Data-Driven Nested Robust Optimization for Generation Maintenance Scheduling Considering Temporal Correlation

Xiao Yang[a], Yuanzheng Li[a,*], Yong Zhao[a], Yaowen Yu[a], Yicheng Lian[a], Guokai Hao[a] and Lin Jiang[b]

[a]School of Artificial Intelligence and Automation, Key Laboratory on Image Information Processing and Intelligent Control of Ministry of Education, Huazhong University of Science and Technology, Wuhan 430074, China

[b]Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool L69 3GJ, United Kingdom

ARTICLE INFO

Keywords:
Generation maintenance scheduling
Nested robust optimization
Temporal correlation Dirichlet process mixture model
Data-driven uncertainty set

ABSTRACT

Traditional power grids are gradually transitioning to smart grids with high penetration of renewable energy, which can realize the efficient utilization of power resources and low carbon emissions. However, the uncertainties of renewable energy (e.g., wind power) and load demand pose considerable challenges to secure operation and cost-effective planning in smart grids, such as generation maintenance scheduling (GMS). In this context, conventional methods including stochastic optimization and robust optimization are adopted to cope with the uncertainties and formulate the GMS plan. Unfortunately, these methods fail to consider the temporal information in uncertain variables, which can introduce extra operational costs brought by the uncertainties. To address this issue, we consider the temporal correlation of the uncertain wind power and load demand, and develop a data-driven two-stage nested robust optimization (NRO) approach for GMS to minimize the total costs of power system operation under uncertain scenarios. In our proposed approach, a temporal correlation Dirichlet process mixture model (TCDPMM) is developed to investigate the temporal information in the wind power and load demand datasets. Then, variational Bayesian inference (VBI) is employed to construct the data-driven uncertainty set, in which the temporal information for the uncertain variables and the correlations between the uncertain variables are considered. Subsequently, combined with this uncertainty set, a two-stage GMS problem is converted to a "min-max-max-min" optimization problem which is solved by the parallel Benders' decomposition algorithm. The effectiveness and superiority of the proposed approach are demonstrated with a six-bus power system and a practical power system in China.

1. Introduction

Generation maintenance scheduling (GMS) is one of the core tasks for power systems operation, which can prolong the useful life of power equipment and improve the utilization of power resources [1, 2]. According to the annual distribution characteristics of load demand, GMS can reasonably arrange the maintenance commitment plan and outage stage of power generation and transmission equipment, which is of great significance for ensuring the economy and security of power system operations.

Recently, GMS has attracted the attention of scholars to improve the flexibility of power systems operation. Considering the benefits of generation companies and independent system operators, a non-cooperative dynamic game-based coordination model for electricity markets is proposed in [3]. Furthermore, Ref. [4] presents an integrated generation and transmission maintenance scheduling model combined with N-1 contingency constraints, which improves the security of the power system operation. On this basis, to ensure the economy of the system, the authors in [5] develop a dynamic game-based bi-level approach, in which a maintenance scheduling mechanism for the integrated energy systems is proposed and the overall revenues are maximized. In general, the above studies formulate reasonable maintenance scheduling plans. However, the uncertainties in power systems (e.g., renewable power and load demand) are ignored, which is disadvantageous [6, 7]. Actually, as the proportion of renewable power in power grids and social power consumption increase year by year, the uncertainty of renewable power (such as wind power) and load demands can bring adverse impacts to power systems operation, such as voltage fluctuation and power flow change, which should not be ignored in GMS.

For formulating the GMS plan with uncertainties, current methods can be divided into stochastic programming (SP) and robust optimization (RO). In SP, the probability distributions of the uncertain variables are assumed to be known. For instance, wind power is usually considered to follow the Weibull distribution [8]. In [9], an integrated stochastic optimization model is proposed to jointly optimize the operation and maintenance of the power system, and the uncertainties from prices, demands and renewable energy generation are explicitly characterized via different scenarios. Considering the uncertainty of renewable energy resources, Ref. [10] formulates a stochastic multi-stage maintenance scheduling model for the active distribution network, which is solved by the stochastic Monte Carlo tree search method. In the above mentioned studies, the uncertain variables such as wind power and load demand are assumed to follow the specific distributions, and their uncertainties are analogized by the large number of scenarios generated by the scenario generation method. Then, SP transforms the uncertain GMS problem into the deterministic one by

*Corresponding author
Yuanzheng_Li@hust.edu.cn (Y. Li)

solving the expected value of the objective function for these scenarios. However, the exact probability distribution of the uncertainty variables in practice is difficult to be obtained directly, and a large number of scenarios sharply increases the computational burden.

To address these issues, robust optimization (RO) is applied to the GMS of power systems, which characterizes the randomness by the boundaries of the uncertainty variables and makes the optimal decision under the worst-case scenario. Ref. [11] develops a robust maintenance scheduling optimization model, in which the wind power uncertainty set is constructed and the forced outages of equipment are also considered. On this basis, considering the uncertainty of load demand, wind power generation and equipment unavailabilities, a reliability constraint based tri-level adaptive robust multi-resolution model for the GMS problem is proposed, and a robust maintenance scheduling for units immunized against the worst case is obtained [12]. In [13], a two-stage robust optimization model for the long-term GMS of a power system is formulated, in which the uncertainties including the load demand, the contingencies in the transmission network and the marginal cost are taken into account. Ref. [14] incorporates the uncertainty of load demand into a multi-scale multi-resolution GMS model based on a stochastic affinely adjustable robust approach. Compared with the SP methods, the above RO methods for GMS do not need to know the specific distribution of the uncertainty variables, and the solution efficiency is also improved. However, there are some drawbacks to the RO methods: 1) relying on the boundary information of the uncertain variables, the results are too conservative, and 2) the probability information of the uncertainty variables is not fully considered. Therefore, the RO method will be limited in making optimal decisions under the worst-case scenario when the uncertain variables dataset, such as wind power and load demand, has the characteristics of high dimensional and data correlation. It is necessary that new methods need to be found to deal with the uncertainties in GMS.

In recent years, it is widely developed that machine learning methods are applied to explore the internal information in the uncertainty dataset [15]. In general, machine learning methods include Gaussian mixture model (GMM), Dirichlet process mixture model (DPMM), Wasserstein metric, generative adversarial network (GAN) and so on. Ref. [16] implements the uncertainty modeling of multivariate wind power based on GMM, and the multi-peaked characteristics of the probability distribution for wind power are portrayed. Further, Ref. [17] utilizes DPMM to fit the probability distribution of wind power and automatically obtains Gaussian mixture components from the updated data. Moreover, Wasserstein metric is used to find the true probability distribution of the uncertain data in Ref. [18]. In Ref. [19], the spatial and temporal information in the wind power dataset is captured based on GAN. The above studies apply machine learning methods to explore the internal information in the uncertainty dataset, including probabilistic, temporal and spatial information, which improves the accuracy of uncertainty characterization with more efficiency. Therefore, the conservativeness of the optimization results decreases when machine learning methods are applied to RO.

Leveraging the strengths of machine learning methods in uncertainty characterization, data-driven robust optimization (DDRO) has been an emerging paradigm for dealing with uncertainties, in which the data-driven uncertainty set is constructed based on machine learning methods and more information in the uncertainty dataset is utilized. It is widely applied to unit commitment problems [20, 21, 22], microgrids [23, 24, 25], virtual power plant [26], energy management [27, 28], reserve dispatch [29], etc. On the one hand, considering the correlation information in the uncertainty dataset, Refs. [22, 26] construct the uncertainty set of the uncertain variables based on support vector machine (SVM) and DPMM, respectively. In Ref. [22], the construction of the uncertainty set for the wind power is inspired by support vector machine (SVM), and the conservativeness of the optimization results is decreased in RO. Further, considering the label information of historical data and correlations among the uncertainties, Ref. [26] develops a two-stage virtual power plant stochastic robust optimization model, in which the ambiguity set is constructed by DPMM. On the other hand, considering the probability information in the uncertainty dataset, Refs. [21, 27, 28, 29] construct the uncertainty set of the probability distributions for the uncertain variables. In [21], a novel distance-based ambiguity set for wind power is developed through Kullback–Leibler divergence, and a distributionally robust optimization model for unit commitment is established. The uncertainty set of probability distributions for wind power generation is constructed based on Wasserstein metric, and the operation costs of the power system are minimized in [27, 28]. On this basis, the one-dimensional wind power generation is extended to the multivariate wind power generation. Ref. [29] constructs the multivariate probability distributions ambiguity set based on the kernel density estimation (KDE), and proposes a risk-averse two-stage stochastic robust optimization method to deal with the reserve dispatch problem with wind power. It can be summarized from the above studies that DDRO based on machine learning methods can well extract the internal information (e.g., probability distribution and correlation) in the uncertainty dataset to construct the data-driven uncertainty set, which improves the data-adaptive flexibility and reduces the conservatism of the optimization results. However, there are few studies on the application of DDRO to GMS, and DDRO approaches for GMS need to be further explored. The reasons and research gaps are shown as follows.

