intelligence Engineering
L.C. Jain et al. (Eds.)

doi:10.3233/FAIA231436

Power System Uncertainty Modeling Based on Gaussian Process

Tianze ZHANG

School of Computer Science and Technology, Xinjiang University, Urumqi, China

Abstract: In order to solve the problem of low accuracy of parameter estimation in expectation maximization algorithm, a modeling method based on Gaussian component number reduction is proposed. Taking the nonparametric kernel density estimation results as the base Gaussian mixture model, the Gaussian mixture model with any number of Gaussian components can be established by reducing the number of Gaussian components by using the density-preserving hierarchical expectation maximization algorithm, which overcomes the problem that the expectation maximization algorithm has low parameter estimation accuracy when there are many Gaussian components. In order to reduce the burden of modeling calculation under large samples, a hierarchical modeling method based on time scale is proposed. In order to solve the problem of Gaussian component number combination explosion of independent random variables, a hierarchical modeling method of "combination-reduction" is proposed. The proposed method is tested by using measured multidimensional wind speed data and load data with complex distribution characteristics. The experimental results show that the Pearson and Spearman correlation coefficients of GMM constructed based on this method are very close to the sample data. The absolute value of Pearson correlation coefficient error is 0.03739, and the root mean square value of error is 0.02388. The absolute value of Spearman correlation coefficient error is 0.11693, and the root mean square error is 0.05797. Conclusion: The accuracy of the proposed method is significantly better than that of Gaussian mixture model and Copula function method based on expectation maximization algorithm.

Keywords: uncertainty analysis; Gaussian mixture model; Electric power system; Multidimensional random variables; Relevance; Wind power

1. Introduction

At present, the development of power system operation control and dispatching depends on "empirical analysis" and "off-line design". The basic idea is to perceive the system situation and analyze the situation, and the dispatcher gives the operation control strategy according to his own knowledge, experience, and off-line strategy. It takes minutes for normal dispatchers to complete the above processes, so traditional manual operation can only solve the problems of system operation and scheduling. For problems that are too late to perceive, analyze and make decisions, such as equipment or system failures, it is necessary to adopt preset protection settings and stable control strategies to solve them. In the traditional power system, this kind of preset protection control depends on experience and deterministic criteria for setting, and its applicable scenarios are limited. With a large number of wind power, photovoltaic and electric vehicles connected to the power grid, the reform of the power market continues to advance, and the influence of the external environment on the power grid increases, the modern power system has strong uncertainty and complexity, and the system has many operation scenarios and strong coupling of protection and control, which leads to a decrease in the matching degree between the setting scheme and the operation scenarios and a decrease in the adaptability of the traditional methods [1-3].

In order to solve the above problems, on the one hand, it is necessary to develop a more robust protection control system, on the other hand, it is even more necessary to quickly adjust the system operation mode, so that the system operation state can be maintained within the applicable range of protection and stable control, so as to realize the system security and stability in the whole time process. However, the ability of ordinary dispatching operators to process information and analyze problems is limited. With the increasing complexity of modern power system, it is more difficult to ensure the safe and economic operation of power system. Therefore, it is the right way to solve this problem by studying the methods of big data, domain knowledge and artificial intelligence (AI) and developing an industrial artificial intelligence system suitable for power system to help dispatchers make decisions more efficiently, quickly, conveniently and accurately.

2. Literature review

The uncertainty and weak controllability of the output of new energy sources such as wind power increase the risk of transmission congestion in the system. The occurrence of transmission congestion will not only affect the realization of unit power dispatching plan and the utilization rate of new energy, but also pose a serious threat to the adequacy and reliability of the system. At present, most congestion management methods for power systems with new energy sources are from the planning level, mainly by improving the topology of medium and long-term transmission networks, improving the weak links and transmission capacity limits of the system. From the short-term operation level, the research on congestion scheduling management methods can be mainly divided into two categories.

The related research thinks that the uncertainty of wind power can be dealt with by increasing the positive and negative rotating reserve capacity, and the reserve capacity demand of the system is set to 20% of the predicted wind power output [4]. In addition, considering the distribution characteristics of new energy forecasting error and reserve constraints, combined with the forced outage rate of the unit, the required rotating reserve capacity is accurately quantified [5]. Shen and J divide the system into regions according to the similarity of transmission congestion distribution factors, define "congestion regions" and ensure the adequacy of spare capacity in each region, so as to reduce the congestion risk of lines in real-time operation [6]. Based on the risk optimization theory, Fan Fan, M put forward Dynamic Reserve at Risk,DRaR) and conditional reserve at risk (DCRAR) as evaluation indexes of system operation adequacy, and quantified the reserve demand level in each period to make the optimal decision-making scheme [7]. Zhang and F describe the uncertainty of wind power through multiple scenarios, and construct a two-stage stochastic planning and dispatching model, thus realizing the overall simulation of unit output planning and real-time power adjustment [8]. Tong, Z, based on the computing framework of SP, established a cooperative optimal scheduling model of source and load with the objective function of minimizing the expected total cost of the system. Because the optimal solution is closely related to the specific scenes generated, it is usually necessary to simulate a large number of scenes to ensure the accuracy of the optimal solution [9]. Therefore, it is proposed to simplify the original scene by scene reduction technology, and replace the original sample set with this subset for calculation [10].