(1) The temporal information of the uncertainty set in DDRO model is not fully considered, which should be further characterized when the uncertainty variables are time series. Note that the temporal information represents that the data at the current period relies on the data from the previous period. For instance, wind

power data from adjacent time periods are correlated, since the constant variation of wind speed can be portrayed by the Markov process [30].

(2) The correlations between the uncertainty variables are not taken into account in the existing GMS model. Actually, GMS is a medium and long-term planning problem with a specific correlation between uncertain variables. Take wind power and load demand as an example, the wind power generation may be massive, and load demand becomes high to consume wind power at a certain period in the GMS planning problem.

(3) Priori's work on GMS problems under uncertainties mainly focuses on SP and RO. The former needs to know the probability distributions for the uncertain variables, and the latter relies on the boundary information of the uncertain variables. The more internal information in the uncertainties datasets has not been fully considered in these methods. Therefore, DDRO method has great potential for application to the GMS problem.

To address these issues, we propose a data-driven two-stage nested robust optimization approach for generation maintenance scheduling, and its framework is shown in Fig. 1. In this figure, the temporal information in the wind power and load demand data is firstly characterized by TCDPMM, and the data-driven uncertainty set is constructed by the VBI results from TCDPMM. Then, these steps are integrated into the two-stage robust optimization model for GMS considering temporal correlation. In this context, the contributions of this paper are as follows:

(1) A temporal correlation Dirichlet process mixture model (TCDPMM) is proposed, and the data-driven uncertainty set based on TCDPMM is constructed, which can portray the temporal correlation of wind power and load demand uncertainties. Here, the temporal correlation not only represents the temporal information in the dataset for the uncertain variables but also contains the correlations between the uncertain variables. This aims to solve research gaps 1 and 2.

(2) In comparison with conventional robust GMS models, a data-driven two-stage NRO GMS model with a "min-max-max-min" four-level structure form is developed. The proposed GMS model is difficult to be solved directly, and the parallel Benders' decomposition algorithm is introduced to solve it. Moreover, since the data-driven uncertainty set is integrated into the proposed GMS model, the Bayesian nonparametric method is leveraged for GMS in the power system. The contribution is targeted to deal with the research gap 3.

(3) The effectiveness and superiority of the proposed approach are verified in the 6-bus power system and the 26-bus Yantai power system in China. Meanwhile, the NRO methods with different data-driven uncertainty sets are compared in terms of the total cost of the power system operation. Here, the uncertainty sets include the DPMM-based data-driven set, the TCDPMM-based data-driven uncertainty set without considering correlation and the proposed TCDPMM-based uncertainty set.

The rest of this paper is organized as follows. The data-driven uncertainty set based on TCDPMM is constructed in Section 2. Section 3 presents the two-stage NRO model for GMS, which integrates the proposed uncertainty set. In Section 4, the two-stage NRO model is solved by the parallel Benders' decomposition algorithm. Comparison case studies are shown in Section 5, and the conclusion is drawn in Section 6.

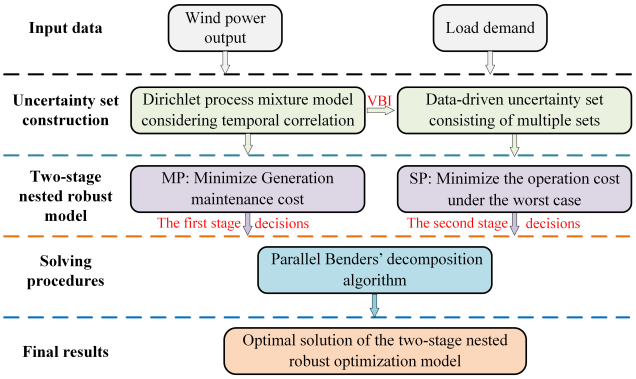


Fig. 1: Overall framework of the proposed approach

2. Data-Driven TCDPMM Based Uncertainty Model

Note that the wind power and load demand data in the power system contain rich information (e.g., temporal characteristic and correlation), which is difficult to be well portrayed by empirical distribution or single distribution [31]. DPMM, an infinite mixture Gaussian model based on Dirichlet process (DP), is usually applied to describe the distribution of input data [32]. The uncertainty and correlation of input data can be well characterized by DPMM [33]. However, the temporal characteristics are not fully considered in DPMM when the input data is time-series, such as the wind power and load demand data in the power system. In this section, we propose a data-driven TCDPMM to explore the temporal information of the wind power and load demand data. Meanwhile, to characterize the correlation between uncertain variables, the data-driven uncertainty set with temporal correlations is constructed based on the TCDPMM.

2.1. Dirichlet Process Mixture Model

DPMM is a non-parametric Bayesian model based on DP, which can be described by a mixture of probabilistic distributions and their corresponding weights [34]. Here, the mixture probabilistic distributions and the corresponding weights are represented by G and $\boldsymbol{\pi}$, respectively. Note that G is the discrete distribution and can be represented by DP distribution if $G \sim DP(\alpha, G_0)$, where α is the concentration

parameter, and G_0 is the base distribution. Generally, DP is expressed explicitly by stick-breaking representation, and that of G is defined as below.

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{D_k} \tag{1}$$

$$\pi_k = \beta_k \prod_{l=1}^{k-1} (1 - \beta_k) \tag{2}$$

where π_k is the random weight of the distribution D_k sampled from the base distribution G_0, and $\sum_{k=1}^{\infty} \pi_k = 1$. Note that the set of D_k stands for the mixture Gaussian distributions G. δ_{D_k} is a Kronecker function. $\delta_{D_k} = 1$ if D_k is chosen by G, and $\delta_{D_k} = 0$ elsewhere. β_k is the proportion of π_k broken off from the remaining the unit length stick, and $\beta_k \sim Beta(1, \alpha)$. In this way, the graphical model of DPMM with sticking-breaking representation is depicted as follows.

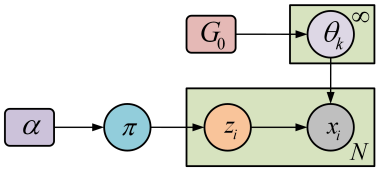


Fig. 2: The graphical model of DPMM

Fig. 2 shows the relations among variables. Here, θ_k is the parameter for Gaussian distribution D_k, and the latent variable z_i indicates that the observation x_i is assigned to which Gaussian distribution. Therefore, the mathematical expression of DPMM with stick-breaking representation is shown in (3).

$$\begin{aligned} &\boldsymbol{\pi}|\alpha \sim GEM(\alpha) \\ &\theta_k|G_0 \sim G_0 \\ &z_i|\boldsymbol{\pi} \sim \boldsymbol{\pi} \\ &x_i|z_i, (\theta_k)_{k=1}^{\infty} \sim F(\theta_{z_i}) \\ &F(\theta_{z_i}) = p(x_i|\boldsymbol{z}, \boldsymbol{\theta}) = \prod_{k=1}^{\infty} \mathcal{N}_k(x_i|\mu_k, \Sigma_k)^{z_k^{[i]}} \end{aligned} \tag{3}$$

where $GEM(\alpha)$ stands for the construction process of the weighted $\boldsymbol{\pi}$, and $F(\theta_{z_i})$ is a Gaussian mixture distribution. The parameter $\boldsymbol{\theta}$ consists of the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$, and $\boldsymbol{\theta} = (\mu_k, \Sigma_k)_{k=1}^{\infty}$. $\mathcal{N}_k(\cdot|\cdot)$ represents the k-th Gaussian distribution. The binary variable $z_k^{[i]}$ indicates that if i-th observation data belongs to k-th Gaussian component, $z_k^{[i]} = 1$, otherwise, $z_k^{[i]} = 0$. The derivation of (3) is presented in Appendix D.1.

2.2. Temporal Correlation Dirichlet Process Mixture Model

In DPMM, the mutually independent latent variable $\boldsymbol{z}$ indicates which Gaussian distribution the observation $\boldsymbol{x}$ is associated with. The temporal characteristics are not fully considered when the observation $\boldsymbol{x}$ is time-series. To explore the temporal correlation of the observation data, we propose a TCDPMM, and its graphical model is shown in Fig. 3.

Different from DPMM, z_i and z_{i-1} are the mutually dependent in Fig. 3. In other words, the value of z_i is related to z_{i-1}, which means that the correlation between x_i and x_{i-1} is characterized through the hidden variable z. In this way, the temporal correlation of $\boldsymbol{x}$ is extracted when the observation $\boldsymbol{x}$ is time-series. Moreover, the same as DPMM represented in Subsection 2.1, the Gaussian mixture distribution of the observation $\boldsymbol{x} = \{x_i\}_{i=1}^{I}$ in the proposed TCDPMM is expressed as follows.

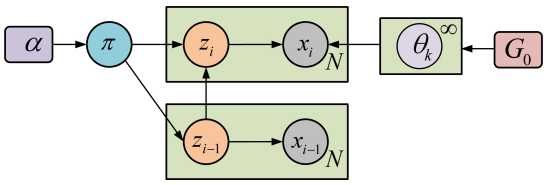


Fig. 3: The graphical model of TCDPMM

$$p(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{\theta}) = \prod_{i=1}^{I} \prod_{k=1}^{\infty} \mathcal{N}_k(x_i|\mu_k, \Sigma_k)^{z_k^{[i]}} \tag{4}$$

where $\boldsymbol{z} = (z_k^{[i]})_{i=1,k=1}^{i=I,k=\infty}$ and $\boldsymbol{\theta} = (\theta_k)_{k=1}^{k=\infty}$ are the set of z_k and θ_k, respectively.

According to the graphical model shown in Fig. 3, the joint distribution of the all random variables is expressed as

$$p(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = p(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) p(\boldsymbol{z}|\boldsymbol{\pi}) p(\boldsymbol{\pi}) p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{5}$$

Traditionally, the relevant parameters of the prior distribution can be easily estimated based on the mathematical and statistical characteristics of the observation data. However, the hyper-parameters $w = (\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ and the latent variable $\boldsymbol{z}$ are difficult to be determined directly due to the multi-dimensional complexity of the observation data $\boldsymbol{x}$. To this end, Variational Bayesian inference (VBI) is employed to address such a dilemma in this paper. The core of VBI is to use the variational distribution q to approximate the posterior distribution p. Based on VBI, the posterior distribution p is approximated by a factorized variational distribution:

$$q(\boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = q(\boldsymbol{z}) q(\boldsymbol{\pi}) q(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{6}$$

where $q(\boldsymbol{\pi})$ is a Dirichlet distribution, and $q(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ follows a Gaussian-Wishart distribution. The specific expressions of $q(\boldsymbol{\pi})$ and $q(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ are shown in the Appendix A. The VBI for TCDPMM is similar to that of DPMM, and it is represented in the Appendix D.2, in which the optimal distributions of $q(\boldsymbol{z})$, $q(\boldsymbol{\pi})$ and $q(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ are determined.

As shown in the Appendix D.2, the optimal distribution of $q(\boldsymbol{z})$ is given by:

$$q^*(\boldsymbol{z}) = \prod_{i=1}^{I} \prod_{k=1}^{K} (r_k^i)^{z_k^{[i]}} \tag{7}$$

$$r_k^i = p(z_k^{[i]} = 1) = \frac{\rho_k^i}{\sum_{j=1}^{K} \rho_j^i} \tag{8}$$

where r_k^i is the weight that the observation x_i belongs to k-th Gaussian component, and ρ_k^i represents the probability that k-th Gaussian component samples observation data x_i, which is shown in Appendix A.

Note that since z_k is correlated with z_{k-1} in TCDPMM when the observation $\boldsymbol{x}$ is the time series, the optimal distribution $q^*(\boldsymbol{z})$ needs to be modified, which is reformulated as follows.

$$q^*(z_k) = q(z_k|z_{k-1}) \tag{9}$$

In this way, the conditional probability of z_k is as below.

$$\begin{aligned} p(z_k^{[i]} = 1) &= p(z_k^{[i]} = 1 | z_{k'}^{[i-1]} = 1) \\ &= \frac{p(z_k^{[i]} = 1, z_{k'}^{[i-1]} = 1)}{p(z_{k'}^{[i-1]} = 1)} \end{aligned} \tag{10}$$

where $k^{'}$ is the maximum weight Gaussian component that observation data x_{i-1} associates with. Specifically, Eq. (10) illustrates that the i-th observation data belongs to the k-th Gaussian component depending on the Gaussian component $(i-1)$-th observation data belongs to. In order to display $p(z_k^{[i]} = 1)$, we define $p(z_k^{[i]} = 1, z_{k'}^{[i-1]} = 1)$ as below.

$$\begin{aligned} &p(z_k^{[i]} = 1, z_{k'}^{[i-1]} = 1) = \\ &\Big(m \cdot p(z_k^{[i]} = 1) + m^{'} \cdot p(z_{k'}^{[i]} = 1)\Big) \times \\ &\Big(m \cdot p(z_k^{[i-1]} = 1) + m^{'} \cdot p(z_{k'}^{[i-1]} = 1)\Big) \\ &= (m \cdot r_k^i + m^{'} \cdot r_{k'}^n) \cdot (m \cdot r_k^{i-1} + m^{'} \cdot r_{k'}^{i-1}) \end{aligned} \tag{11}$$

where $m = \frac{\pi}{\pi+\pi^{'}}$ and $m^{'} = \frac{\pi^{'}}{\pi+\pi^{'}}$ represent the proportion of k-th Gaussian component and $k^{'}$-th Gaussian component, respectively. Therefore, ρ_k^n is reformulated as.

$$\begin{aligned} \rho_{k'}^i &= \frac{p(z_k^{[i]} = 1, z_{k'}^{[i-1]} = 1)}{p(z_{k'}^{[i-1]} = 1)} \\ &= (m \cdot r_k^i + m^{'} \cdot r_{k'}^i) \cdot (m \cdot \frac{r_k^{i-1}}{r_{k'}^{i-1}} + m^{'} \cdot r_{k'}^{i-1}) \end{aligned} \tag{12}$$

The weight $r_k^{'i}$ is modified by normalization $\rho_{k'}^i$.

$$r_k^{'i} = \frac{\rho_{k'}^i}{\sum_{j=1}^{\infty} \rho_{j'}^i} \tag{13}$$

Therefore, $q^*(\boldsymbol{z})$ is reformulated as below.

$$q^*(\boldsymbol{z}) = \prod_{i=1}^{I} \prod_{k=1}^{K} (r_k^{'i})^{z_k^{[i]}} \tag{14}$$

In this way, the optimal distribution of $q(\boldsymbol{z})$ in TCDPMM can be calculated by (14). Moreover, the optimal distributions of $q(\boldsymbol{\pi})$ and $q(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ can be calculated by (D.17) and (D.18). In order to make the variational distribution q as close as possible to the posterior distribution p, we implement the iterative update of the parameters using the Expectation-Maximization (EM) algorithm [35]. The optimal parameter w^* and optimal distribution $q^*(\boldsymbol{z})$ in TCDPMM can be obtained when the EM algorithm reaches convergence. That is, the optimal variational distribution over the optimal parameters is determined, which is the closest to the posterior distribution p. The detailed flow of the EM algorithm can be referred to Appendix B.

2.3. The Construction of Data-driven Uncertainty Set

In this paper, the observation $\boldsymbol{x} = \{x_i\}_{i=1}^{i=I}$ consists of wind power and load demand, in which $x_i = \{P_{wt,i}, P_{load,i}\}$. Moreover, the observation $\boldsymbol{x}$ is considered to follow a mixture of Gaussian distributions. Based on TCDPMM, the temporal information in the observation $\boldsymbol{x}$ can be characterized. Meanwhile, the mean, the covariance and the correlation parameters can be obtained according to the variational inference for TCDPMM.

In order to capture the uncertainty of wind power and load demand, the data-driven uncertainty set is constructed based on the results from TCDPMM variational inference, which is shown as below.

$$U = U_1 \cup U_2 \cup \cdots \cup U_i \tag{15}$$

$$U_i = \{\mathbf{u} | \mathbf{u} = \boldsymbol{\mu}_i + \omega_i {\boldsymbol{\Psi}_i}^{1/2} \Lambda_i \mathbf{Z}, \|\mathbf{Z}\|_\infty \leq 1, \|\mathbf{Z}\|_1 \leq \boldsymbol{\Phi}_i\} \tag{16}$$

where U_i is the uncertainty set for the i-th Gaussian component. $\boldsymbol{\mu}_i$, ω_i, $\boldsymbol{\Psi}_i$ can be obtained from the inference results of the i-th Gaussian component using the variational inference algorithm [36]. Λ_i is the scaling factor. $\mathbf{Z}$ is the auxiliary variable, which can control the correlation among uncertainties. $\boldsymbol{\Phi}_i$ stands for the uncertainty budget.

It is found from (15) that the data-driven uncertainty set U is composed of multiple basic uncertainty sets. The number of basic uncertainty sets depends on the number of clustered components of the observed data, which is obtained by the VBI for TCDPMM. Meanwhile, the basic uncertainty set U_i corresponding to component i is explicitly shown in (16). Since the proposed data-driven uncertainty set is constructed based on the VBI for TCDPMM, it can capture more internal information in the uncertain variables dataset. There are several advantages of the proposed uncertainty set. First, the temporal information for the uncertain variables is mined in the proposed data-driven uncertainty set, since TCDPMM considers the temporal correlation of the observation data. Second, the correlation between the uncertain variables is considered in the proposed uncertainty set, which is subject to the auxiliary variable $\mathbf{Z}$. Finally, the proposed uncertainty set is constructed by the Gaussian components, which utilizes the probabilistic information of the uncertain variables.

The data-driven uncertainty set considering the temporal information and correlation between the uncertain variables will be integrated into the two-stage generation maintenance scheduling model in the next section.