This paper analyzes the problem that the parameter estimation accuracy of multidimensional Gaussian mixture model, GMM) based on expectation-maximization (EM) algorithm is not high. A GMM modeling method based on nonparametric kernel density estimation and density-preserving hierarchical expectation maximization (KDE-DPHEM) is proposed. Based on the results of multidimensional kernel density estimation, KDE), GMM modeling is realized by reducing the number of Gaussian components, and the accuracy of GMM is significantly improved by increasing the number of Gaussian components, which overcomes the problem that EM algorithm is difficult to obtain high-precision GMM when there are many Gaussian components. On this basis, in order to reduce the computational burden of probabilistic modeling under large samples, a hierarchical modeling method based on time scale is proposed. In order to solve the problem of Gaussian component number combination explosion after multiple independent random variables are combined, a hierarchical modeling method of "combination-reduction" is proposed. Based on KDE-DPHEM algorithm, this paper systematically puts forward the modeling method of power system probability model based on multi-dimensional GMM, and comprehensively tests the proposed method by using the actually collected multi-dimensional wind speed and multi-dimensional load data. The results verify the effectiveness and advantages of the proposed method.

3. Research methods

3.1. GMM method based on KDE-DPHEM algorithm

Using KDE method to model probability density function does not need prior information of random variable distribution type, and it can be used for arbitrary d Dimensional random variable x, based on n The joint probability density function estimated by KDE method is shown in Formula (1) [11-13].

$$f(x)=\frac{1}{n}\sum_{i=1}^{n}K_{\mathrm{H}}\left(x-x_i\right) \tag{1}$$

Where: f(x) for x Joint probability density function of; x_i for x DediiSamples; n Is the number of samples; KHIs a kernel function.

when K_H When it is a Gaussian kernel function, KDE obtains a special GMM with the number of Gaussian components equal to the number of samples, and takes this model as the base GMM, as shown in Formula (2).

$$f^{(\mathrm{b})}(\mathbf{x})=\sum_{i=1}^{n}\frac{1}{n}\mathrm{N}(\mathbf{x} \mid \mathrm{x}_{\mathrm{i}},\mathbf{H}_{\mathrm{i}})=\sum_{\mathrm{i}=1}^{K_{\mathrm{b}}}\omega_{\mathrm{i}}^{(\mathrm{b})}\mathrm{N}\left(\mathrm{x} \mid \mu_{\mathrm{i}}^{(\mathrm{b})},\Sigma_{\mathrm{i}}^{(\mathrm{b})}\right) \tag{2}$$

Where: $f^{(b)}(x)$ Based on GMMx Joint probability density function of; K_b Is the number of Gaussian components of the base GMM, which is equal ton; $\omega_i^{(b)}$ Weidii The weight of a Gaussian component is equal to1/n; $\mu_i^{(b)}$ Weidii The mean vector of Gaussian components is equal to the first.i Samplex_i; $\Sigma_i^{(b)}$ Weidii The covariance matrix of Gaussian components is equal to the bandwidth matrix.H_i.

When the samples are sufficient and the bandwidth is appropriate, the base GMM based on KDE method is the most accurate GMM, and this model is often used as a benchmark model to evaluate the accuracy of parameter estimation methods. The Gaussian component number of base GMM is extremely large, which is not practical in the subsequent probabilistic uncertainty analysis of power system. Further, the basis GMM of the formula (2) is reduced to the basis GMM of the formula (3) $K_r(K_r<K_b)$The simplified GMM of Gaussian components adopts DPHEM algorithm in the reduction process.

$$f^{(\mathrm{r})}(x)=\sum_{j=1}^{K_{\mathrm{r}}}\omega_j^{(\mathrm{r})}\mathbf{N}\left(x|\ \boldsymbol{\mu}_j^{(\mathrm{r})},\boldsymbol{\Sigma}_j^{(\mathrm{r})}\right) \tag{3}$$

DPHEM algorithm transforms the reduction of base GMM into the following optimization problem: satisfy the base model for probability distribution $f^{(b)}(x)$ Virtual sample set based onX, beg$f^{(b)}(x)$ Parameters to simplify the model.$f^{(b)}(x)$ The expected value of log-likelihood function of is the largest, as shown in Equation (4).