3. Two-Stage NRO Model for GMS

Generally, GMS problem is a sequential decision-making problem, and it contains two stages [37]. In the first stage, the maintenance decisions are determined to obtain the number of available units. On this basis, the generation scheduling for the available units is optimized in the second stage. Apparently, GMS is a typical two-stage programming problem. However, the GMS becomes an uncertain mathematical problem when the uncertain wind power and load

demand are integrated into the two-stage model. To cope with the uncertainties, a TCDPMM-based two-stage NRO optimization model for the GMS of the power system is presented as follows.

1) Objective Function

The objective of the two-stage robust optimization model for GMS is to minimize the total cost of the power system under the worst case uncertainty scenario, which is shown in (17).

$$\min_{x,g} \sum_{i=1}^{n_G}\sum_{t=1}^{T} \overbrace{(c_{M,i}x_{i,t} + c_{R,i}(1-g_{i,t}))}^{\textbf{The first term}} + \max_{P_{wt},P_{load}\in U} \min_{P_G} \sum_{i=1}^{n_G}\sum_{t=1}^{T} \underbrace{(c_{O,Gi}+c_{F,Gi})P_{Gi}}_{\textbf{The second term}} \tag{17}$$

where the first term represents the cost in the first stage, which consists of maintenance cost and reserve cost for all generation units in the power system. The second term is the operation cost in the second stage, including generation cost and fuel cost. The i-th unit generation maintenance status $x_{i,t}$ and the i-th unit operation status $g_{i,t}$ are the first stage decision variables, and the output power P_{Gi} of i-th unit is the second stage decision variable. $c_{M,i}$ and $c_{R,i}$ are the maintenance factor and reserve factor of i-th unit, respectively. $c_{O,Gi}$ and $c_{F,Gi}$ are the generation factor and fuel factor of i-th unit. It should be noted that the output power P_{wt} of WT and load demand P_{load} are uncertain variables, which are subject to the data-driven uncertainty set U constructed in (15) and (16).

Usually, an uncertainty set corresponds to a "max-min" problem in RO. Therefore, the second stage optimization problem contains multiple "max-min" problems when the data-driven uncertainty sets for wind power and load demand consisting of multiple subsets are embedded in the two-stage optimization model for GMS. Note that we need to find the minimum operation cost under the worst-case scenario. That is to say, our goal is to find the problem with the largest value of the objective function from the multiple "max-min" problems, which corresponds to the worst-case scenario and is represented by the "max-max-min" form. Therefore, the two-stage robust optimization model for GMS is converted to the NRO model with a "min-max-max-min" form, which is defined as below.

$$\min_{x,g} \sum_{i=1}^{n_G}\sum_{t=1}^{T} \overbrace{(c_{M,i}x_{i,t} + c_{R,i}(1-g_{i,t}))}^{\textbf{The first term}} + \max_{n=\{1,\ldots,i\}} \max_{P_{wt},P_{load}\in U_n} \min_{P_G} \sum_{i=1}^{n_G}\sum_{t=1}^{T} \underbrace{(c_{O,Gi}+c_{F,Gi})P_{Gi}}_{\textbf{The second term}} \tag{18}$$

2) Constraints for Units Maintenance

To improve the performance of units, the units maintenance duration constraint should be enforced by (19). Meanwhile, in order to ensure the reliability of the power supply, the number of units maintenance simultaneously is limited in (20), and the uninterrupted units maintenance constraint is represented in (21).

$$\sum_{t=1}^{T} x_{i,t} = D_i, \forall i \tag{19}$$

$$\sum_{i=1}^{n_G} x_{i,t} \le R_t, \forall t \tag{20}$$

$$x_{i,t} - x_{i,t-1} \le x_{i,t+D_i-1}, \forall i, \forall t = 2, ..., T+1-D_i \tag{21}$$

where D_i is the maintenance duration of i-th unit, and R_t is the maximum number of unavailable units at time t.

3) Constraints for Logic Relationships and Reserve

The logic relationships between the generation maintenance status and the operation status are described in (22). Constraint (23) enforce the reserve of the power system.

$$x_{i,t} \ge m_{i,t}, \forall i, \forall t \tag{22}$$

$$\sum_{i=1}^{n_G} P_{\max,Gi}(1-x_{i,t}) \ge (1+r_t)P_{load,t}, \forall t \tag{23}$$

where $P_{\max,Gi}$ is the maximum power output of i-th unit, and r_t is the reserve coefficient at time t. The load demand of time t is represented by $P_{load,t}$.

Note that $x_{i,t}$ and $g_{i,t}$ are binary variables, which are shown in (24).

$$x_{i,t}, g_{i,t} \in \{0,1\}, \forall i, \forall t. \tag{24}$$

where the unit i is in maintenance state/shutdown state at time t, if $x_{i,t} = 1$ / $g_{i,t} = 1$, the unit i is in available state/operation state, otherwise.

4) Constraints for Generation and Ramping Limits

In second stage, the supply and demand balance of the system should be satisfied, so the power balance constraint is described in (25). Meanwhile, the minimum and maximum outputs of each unit are specified in (26), and the ramping rate limits of units are presented in (27) and (28).

$$\sum_{i=1}^{n_G} P_{Gi,t} + \sum_{j=1}^{n_W} P_{wtj,t} = P_{load,t}, \forall t \tag{25}$$

$$P_{\min,Gi}(1-g_{i,t}) \le P_{Gi,t} \le P_{\max,Gi}(1-g_{i,t}) \tag{26}$$

$$P_{Gi,t} - P_{Gi,t-1} \le RU_i, \forall i, \forall t \tag{27}$$

$$P_{Gi,t-1} - P_{Gi,t} \le RD_i, \forall i, \forall t \tag{28}$$

where $P_{Gi,t}$ and $P_{wtj,t}$ are the power output of i-th unit and j-th WT, respectively. n_G and n_W are the number of units and WT. $P_{\min,Gi}$ is the minimum power output of i-th unit. RU_i and RD_i are the ramp up rate and the ramp down rate of i-th unit.

5) Constraints for Transmission Lines

The transmitted power is affected by the capacity of transmission lines, which is described in (29) and (30).

$$\sum_b k_{lb} \cdot \left(\sum_{i\in b} P_{Gi,t} + \sum_{j\in b} P_{wj,t} - \sum_{d\in b} P_{d,t} \right) \le f_{l\max} \tag{29}$$

$$\sum_b k_{lb} \cdot \left(\sum_{i\in b} P_{Gi,t} + \sum_{j\in b} P_{wj,t} - \sum_{d\in b} P_{d,t} \right) \ge -f_{l\max} \tag{30}$$

where $f_{l\max}$ is the maximum power flow of the line l, and k_{lb} is the power flow distribution factor for the transmission line l due to the net injection at bus b.

The two stages in the proposed NRO model for GMS are defined as the "here-and-now" and "wait-and-see" stages. Specifically, in the first stage, "here-and-now" variables including maintenance status $x_{i,t}$ and operation status $g_{i,t}$ can be optimized, which helps the power system operation to determine the number of available units. On this basis, "wait-and-see" variable P_{Gi} will be obtained once the uncertain variables P_{wt} and P_{load} are revealed in the second stage. Therefore, the sequential decision-making process of GMS in the power system operation is captured. In particular, it should be noticed that the uncertain variables determined at the second stage are from the data-driven uncertainty set which characterizes the temporal information in the uncertainties dataset and the correlations between the uncertain variables. Meanwhile, the model combined with the Bayesian nonparametric model-TCDPMM can adapt to the complexity of wind power and load demand datasets, which fully leverages the internal information in the datasets to achieve more efficient power systems operations.

4. Model Solution

The two-stage NRO model for GMS constructed in Section III is a multi-level structure optimization problem, which is reformulated in the form of the general compact matrix for description brevity.

$$\min_{\mathbf{x}} \mathbf{c}^T\mathbf{x} + \max_{n=\{1,\dots,i\}} \max_{\mathbf{u}\in U_n} \min_{\mathbf{y}\in\Omega(\mathbf{x},\mathbf{u})} \mathbf{b}^T\mathbf{y} \tag{31}$$

$$\mathbf{s.t.}\ \mathbf{Ax} \geq \mathbf{d},\ \mathbf{x} \in \mathbf{s_x} \tag{32}$$

$$\Omega(\mathbf{x},\mathbf{u}) = \{\mathbf{y} \in \mathbf{s_y} : \mathbf{Wy} \geq \mathbf{h} - \mathbf{Tx} - \mathbf{Mu}\} \tag{33}$$

$$U = U_1 \cup U_2 \cup \cdots \cup U_i \tag{34}$$

$$U_i = \{\mathbf{u}|\mathbf{u} = \boldsymbol{\mu}_i + \omega_i \boldsymbol{\Psi}_i^{1/2} \Lambda_i \mathbf{Z}, \|\mathbf{Z}\|_\infty \leq 1, \|\mathbf{Z}\|_1 \leq \boldsymbol{\Phi}_i\} \tag{35}$$

where $\mathbf{x}$ and $\mathbf{y}$ are the first-stage decision variables and the second-stage decision variables, which can be described by $\mathbf{x} = \{x, g\}$ and $\mathbf{y} = \{P_{Gi}\}$, respectively. $\mathbf{u}$ represents the uncertain variables, and $\mathbf{u} = \{P_{wt}, P_{load}\}$. The objective function (18) is represented as (31), and constraint (32) is the set of all the first-stage constraints including maintenance constraints and reserve constraints that are formulated in (19)-(24). The set of second-stage constraints (25)-(30) is shown in (33). Constraints (34)-(35) are the data-driven uncertainty set defined in Section II. Note that $\mathbf{s_x}$ and $\mathbf{s_y}$ represent the feasibility set of $\mathbf{x}$ and $\mathbf{y}$.

In general, the conventional two-stage robust optimization model for GMS is a "min-max-min" model. To solve this model, it is decomposed into a master problem (MP) and a sub-problem (SP) in the conventional method. Since the objective function values of MP and SP can converge to the same value, the optimal solutions for the model can be obtained by iterative solving MP and SP. However, the nested model in (31) is a "min-max-max-min" model, in which the uncertainty set U is a non-convex set consisting of multiple subsets. Note that a subset corresponds to a SP. It means that problem (31) can be decomposed into a MP and multiple SPs. Therefore, the two-stage NRO model for GMS in this paper is difficult to be solved directly by the conventional method. To address this issue, we introduce a parallel Benders' decomposition-based approach to solve the two-stage NRO for GMS. The parallel Benders' decomposition-based approach is described in **Algorithm** 1.

As shown in **Algorithm** 1, the proposed approach decomposes the two-stage problem into a master problem and multiple sub-problems. The specific expression of MP is as below.

$$\begin{aligned} \mathbf{MP}:\ &\min_{\mathbf{x},\boldsymbol{\eta}} \mathbf{c}^T\mathbf{x} + \eta \\ &\mathbf{s.t.}\ \mathbf{Ax} \geq \mathbf{d} \\ &\eta \geq (\mathbf{h} - \mathbf{Ey} - \mathbf{Mu}_l{}^*)^T \boldsymbol{\pi}_l^*, \forall l \leq k \\ &\mathbf{x} \in \mathbf{s_x}, \mathbf{y} \in \mathbf{s_y}, \eta \in \mathbf{R} \end{aligned} \tag{36}$$

where η is an auxiliary variable. $\mathbf{u}_l{}^*$ and $\boldsymbol{\pi}_l^*$ are the optimal solutions for SP at the l-th iteration, which represent the worst-case uncertainties realization and the value of the dual variable, respectively. The MP is to make decisions for units maintenance and reserve to minimize the total cost of the power system when the uncertainties are captured from the SP in the previous iteration.

After obtaining $\mathbf{x}^*$ in the MP, the SP can determine the worst-case wind power and load demands. Meanwhile, the optimality cuts associated with the determined worst-case are added to MP for the next iteration. Specially, we should notice that each uncertainty set U_i corresponds to a subproblem $\mathbf{SP}_i$, which forms a "max-max-min" problem combined with the worst property of robust optimization. To deal with the tri-level structure, we transform "max-max-min" problem into the multiple subproblems ($\mathbf{SP}_n, n = 1, ..., i$) and solve them in parallel. In each iteration, the largest one is selected by comparing the optimal objective values of all subproblems such that the worst-case uncertainties can be revealed and determine the optimality cuts added to MP. The subproblem ($\mathbf{SP}_i$) is expressed as below.

$$\begin{aligned} \mathbf{SP}_i:\ &\max_{\mathbf{u}\in U_i} \min_{\mathbf{y}} \mathbf{b}^T\mathbf{y} \\ &\mathbf{s.t.}\ \mathbf{Wy} \geq \mathbf{h} - \mathbf{Tx}^* - \mathbf{Mu} \\ &\mathbf{y} \in \mathbf{s_y} \end{aligned} \tag{37}$$

where $\mathbf{x}^*$ is the optimal decision obtained in the MP. In general, the "max-min" problem can be converted into a single-level problem based on the dual theory. In this paper, we introduce the dual variable π and transfer the inner min problem of $\mathbf{SP}_i$ into the max problem such that the problem (37) is redefined as below.

$$\begin{aligned} \mathbf{SP}_i:\ &Q_i(\mathbf{x}^*) = \max_{\mathbf{u},\boldsymbol{\pi}} (\mathbf{h} - \mathbf{Tx}^* - \mathbf{Mu})^T \boldsymbol{\pi} \\ &\mathbf{s.t.}\ \mathbf{G}^T\boldsymbol{\pi} \leq \mathbf{b} \\ &\boldsymbol{\pi} \geq 0,\ \mathbf{u} \in U_i \end{aligned} \tag{38}$$

where $Q_i(\mathbf{x}^*)$ is the objective function value of i-th **SP**. Note that the product of $\mathbf{u}$ and $\boldsymbol{\pi}$ in (38) contains the bilinear terms $\mathbf{Z}\boldsymbol{\pi}$, which are linearized by employing the big-M method in this paper. In detail, the transformation of $\mathbf{Z}\boldsymbol{\pi}$ is shown

below.

$$\begin{aligned} 0 \le \mathbf{F} \le M_0 \cdot \mathbf{Z} \\ \boldsymbol{\pi} - M_0(1-\mathbf{Z}) \le \mathbf{F} \le \boldsymbol{\pi} \end{aligned} \tag{39}$$

where M_0 is a very large positive number, and $\mathbf{F} = \mathbf{Z}\boldsymbol{\pi}$. In this way, $\mathbf{SP}_i$ is converted into a linear programming model which can be solved by commercial solvers.

Algorithm 1 Parallel Benders' Decomposition Algorithm

Set $LB \leftarrow -\infty$, $UB \leftarrow \infty$, $k \leftarrow 0$, and $\varepsilon \leftarrow 10^{-3}$
while $\left|\frac{UB-LB}{UB}\right| \ge \varepsilon$ **do**
  Solve **MP** to obtain $\mathbf{x}^*_{k+1}, \eta^*_{k+1}, \mathbf{y}^{1*}, ..., \mathbf{y}^{k*}$;
  Update $LB \leftarrow c^T x^*_{k+1} + \eta^*_{k+1}$;
  for $i = 1$ to n **do**
    Solve $\mathbf{SP}_i$ to obtain $\mathbf{u}_i^{k+1}$ and $Q_i(\mathbf{x}^*_{k+1})$;
  end for
  $i^* \leftarrow \arg\max_i Q_i(\mathbf{x}^*_{k+1})$;
  $\mathbf{u}^{k+1} \leftarrow \mathbf{u}_{i^*}^{k+1}$ and $Q(\mathbf{x}^*_{k+1}) \leftarrow Q_{i^*}(\mathbf{x}^*_{k+1})$;
  Update $UB \leftarrow \min\left\{UB, \mathbf{c}^T\mathbf{x}^*_{k+1} + Q(\mathbf{x}^*_{k+1})\right\}$;
  Create second-stage variables $\mathbf{y}^{k+1}$ and add cuts $\eta \ge \mathbf{b}^T\mathbf{y}^{k+1}$, $\mathbf{Tx} + \mathbf{Wy}^{k+1} \ge \mathbf{h} - \mathbf{Mu}^{k+1}$ to **MP**;
  $k \leftarrow k+1$;
end while
return UB.

It is notable that UB and LB are the upper and the lower bounds, which can be obtained from MP and SP, respectively. ε represents the optimality gap that is used to determine when the algorithm stops. When the gap between UB and LB is smaller than the ε, the algorithm converges and the optimal solution of the two-stage NRO for GMS is determined.

5. Simulations and Results

In this section, we conduct case studies to verify the effectiveness of the proposed approach. The data-driven two-stage NRO model for the GMS is solved using GUROBI within MATLAB 2021b. In Subsection 5.1, the parameters and data for the power system are represented. Subsection 5.2 illustrates the performance of the proposed TCDPMM model. On this basis, Subsection 5.3 analyzes the results from the data-driven two-stage NRO approach for GMS on a six-bus system. Finally, the proposed approach is applied to the 26-bus Yantai power system to demonstrate its performance and superiority.

5.1. Data and Parameter Settings

In this paper, all case studies are carried out on a modified IEEE 6-bus system and a 26-bus Yantai power system in China. The topology of the 6-bus power system is shown in Fig. 4, which contains 2 wind turbines, 4 generators, and 4 load demands. The parameters of all units are taken from [13], which are shown in Table 1. Note that the reserve cost of units is set to 1% of the maintenance cost, and the fuel cost of units is assumed to be 14.4 $/MW. The parameters of the transmission lines for IEEE 6-bus system are obtained from [38]. Wind power and load demand datasets are derived from the real annual data of Austria in 2015. The time step for GMS is one week and the operation horizon is one year which can be divided into 52 weeks. The weekly load demand and wind power generation are shown in Fig. 5. Furthermore, the optimality gap ε is predefined as 10^{-4}.