$$\underset{\left(\omega_j^{(n)},\mu_j^{(p)},\Sigma_j^{(\mathrm{r})}\right)}{\arg}\ \max\mathbb{E}_{X-f^{(\mathrm{b})}(x)}\left[\ln f^{(\mathrm{r})}(X)\right] \tag{4}$$

Where: $f^{(r)}(x)$ To simplify GMMx Joint probability density function of; $\omega_j^{(r)}$ Weidij The weight of Gaussian components; $\mu_{\mathrm{j}}^{(\mathrm{r})}$ Weidij Mean vector of Gaussian components; $\Sigma_{\mathrm{j}}^{(\mathrm{r})}$ Weidij Covariance matrix of Gaussian components; K_r In order to simplify the Gaussian component number of GMM; Virtual sample setXbym Virtual samples $\{X_1,X_2,\ldots,X_m\}$ Composition, in the actual solution process does not need to really generate the sample; $\mathbb{E}$For expectation.

Equation (4) is solved by variational EM algorithm. The iterative steps of DPHEM algorithm are basically the same as those of conventional EM algorithm, which are divided into E-step and M-step, as shown in equations (5)-(9).

E-step:

$$E_{i,j}=\ln\left[\mathbf{N}\left(\mu_i^{(\mathrm{b})}|\ \mu_j^{(\mathrm{r})},\Sigma_j^{(\mathrm{r})}\right)\right]-\frac{1}{2}\mathrm{tr}\left[\left(\Sigma_j^{(\mathrm{r})}\right)^{-1}\Sigma_i^{(\mathrm{b})}\right] \tag{5}$$

$$z_{ij}=\frac{\omega_j^{(r)}\exp\left(mE_{i,j}\right)}{\sum_{l=1}^{K_r}\omega_j^{(r)}\exp\left(mE_{i,l}\right)} \tag{6}$$

M-step:

$$\omega_j^{(\mathrm{r})}=\sum_{i=1}^{K_{\mathrm{b}}}z_{ij}\omega_i^{(\mathrm{b})} \tag{7}$$

$$\mu_j^{(\mathrm{r})} = \frac{1}{\omega_j^{(\mathrm{r})}} \sum_{i=1}^{K_\mathrm{b}} z_y \omega_i^{(\mathrm{b})} \mu_i^{(\mathrm{b})} \tag{8}$$

$$\Sigma_j^{(\mathrm{s})} = \frac{1}{\omega_j^{(\mathrm{r})}} \sum_{i=1}^{K_\mathrm{b}} z_{ij} \omega_i^{(\mathrm{b})} \left[\Sigma_i^{(\mathrm{b})} + \left(\mu_i^{(\mathrm{b})} - \mu_j^{(\mathrm{r})} \right) \left(\mu_i^{(\mathrm{b})} - \mu_j^{(\mathrm{r})} \right)^\mathrm{T} \right] \tag{9}$$

Where: Z_{ij} Is a variational parameter; subscripti $\in[1,K_b]$; subscriptl、j $\in[1,K_r]$; tr Is the trace of the matrix; The number of virtual samples m is generally 10 times.K_b; Other symbols have the same meanings as Formula (2)-(4).

Because the basis GMM in this paper is obtained by KDE method, large numbersm Will lead to the formula (6) exp ($mE_{i,j}$)The numerical value overflows, and there is a numerical problem of 0/0. Based on equations (5) and (6), this paper proposes an improved E-step, as shown in equations (10)-(12). Compared with the Log-Sum-Exp technique.ln(z_{ij}) , and then calculate exp [ln(z_{ij})]The algorithm in this paper has higher computational efficiency and better numerical stability [14-15].

$$\begin{aligned} A_{i,j} = \quad & \frac{1}{2} tr\left[\left(\Sigma_j^{(\mathrm{r})} \right)^{-1} \Sigma_i^{(\mathrm{b})} \right] + \frac{1}{2} \ln \left| \Sigma_j^{(\mathrm{r})} \right| + \\ & \frac{1}{2} \left(\mu_i^{(\mathrm{b})} - \mu_j^{(\mathrm{r})} \right)^\mathrm{T} \left(\Sigma_j^{(\mathrm{r})} \right)^{-1} \left(\mu_i^{(\mathrm{b})} - \mu_j^{(\mathrm{r})} \right) \end{aligned} \tag{10}$$

$$l_{i,\min} = \arg\min_l \left(\Lambda_{i,l} \right), l = 1, 2, \cdots, K_\mathrm{r} \tag{11}$$

$$z_{ij} = \frac{\omega_j^{(\mathrm{r})} \exp\left[m \left(\Lambda_{i,l_{,\min}} - \Lambda_{i,j} \right) \right]}{\sum_{l=1}^{K_1} \omega_l^{(\mathrm{r})} \exp\left[m \left(\Lambda_{i,l_{,\min}} - \Lambda_{i,l} \right) \right]} \tag{12}$$

KDE and DPHEM constitute a two-stage KDEDPHEM algorithm, which can realize the modeling of any Gaussian component GMM and overcome the bottleneck problem of parameter estimation of EM algorithm for more Gaussian components GMM.

3.2. Hierarchical modeling of time scale for large samples

The sample size of power data is increasing. In order to reduce the computational burden of modeling, a hierarchical modeling method based on GMM model characteristics and DPHEM algorithm is further proposed. Taking the two time scales of month and year as an example, the monthly scale probability model is first constructed, and each monthly scale sample is the conditional probability sample of the annual scale sample. Further, the 12-month scale model is accumulated into the annual scale model with the proportion of the number of samples in each month as the weight. Fig. 1 is a schematic diagram of the modeling flow of a two-layer structure. Suppose a sample X The sample size of is n, according to the monthly scale samples. X be split into M Block, each sample is recorded asX_{b1}、X_{b2}、 … 、X_{bM} , the number of samples per block is: n_1、n_2、 … 、n_M. GMM is established for each sample by KDE-DPHEM algorithm, and the number of Gaussian components of GMM for each sample is as follows:K_{b1}、K_{b2}、 … 、K_{bM}, the first M The joint probability density function of the blocks is shown in Formula (13). According to the total probability formula, X The joint probability density function of is composed of X_{b1}、X_{b2}、 … 、X_{bM} The joint probability density function and the corresponding sample number ratio are accumulated as weight coefficients, and the annual scale based GMM model is shown in Formula (14). Gauss component number of formula (14) $K = Kb_1 + Kb_2 + \cdots + K_{bM}$; such as K If it is too large, further reduction is carried out by using DPHEM algorithm, and the annual scale simplified GMM shown in equation (15) is obtained. Consider that more general case, the sample X Divided into multiple layers according to the time scale, we only need to model the bottom time scale directly, and then aggregate from the bottom time scale layer by layer until the top time scale. The time scale layering method not only reduces the computational burden of large sample modeling, but also obtains multi-time scale probability models, and the modeling process can be parallelized [16-19].

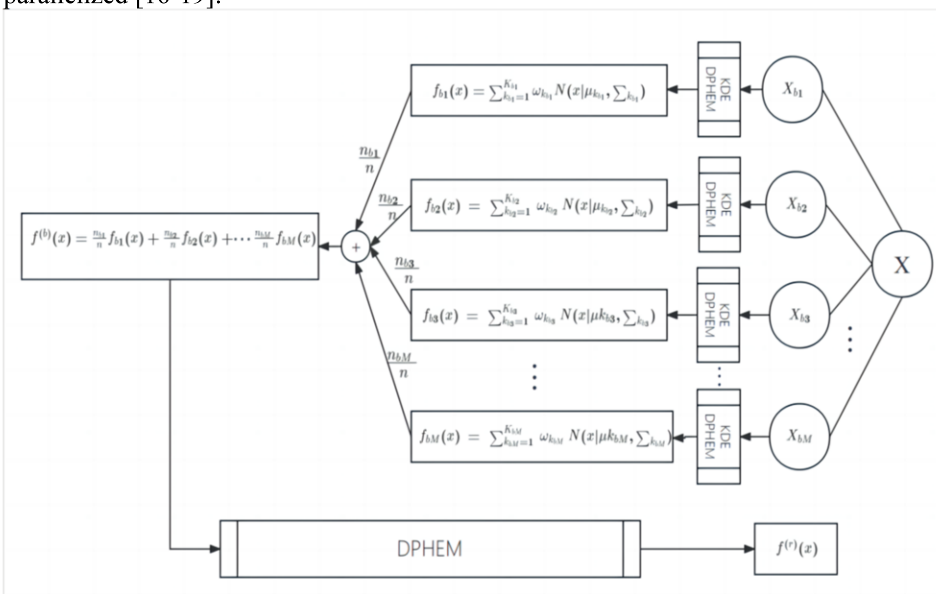

Fig. 1 Time-scale hierarchical modeling method for large samples

$$f_{\mathrm{b}M}(\mathbf{x}) = \sum_{k_{\mathrm{b}M}-1}^{K_{\mathrm{bM}}} \omega_{k_{\mathrm{bM}}} \mathrm{N}\left(\mathbf{x} \middle| \ \mu_{k_{\mathrm{bM}}}, \Sigma_{k_{\mathrm{bM}}}\right) \tag{13}$$

$$f^{(\mathrm{b})}(\mathbf{x}) = \frac{n_1}{n} f_{\mathrm{b1}}(\mathbf{x}) + \frac{n_2}{n} f_{\mathrm{b2}}(\mathbf{x}) + \cdots \frac{n_M}{n} f_{\mathrm{b}M}(\mathbf{x}) \tag{14}$$

$$f^{(\mathrm{r})}(\mathbf{x}) = \mathrm{DPHEM}\left[f^{(\mathrm{b})}(\mathbf{x}) \middle| \right] \tag{15}$$