TABLE 1
The parameters of all units

Unit	c_M ($)	$c_{O,G}$ ($/MWh)	P_{max}/P_{min}(MW)	D(h)	RU/RD(MW/h)
P_{G1}	40000	63	520/0	4	350
P_{G2}	30000	73	360/0	5	250
P_{G3}	10000	98	200/0	2	100
P_{G4}	40000	73	360/0	5	250

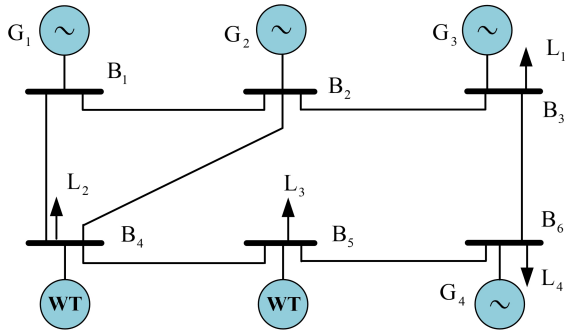


Fig. 4: Six bus power system

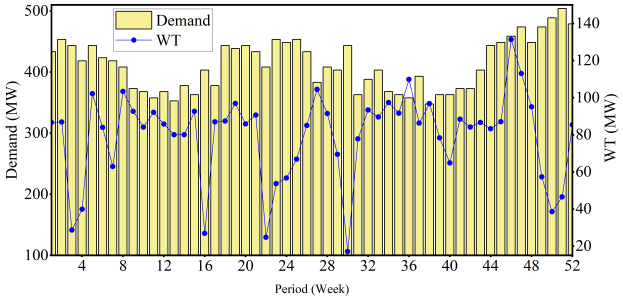
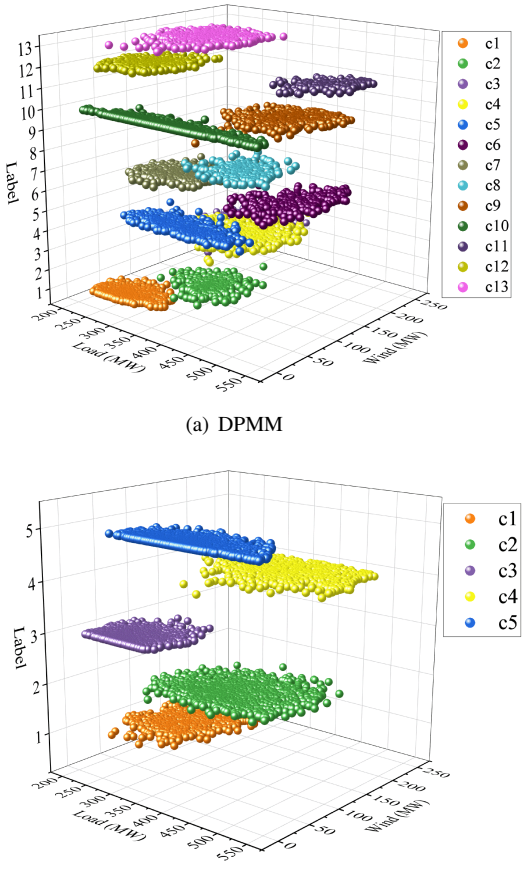


Fig. 5: Load demand and wind generation

5.2. Effectiveness of TCDPMM Model

To verify the effectiveness of the TCDPMM model constructed in Section 2, we compare the component results that TCDPMM and DPMM explore the internal information of wind power and load demand datasets. The components based on TCDPMM and DPMM are shown in Fig. 6, and the data in the same component has common intrinsic features. From Fig. 6(a) and Fig. 6(b), we can find that the number of components based on DPMM is 13, and the number of components based on TCDPMM is 5. Obviously, the number of components for the datasets is reduced after taking into account the temporal correlation.

Further, in order to exhibit the difference between DPMM-based components and TCPMM-based components clearly, the probabilistic information of components is represented in Fig. 7. We can see that the probability of TCDPMM-based components becomes larger than the probability of DPMM-based components. This means that the typical features of the datasets can be better mined due to



(a) DPMM

(b) TCDPMM

Fig. 6: Components for TCDPMM and DPMM

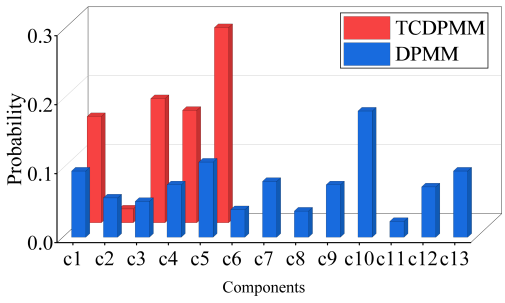
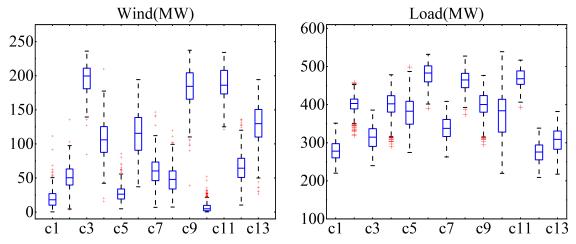


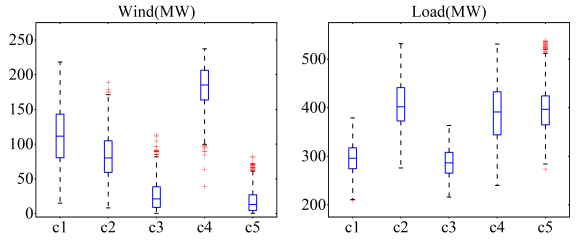
Fig. 7: Probability of components

the fewer quantities of lower probability components, when the temporal information of datasets is considered.

Based on the component results obtained from TCDPMM and DPMM, the data-driven uncertainty sets of wind power and load demand are represented as boxplots in Fig. 8. Each subfigure in Fig. 8 contains two boxes representing wind power and load demand, respectively. The horizontal coordinate of all boxes represents the components, and the vertical coordinate represents the corresponding value of the uncertainty set. Note that the red symbols in all boxes represent data outliers that are



(a) DPMM



(b) TCDPMM

Fig. 8: Data-driven uncertainty set based on DPMM and TCDPMM

contained in the corresponding uncertainty set. Comparing Fig. 8(a) and Fig. 8(b), it can be obtained that the number of TCDPMM-based data-driven uncertainty sets is lower than the number of DPMM-based data-driven uncertainty sets, and the number of data outliers in Fig. 8(b) is also lower than the number of data outliers in Fig. 8(a). Generally, the data outliers in the uncertainty set lead to higher conservatism in robust optimization. Taking the load demand uncertainty set corresponding to c_9 in Fig. 8(a) as an example, the uncertainty set can be expressed as $320 \leq u \leq 480$, but it will be reformulated as $280 \leq u \leq 480$ when the data outliers are considered. The range of uncertainty sets becomes more extensive, which will improve the conservativeness of the robust optimization results. In addition, each component corresponds to a data-driven uncertainty set. It means that the data-driven uncertainty sets have the same probabilistic information as the corresponding components. The data-driven uncertainty set with small probability information will lead to more conservative results in robust optimization. Therefore, it can be concluded that the TCDPMM-based data-driven uncertainty set can lead to lower conservatism, which can be verified in Subsections 5.3 and 5.4.

5.3. Six-Bus Power System

In this subsection, two case studies are conducted, where the data-driven two-stage NRO approach for GMS is applied to the modified six-bus power system, and the NRO methods with various uncertainty sets are compared in terms of the total cost of the power system operation.

Case 1 – (Data-driven Two-stage NRO for GMS on a Six-Bus System): In order to represent the outperformance of the TCDPMM-based two-stage NRO for GMS, we conduct the

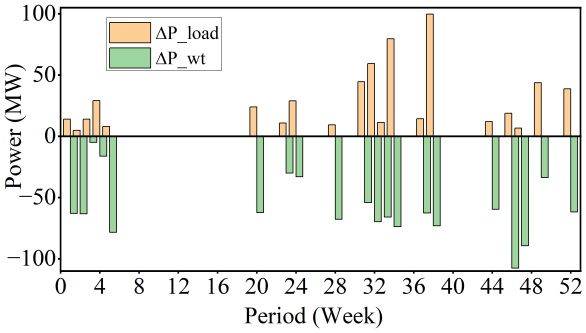


Fig. 9: The deviation of wind power and load demand compared to DM method

TABLE 2
The optimal maintenance decisions of units in different methods

Method	unit1 (weeks)	unit2 (weeks)	unit3 (weeks)	unit4 (weeks)
DM	38-41	3-7	50-51	45-49
CRO	8-11	2-6	20-21	25-29
NRO	1-4	43-47	35-36	48-52

TABLE 3
Cost comparison of different methods

Method	Total cost ($)	the first stage ($)	the second stage ($)	Conservativeness
DM	1.8665×10^6	5.0640×10^5	1.3601×10^6	-
CRO	2.2671×10^6	5.4020×10^5	1.7269×10^6	21.5%
NRO	$\mathbf{2.1060}\times 10^6$	$\mathbf{5.0680}\times 10^5$	$\mathbf{1.5992}\times 10^6$	**12.8%**

simulations to compare with the deterministic model (DM) and the conventional two-stage robust optimization model (CRO) for GMS, and the conservativeness is introduced as a measure indicator to quantify the performance of the different robust optimization methods. Note that the uncertainties of wind power generation and load demand are not considered in the DM, and the specific formulation representations of the DM and the CRO are placed in Appendix C. Moreover, the budget value Φ of the TCDPMM-based uncertainty set is set to be 20 in this case, and the values of Φ_{wt} and Φ_{Load} are also set to be 20 in the CRO method.