3.3. "Combination-Reduction" Hierarchical Modeling of Combination Explosion

In the actual power system, variables may be correlated or independent. Assume that the system has P Group multidimensional random variables $\xi^{(1)}$、$\xi^{(2)}$、 … 、$\xi^{(P)}$ Groups are independent of each other, forming a multidimensional random variable. ξAs shown in formula (16) [20].

$$\boldsymbol{\xi} = \left[\boldsymbol{\xi}^{(1)\mathrm{T}}, \boldsymbol{\xi}^{(2)\mathrm{T}}, \cdots, \boldsymbol{\xi}^{(P)\mathrm{T}}\right]^{\mathrm{T}} \tag{16}$$

Arbitrary firstp Group random variable $\xi^{(p)}$ The joint probability density function of is α_p The GMM model composed of 10 Gaussian components is shown in Formula (17).

$$f\left(\xi^{(p)}\right) = \sum_{k=1}^{\alpha_p} \omega_k^{(p)} \mathbf{N}\left(\xi^{(p)} \mid \boldsymbol{\mu}_k^{(p)}, \boldsymbol{\Sigma}_k^{(p)}\right) \tag{17}$$

Where: $f\left(\xi^{(p)}\right)$ for $\boldsymbol{\xi}^{(p)}$ Joint probability density function based on GMM; $\omega_k^{(p)}$ for $f\left(\xi^{(p)}\right)$ sequencek The weight of Gaussian components; $\boldsymbol{\mu}_k^{(p)}$ for $f\left(\xi^{(p)}\right)$ sequencek Mean vector of Gaussian components; $\boldsymbol{\Sigma}_k^{(p)}$ for $f\left(\xi^{(p)}\right)$ sequence k Covariance matrix of Gaussian components.

According to probability theory, ξThe joint probability density function of can be derived as shown in formula (18), and the number of Gaussian components is α. The corresponding Gaussian component parameters are shown in equations (19)-(21).

$$f(\boldsymbol{\xi}) = \sum_{k=1}^{\alpha} \omega_k \mathrm{N}\left(\boldsymbol{\xi} \mid \boldsymbol{\mu}_{\mathrm{k}}, \boldsymbol{\Sigma}_{\mathrm{k}}\right), \alpha = \prod_{1}^{\mathrm{P}} \alpha_{\mathrm{p}} \tag{18}$$

$$\mu_{l_1 l_2 \cdots l_P} = \left[\mu_{l_1}^{(1)\mathrm{T}}, \mu_{l_2}^{(2)\mathrm{T}}, \cdots \mu_{l_P}^{(P)\mathrm{T}}\right]^{\mathrm{T}} \tag{19}$$

$$\Sigma_{l_1 l_2 \cdots l_p} = \begin{bmatrix} \Sigma_{l_1}^{(1)} & 0 & 0 & 0 \\ 0 & \Sigma_{l_2}^{(2)} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \Sigma_{l_p}^{(P)} \end{bmatrix} \tag{20}$$

$$\omega_{l_1 l_2 \cdots l_P} = \omega_{l_1}^{(1)} \omega_{l_2}^{(2)} \cdots \omega_{l_P}^{(P)} \tag{21}$$

Where: f(ξ) for ξ Joint probability density function based on GMM; ω_k for f(ξ) sequencek The weight of Gaussian components; $\boldsymbol{\mu}_k$ for f(ξ) sequence k Mean vector of Gaussian components; Σ_k for f(ξ) sequence k Covariance matrix of Gaussian components; lp ∈[1, αp], p ∈[1, P].

ξGaussian component number of α Will follow P Increasing the problem of combined explosion leads to a great amount of calculation in the analytical method of probability uncertainty analysis of power system. Taking the independent load of 10 nodes as an example, that is, the dimension of each group of random variables is 1, assuming that each load GMM model only uses 4 Gaussian components, the combination number of 410 will make the analytical method lose its efficiency advantage. In the scene with more complicated probability distribution and more nodes, the combined Gaussian components will be astronomical.

Aiming at the above-mentioned combined explosion problem, a hierarchical modeling method of "combination-reduction" is proposed based on GMM model characteristics and DPHEM algorithm. Only two groups of random variables are combined at a time: $f(\xi^{(1)})$ and $f(\xi^{(2)})$ Combine $f(\xi^{(1,2)})$; $f(\xi^{(3)})$ and $f(\xi^{(4)})$ Combine $f(\xi^{(3,4)})$; …; $f(\xi^{(P-1)})$ and $f(\xi^{(P)})$ Combine $f(\xi^{(P-1,P)})$. After combination $f(\xi^{(1,2)})$、$f(\xi^{(3,4)})$、…$f(\xi^{(P-1,P)})$ It is still a multi-dimensional GMM model and still independent of each other, and the above process can be further repeated. Before combining again, the last combination result is reduced by DPHEM algorithm, that is $f(\xi^{(1,2)})$、$f(\xi^{(3,4)})$、…$f(\xi^{(P-1,P)})$ Reduce to $f^{(r)}(\xi^{(1,2)})$、$f^{(r)}(\xi^{(3,4)})$、…$f^{(r)}(\xi^{(P-1,P)})$Then combine in pairs, so as to avoid combination explosion. For P groups of multidimensional random variables, the above method only needs P-1 "combination-reduction" operation, and the computational complexity is linear with P. The reduction step of the above method will produce a certain precision loss. The smaller the error between the joint cumulative probability of the obtained model and the original model, and the closer the Pearson correlation coefficient of any two groups of independent random variables is to zero, the smaller the precision loss is. The intermediate GMM in the reduction process and the Gaussian component number of the GMM finally obtained can be configured according to the accuracy requirements and computing power.

4. Result analysis

4.1. performance verification of GMM method based on KDE-DPHEM algorithm

The wind speed of NaselleRidge and Megler in BPA Power Bureau in January 2016 was taken as the test data, and the wind power output was converted from wind speed according to IEC Class II fan model. NaselleRidge and Megler are about 18km apart, and the wind speed has a significant correlation. The joint probability density functions of wind speed and wind power are modeled respectively, and the errors of GMM method based on EM algorithm, GMM method based on KDE-DPHEM algorithm and Copula function method are compared with the KDE results. GMM methods based on EM and KDE-DPHEM algorithms are tested by different Gaussian components. The Copula function method adopts Gaussian, T, Clayton, Frank, Gumbel and many kinds of mixed Copula respectively. The root-mean-square error, RMSE) of joint probability density function is used for comparison, and 40,000 sampling points are evenly divided into 200×200 grids in the defined domain, and the root-mean-square error is obtained by statistical calculation of each point error.

The optimal Copula is t-Frank-Gumbel mixed Copula. When the number of Gaussian components is less than 10, the GMM errors based on EM algorithm and KDE-DPHEM algorithm are greater than the optimal Copula function. When the number of Gaussian components is greater than 10, the GMM error based on EM algorithm decreases slowly with the increase of Gaussian components. When the number of Gaussian components is greater than 20, the error level of GMM based on EM algorithm has not been significantly improved. When the number of Gaussian components is greater than 20, the error of GMM based on KDE-DPHEM algorithm is obviously lower than that of EM algorithm, and the error is obviously reduced with the increase of Gaussian components. When the number of Gaussian components is greater than 40, the error is less than half of the optimal Copula function. When the number of

Gaussian components is increased to 100, the error is only 18.54% of the optimal Copula and 20.56% of the EM algorithm.

The joint probability density function of wind speed and wind power output obtained by different methods, the Gaussian component number of GMM is 100. In the GMM modeling of wind power output based on EM algorithm, the covariance matrix is ill-conditioned in the iterative process, and the GMM result based on EM algorithm is better before the ill-conditioned problem appears. The whole and details of GMM based on KDE-DPHEM algorithm are highly similar to those of KDE and histogram, and the density characteristics of each local area are described in detail. The outer contour of GMM results based on EM algorithm is similar to that of KDE, but the detail error is large. The high-density region of t-Frank-Gumbel mixed Copula is similar to that of KDE and histogram, but it also has the problems of overall similarity and large detail error. Compared with KDE and histogram, Gaussian Copula is quite different, which shows that Gaussian Copula is not suitable for describing the correlation structure of wind speed and wind power output in NaselleRidge and Megler. The above results show that GMM method based on KDE-DPHEM algorithm has the advantage of high accuracy, and the modeling accuracy is significantly better than GMM method based on EM algorithm and the optimal Copula function method.

4.2. Verification of Time Scale Hierarchical Modeling Method

The 5-minute wind speed of NaselleRidge and Megler wind stations from January 2015 to December 2017 is selected as the test data, and the joint probability distribution of wind power output of the two stations is taken as the modeling goal. The total sample is about 315,000 sets of two-dimensional data. The data are divided into 12 blocks by month, and each sample has an average of about 26,000 sets of data. The time-scale hierarchical method is used to model. Based on the results of KDE, the annual scale GMM obtained by time scale layering method is compared with the annual scale GMM obtained by direct modeling, and the k of annual scale GMM and monthly scale GMM are set to 100.