The purposes of the proposed NRO approach for GMS are to make the optimal generation maintenance plans under the worst case. Specifically, at the expense of certain operating costs, it can seek the optimal maintenance decisions for units to endure the normal operation of the system under the worst case, such as the maximum load demand and the minimum wind generation. The effectiveness of the proposed method is verified by the comparative results shown in Fig. 9, Table 2 and Table 3.

Comparing the TCDPMM-based NRO method, the CRO method and DM method, Fig. 9 illustrates the volatility of wind power and load demand, and the comparison results of maintenance decisions are shown in Table 2. As can be seen from Fig. 9, the wind output power is always reduced, and the load demand power is always increased in the TCDPMM-based NRO method when the budget value $\Phi = 20$, which is consistent with the worst uncertainty case in robust optimization. It indicates that the units power generation is higher and wind power generation is lower in the TCDPMM-based NRO method compared to the DM method. Moreover, to cope with the uncertainties of wind power and load demand, the optimal maintenance decisions for the units in the CRO method change from weeks 38-41, 3-7, 50-51 and 45-49 to weeks 8-11, 2-6, 20-21 and 25-29, respectively. The optimal maintenance decisions for the units in the NRO method change from weeks 38-41, 3-7, 50-51 and 45-49 to weeks 1-4, 43-47, 35-36 and 48-52. Note that the optimal maintenance decisions in the DM and NRO methods are shown in Table 2.

On the basis of the optimal maintenance decisions, the total cost of the power system operation that contains the first stage cost and the second stage cost is obtained, which is shown in Table 3. On the one hand, the total cost for TCDPMM based NRO method is higher, which is US$ 2.106×10^6 compared with US$ 1.8665×10^6 regarding that of the DM method. The reason is that the system sacrifices some economic costs to deal with the uncertainty of wind power and load demand, which can improve the security and stability of the system operation. On the other hand, it should be noticed that the first stage cost of the DM method is US$ 5.064×10^5, lower than US$ 5.068×10^5 in the NRO method. This is because that the TCDPMM-based NRO method enhances the reserve capacity of the system and increases the reserve cost in the first stage. Moreover, its operation cost of the second stage is US$ 1.5992×10^6, higher than US$ 1.3601×10^6 regarding the DM method. Consequently, we can draw a conclusion from the above simulation results that the TCDPMM-based NRO for GMS can improve the robustness of the power system at the expense of lower economic cost, since its total cost increases by only US$ 2.395×10^5.

To quantify the performance of the proposed method, the cost in the DM method is regarded as a baseline. We can find from Table 3 that the total cost for TCDPMM based NRO method is lower, which is US$ 2.106×10^6 compared with US$ 2.2671×10^6 regarding that of the CRO method. The conservativeness of the CRO and the NRO can be calculated by $\left(2.2671 \times 10^6 - 1.8665 \times 10^6\right) / 1.8665 \times 10^6$ and $\left(2.106 \times 10^6 - 1.8665 \times 10^6\right) / 1.8665 \times 10^6$, and the values of the conservativeness are 21.5% and 12.8%, respectively. Apparently, compared with the CRO, the conservativeness of the proposed NRO method is reduced by 8.7%, which means that the proposed NRO method reduces the cost of the power system to deal with the uncertainties. Meanwhile, the first stage cost of the NRO method is US$ 5.068×10^6, lower than US$ 5.402×10^6 regarding the CRO method. It is indicated that the reserve cost in the NRO method is reduced, compared to the CRO method. Moverover, its operation cost of the second stage is US$ 1.5992×10^6, lower than US$ 1.7269×10^6 regarding the CRO method. The above results show that in comparison with the CRO method, the NRO method copes with the uncertainties by means of the lower reserve capacity, and the optimal results is less conservative. Therefore, the performance of the NRO to cope with the uncertainties is better than the CRO method.

Case 2 – (Comparative Analysis of Various Uncertainty Sets under NRO framework): In this case, we compare the experimental results for different uncertainty sets under the NRO framework. The uncertainty sets contain DPMM-based uncertainty set, TCDPMM-based uncertainty set without considering the correlation between variables, and the proposed TCDPMM-based uncertainty set. For brevity, these uncertainty sets under NRO are represented by Method 1, Method 2 and Method 3. To verify the advantages of Method 3, the impacts of budget value for the uncertainty set on the total cost are elaborately analyzed by setting the budget value to be 10, 15 and 20, respectively. The optimization results with different budget values are represented in Table 4.

It can be observed from Table 4 that the total cost of the system operations increases as the budget value increases. It can be revealed that the higher the value of Φ is, the more severe the worst uncertainty case is, which leads to a conservatism increase in the system operation. Moreover, we can notice that the total cost in Method 3 is US$ 2.0671× 10^6, US$ 2.0875× 10^6 and US$ 2.106× 10^6, lower than US$ 2.2231× 10^6, US$ 2.2414× 10^6 and US$ 2.2593× 10^6 regarding Method 1, when budget value Φ is set to 10, 15, 20. Apparently, the economic performance of Method 3 is better than Method1. The reasons mainly include two points: 1) the TCDPMM-based uncertainty set contains fewer data outliers than the DPMM-based uncertainty set, while the data outliers lead to higher conservatism, 2) DPMM-based components contain some small probability of components, such as the probability of c_8 is only 0.038. It means that the DPMM-based uncertainty set contains a large number of small probability scenarios, which leads to a higher cost for the system to cope with uncertainty. In addition, it is notable that the total cost in Method 3 decreased by 7%, 6.9% and 6.8% compared to Method 1. It indicates that the optimization results of Method 3 are much lower than Method 1 in terms of conservativeness.

The advantage of considering correlations between variables is verified by comparing Method 3 with Method 2. It can be found from Table 4 that the total cost for Method 2 is higer than that of Method 3. This is because that the range of the uncertainty set based on TCDPMM becomes smaller after considering the correlation between variables. To be specific, taking the wind power uncertainty set in week 43 as an example, the uncertainty set is represented by $P_{\text{wt}} = 18.04+5.54z_1, |z_1| \leq 1$ when the correlation between wind power and load demand is considered. On the contrary, the uncertainty set can be represented by $P_{\text{wt}} = 18.04 + 16.39z_1, |z_1| \leq 1$ when the correlation is not considered. For ease of understanding, these two uncertainty sets are further elaborated by Fig. 10. As shown in Fig. 10, the upper bound of the uncertainty set becomes smaller and the lower bound becomes larger after considering the correlation between the uncertain variables. Therefore, the output range of wind power becomes smaller, which reduces the cost of the system to cope with the uncertainty and reduces the conservativeness of robust optimization for GMS. In summary, the TCDPMM-based two-stage NRO method not only considers the temporal information in the dataset and the correlation between the uncertain variables, but also reduces the conservativeness of robust optimization, which decreases the cost of dealing with the uncertainties of wind power and load demand.

Fig. 10: Comparison of the uncertainty sets

TABLE 4
Cost comparison of different methods

Budget	Φ=10	Φ=15	Φ=20
Method 1 ($)	2.2231× 10^6	2.2414× 10^6	2.2593× 10^6
Method 2 ($)	2.0852× 10^6	2.1131× 10^6	2.1413× 10^6
Method 3 ($)	**2.0671 × 10^6**	**2.0875× 10^6**	**2.1060× 10^6**

5.4. The 26-Bus Yantai Power System

In this subsection, we consider a 26-bus Yantai power system in China to verify the effectiveness of the proposed data-driven NRO model based on TCDPMM. The topology of the system is shown in Fig. 11, which consists of 26 buses, 4 generators, 39 transmission lines, and 19 loads. Moreover, two wind farms are installed at Buses 2 and 12. The units parameters are origin from Ref. [39], and the line parameters of the Yantai power grid can be found in Ref. [40]. The weekly load demand and wind power generation are shown in Fig. 12, and the data-driven uncertainty set is the same as in the previous study on the six-bus system.