The error analysis of the joint probability density function obtained by modeling is shown in Table 1. The results obtained by time scale method have high similarity with KDE and histogram, but little difference with direct method. Table 1 shows that the error of the time-scale hierarchical method is in the same order of magnitude as that of the direct method, the calculation time required is 44.39% of that of the direct method, and the amount of data to be processed by the single DPHEM algorithm is about 1/12 of that of the direct method, and the calculation complexity is significantly reduced. The 12-month monthly scale probability model obtained by time scale stratification method and its aggregated annual scale basis GMM belong to the intermediate model of time scale stratification method. The monthly scale model can be used for the probability analysis of monthly scale; The error of annual scale basis GMM is about half of the final simplified GMM, which can be used when higher accuracy is required.

Table 1 Comparison of annual scale model errors and time consumption of different methods

way	Root Mean Square Error (RMSE)		Time /s
	probability density	cumulative probability	
YB model	3.17×10^{-2}	3.05×10^{-4}	239.41
YR model	6.39×10^{-2}	8.08×10^{-4}	233.02
YD model	8.56×10^{-2}	6.22×10^{-4}	482.17

4.3. High-dimensional correlation modeling verification

In order to further verify the high-dimensional modeling ability of the method proposed in this paper, all related random variables in IEEE33-node distribution network are modeled by multi-dimensional GMM. The correlation load data of 32 nodes connected with load in IEEE33-node distribution network consists of 192 residential loads superimposed in turn every 6 households, and scaled to the average value consistent with the original deterministic load; Four groups of wind turbines with rated power of 0.4MW are connected to four different nodes, and the relevant wind power output is obtained by converting the wind speed of four wind stations (Naselle Ridge, Megler, Troutdale and BiddleButte) of BPA Power Bureau according to the method in Section 4.1. Due to a large number of missing wind speed data in the first half of 2010, the load and wind power data are unified in the second half of 2010, with a data resolution of 10min, about 26,000 pieces, and a data dimension of 36 dimensions. Random variables in each dimension are recorded as x_1、x_2、…、x_{36}, x_1—x_{32} Corresponding to the active load of 2-33 nodes, x_{33}—x_{36} Corresponding to the active output of four groups of wind power, the data takes the load as positive and the output as negative.

The computer configuration used for modeling is Intel Core i7-10510UCPU and 16GB memory. The time scale layering method is adopted, and the Gaussian component number of the monthly scale GMM is set to 1000, and the Gaussian component number of the semi-annual scale GMM is set to 100. The total modeling time is 608.02s. Pearson and Spearman correlation coefficients of GMM constructed based on this method are very close to the sample data. The absolute value of Pearson correlation coefficient error is 0.03739, and the root mean square value of error is 0.02388. The maximum absolute error of Spearman correlation coefficient is 0.11693, and the root mean square error is 0.05797, which proves that the method proposed in this paper has good high-dimensional correlation modeling ability. In the actual larger-scale system, the correlation random variable has regional characteristics, and the correlation within the region is significant, but the correlation between regions is low. At this time, it can be modeled by regions, and then a higher-dimensional GMM model of the whole system can be formed by the "combination-reduction" method.

5. Conclusion

In this paper, based on KDE-DPHEM algorithm, the probabilistic modeling method of power system based on multidimensional GMM is systematically improved, and the effectiveness of the proposed method is verified by an example test. The main conclusions are as follows:

1) KDE-DPHEM algorithm can continuously improve the accuracy of GMM model with the increase of Gaussian components, which overcomes the bottleneck that the accuracy of EM algorithm is difficult to improve after the number of Gaussian components is greater than 10. By increasing the number of Gaussian components, GMM model approaching KDE results can be obtained, and the modeling accuracy is significantly higher than that of Copula function method and EM algorithm.

2) The hierarchical modeling method of time scale effectively reduces the computational burden of GMM modeling under large sample data and can obtain multi-time scale intermediate models.

3) The hierarchical modeling method of "combination-reduction" effectively solves the problem of Gaussian component number explosion after the combination of independent random variables.

4) The proposed method accurately describes the complex correlation structure of high-dimensional random variables, and the large-scale correlation coefficient matrix of the high-dimensional model is close to the actual value.

5) The proposed method studies the modeling method of GMM with a given number of Gaussian components. According to the experience of an example, setting the number of Gaussian components to 100 orders of magnitude can get a higher precision result. In the future, we will further study how to balance the number of Gaussian components and model accuracy and realize the adaptive configuration of Gaussian components.

References

[1] He, Y., Li, H., Wang, S., & Yao, X. (2021). Uncertainty analysis of wind power probability density forecasting based on cubic spline interpolation and support vector quantile regression. Neurocomputing, 430, 121-137.

[2] Gu, B., Zhang, T., Meng, H., & Zhang, J. (2021). Short-term forecasting and uncertainty analysis of wind power based on long short-term memory, cloud model and non-parametric kernel density estimation. Renewable Energy, 164, 687-708.

[3] Dall'Armi, C., Pivetta, D., & Taccani, R. (2022). Uncertainty analysis of the optimal health-conscious operation of a hybrid PEMFC coastal ferry. International Journal of Hydrogen Energy, 47(21), 11428-11440.

[4] Petkov, I., & Gabrielli, P. (2020). Power-to-hydrogen as seasonal energy storage: an uncertainty analysis for optimal design of low-carbon multi-energy systems. Applied Energy, 274, 115197.

[5] Zhao, X., Ge, C., Ji, F., & Liu, Y. (2021). Monte Carlo method and quantile regression for uncertainty analysis of wind power forecasting based on Chaos-LS-SVM. International Journal of Control, Automation and Systems, 19, 3731-3740.

[6] Tong, Z., Chen, X., Tong, S., & Yang, Q. (2022). Dense Residual LSTM-Attention Network for Boiler Steam Temperature Prediction with Uncertainty Analysis. ACS omega, 7(13), 11422-11429.

[7] Zhang, F., Wang, X., Wang, W., Zhang, J., Du, R., Li, B., & Liu, W. (2023). Uncertainty analysis of photovoltaic cells to determine probability of functional failure. Applied Energy, 332, 120495.

[8] Fan, M.,Li, Z., Ding, T., Huang, L., Dong, F., Ren, Z., & Liu, C. (2021). Uncertainty evaluation algorithm in power system dynamic analysis with correlated renewable energy sources. IEEE Transactions on Power Systems, 36(6), 5602-5611.

[9] Shen, J., ZWang, Q., Ma, Z., & He, Y. (2021). Nonlinear optimization strategy for state of power estimation of lithium-ion batteries: A systematical uncertainty analysis of key impact parameters. IEEE Transactions on Industrial Informatics, 18(10), 6680-6689.

[10] Hosseini, S. H. R., Allahham, A., Walker, S. L., & Taylor, P. (2021). Uncertainty analysis of the impact of increasing levels of gas and electricity network integration and storage on Techno-Economic-Environmental performance. Energy, 222, 119968.

[11] Medrano, M., Liu, R., Zhao, T., Webb, T., Politte, D. G., Whiting, B. R., ... & Williamson, J. F. (2022). Towards subpercentage uncertainty proton stopping‐power mapping via dual‐energy CT: Direct experimental validation and uncertainty analysis of a statistical iterative image reconstruction method. Medical physics, 49(3), 1599-1618.

[12] Dash, R. C., Sharma, N., Maiti, D. K., & Singh, B. N. (2022). Uncertainty analysis of galloping based piezoelectric energy harvester system using polynomial neural network. Journal of Intelligent Material Systems and Structures, 33(16), 2019-2032.

[13] Yin, T., Li, W. T., Li, K., & He, Z. Z. (2021). Multi-parameter optimization and uncertainty analysis of multi-stage thermoelectric generator with temperature-dependent materials. Energy Reports, 7, 7212-7223.

[14] Deng, H., Chen, W., Cao, D., Chen, J., & Hu, W. (2020). Uncertainty analysis and robust control of fuel delivery systems considering nitrogen crossover phenomenon. International Journal of Hydrogen Energy, 45(56), 32367-32387.

[15] Shen, X., Zhang, Y., Zhang, J., & Wu, X. (2022). An Interval Analysis Scheme Based on Empirical Error and MCMC to Quantify Uncertainty of Wind Speed. IEEE Transactions on Industry Applications, 58(6), 7754-7763.

[16] Zhang, H., Liu, Y., Yan, J., Han, S., Li, L., & Long, Q. (2020). Improved deep mixture density network for regional wind power probabilistic forecasting. IEEE Transactions on Power Systems, 35(4), 2549-2560.

[17] Petkov, I., & Gabrielli, P. (2020). Power-to-hydrogen as seasonal energy storage: an uncertainty analysis for optimal design of low-carbon multi-energy systems. Applied Energy, 274, 115197.

[18] Zhou, W., Zhao, Z., Nielsen, J. B., Fritsche, L. G., LeFaive, J., Gagliano Taliun, S. A., ... & Lee, S. (2020). Scalable generalized linear mixed model for region-based association tests in large biobanks and cohorts. Nature genetics, 52(6), 634-639.

[19] Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., ... & Nahavandi, S. (2021). A review of uncertainty quantification in deep learning: Techniques, applications and challenges. Information fusion, 76, 243-297.

[20] Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H., Teplitsky, C., ... & Araya‐Ajoy, Y. G. (2020). Robustness of linear mixed‐effects models to violations of distributional assumptions. Methods in ecology and evolution, 11(9), 1141-1152.