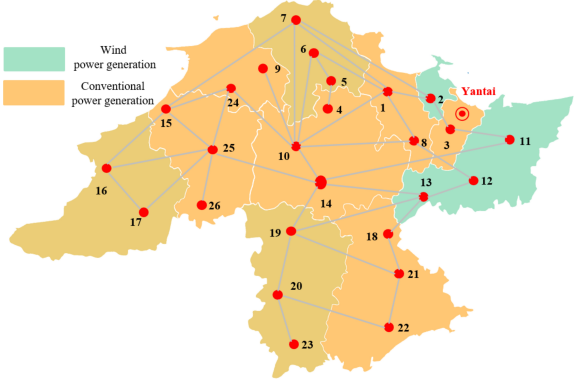


Fig. 11: 26-bus Yantai power system

For brevity, Φ is set to be 15, and Fig. 13 presents the deviation of wind power generation and load demand when the proposed NRO approach is applied to the Yantai power grid. It can be found from Fig. 13 that the volatility of load demand is positive, and that of wind power generation is

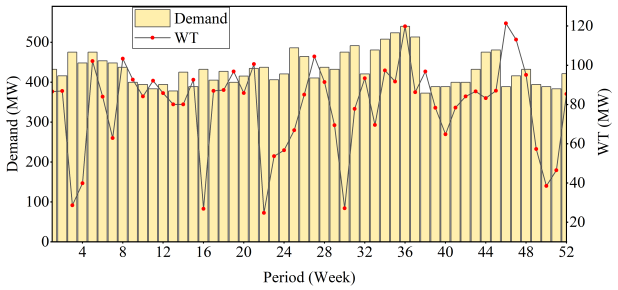


Fig. 12: Load demand and wind generation

negative. This indicates that the load demand is increased and the wind power generation is reduced in the proposed NRO model for GMS, which corresponds to the worst-case uncertainty scenario in the RO method.

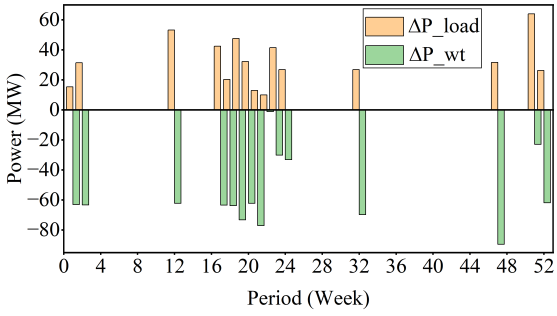


Fig. 13: The deviation of load demand and wind generation when $\Phi = 15$

To verify the economics of the proposed NRO approach, the total costs of different uncertainty sets under the NRO framework are presented in Table 5. Note that Method 1, Method 2 and Method 3 in Table 5 are consistent with the methods in Case 2, and Φ is set to be 10, 15, and 20, respectively. By comparing the total costs, we can see from Table 5 that the total cost in Method 2 is US$ 8.74× 10^5, US$ 8.98× 10^5 and US$ 9.22× 10^5, lower than US$ 9.69× 10^5, US$ 9.88× 10^5 and US$ 1.01× 10^6 regarding Method 1. The percentage of cost reduction is 9.8%, 9.1% and 8.7%,respectively. This reveals that the cost of the system coping with the uncertainty of wind power and load demand can be reduced when the temporal information of the uncertainty variables is considered.

Moreover, we can find from Table 5 that the total cost determined by Method 3 is US$ 8.54× 10^5, US$ 8.69× 10^5 and US$ 8.83× 10^5, lower than US$ 8.74× 10^5, US$ 8.98× 10^5 and US$ 9.22× 10^5 regarding that of Method 2. Obviously, the total cost in Method 3 is further reduced. It is because that the correlation between the uncertain variables is characterized in the proposed data-driven uncertainty set, which reduces the conservativeness of robust optimization. Therefore, it can be concluded that considering the temporal correlation of wind power and load demand, the TCDPMM-based two-stage NRO method for GMS has better economic performance, since the cost of tackling the uncertainties is reduced. In this way, the effectiveness and superiority of the proposed approach are verified in the Yantai power system.

TABLE 5
Cost comparison of different methods

Budget	Φ=10	Φ=15	Φ=20
Method 1 ($)	9.69× 10^5	9.88× 10^5	1.01× 10^6
Method 2 ($)	8.74× 10^5	8.98× 10^5	9.22× 10^5
Method 3 ($)	**8.54 × 10^5**	**8.69× 10^5**	**8.83× 10^5**

6. CONCLUSION

In this paper, we present a data-driven two-stage NRO framework for GMS in the power system to cope with the uncertainties of wind power and load demand. As the limitations of traditional DPMM information characterization, the TCDPMM, an improved DPMM method, is developed to characterize the temporal correlation of the uncertain variables. With the Gaussian component results of TCDPMM and variational Bayesian inference, the data-driven uncertainty set is constructed, which captures the temporal information in the uncertainties datasets and correlations between wind power generation and load demand. Moreover, with data-driven uncertainty set embedded in the robust optimization model, the two-stage problem for GMS is reformulated as a "min-max-max-min" problem, in which the NRO model is solved by the introduced parallel Benders' decomposition algorithm, and the optimal maintenance decisions are determined under the worst case. The simulation results show that 1) the proposed TCDPMM has a better performance in terms of mining the temporal information in the datasets compared to the DPMM method, since the inference results of TCDPMM contain fewer components, 2) compared to the deterministic GMS model that does not consider the uncertainties, the developed NRO model is better to improve the security and stability of the system at the cost of lower economic cost, and 3) the proposed NRO method for GMS has a lower conservatism compared to the DPMM based NRO method and TCDPMM based NRO method without considering correlation, which helps reduce the operation cost to deal with uncertainties.

The proposed NRO for GMS method has two limitations, as follows:

(1) There exists the spatial correlation of wind power generator output in the same region of the power system. However, it is not considered in the proposed two-stage NRO model for GMS, which may impede the economic operation of the power system.

(2) The uncertainties in GMS are from wind power, photovoltaic power generation, load, and the contingencies in the transmission network, etc. Only wind power and load uncertainties are considered in the proposed model, and the correlations among multiple uncertain variables needs to be further studied.

Therefore, the future work will focus on the above two limitations.

References

[1] M. Z. Gargari, M. T. Hagh, S. G. Zadeh, Preventive maintenance scheduling of multi energy microgrid to enhance the resiliency of system, Energy 221 (2021) 119782.

[2] N. Hoseini, A. Sheikholeslami, T. Barforoushi, M. A. Latify, Preventive maintenance mid-term scheduling of resources in multi-carrier energy systems, Energy 197 (2020) 117164.

[3] C. G. Min, M. K. Kim, J. K. Park, Y. T. Yoon, Game-theory-based generation maintenance scheduling in electricity markets, Energy 55 (15) (2013) 310–318.

[4] Y. Wang, H. Zhong, Q. Xia, D. S. Kirschen, C. Kang, An approach for integrated generation and transmission maintenance scheduling considering n-1 contingencies, IEEE Transactions on Power Systems 31 (3) (2016) 2225–2233.

[5] C. Wang, Z. Wang, Y. Hou, K. Ma, Dynamic game-based maintenance scheduling of integrated electric and natural gas grids with a bilevel approach, IEEE Transactions on Power Systems 33 (5) (2018) 4958–4971.

[6] Y. Li, Y. Ding, Y. Liu, T. Yang, P. Wang, J. Wang, W. Yao, Dense skip attention based deep learning for day-ahead electricity price forecasting, IEEE Transactions on Power Systems (2022) 1–19.

[7] L. Ge, H. Liu, J. Yan, X. Zhu, S. Zhang, Y. Li, Optimal integrated energy system planning with dg uncertainty affine model and carbon emissions charges, IEEE Transactions on Sustainable Energy 13 (2) (2022) 905–918.

[8] A. R.-K. M. A. Tajeddini, A. Soroudi, Risk averse optimal operation of a virtual power plant using two stage stochastic programming, Energy 73 (2014) 958–967.

[9] F. Fallahi, M. Yildirim, J. Lin, C. Wang, Predictive multi-microgrid generation maintenance formulation and impact on operations & resilience, IEEE Transactions on Power Systems 36 (6) (2021) 4979–4991.

[10] Y. Shang, W. Wu, J. Liao, J. Guo, J. Su, W. Liu, Y. Huang, Stochastic maintenance schedules of active distribution networks based on monte-carlo tree search, IEEE Transactions on Power Systems 35 (5) (2020) 3940–3952.

[11] G. Ji, W. Wu, B. Zhang, Robust generation maintenance scheduling considering wind power and forced outages, IET Renewable Power Generation 10 (5) (2016) 634–641.

[12] B. Bagheri, N. Amjady, Adaptive-robust multi-resolution generation maintenance scheduling with probabilistic reliability constraint, IET Generation Transmission & Distribution 13 (15) (2019) 3292–3301.

[13] S. D. Manshadi, M. E. Khodayar, Risk-averse generation maintenance scheduling with microgrid aggregators, IEEE Transactions on Smart Grid 9 (6) (2018) 6470–6479.

[14] B. Bagheri, N. Amjady, S. Dehghan, Multiscale multiresolution generation maintenance scheduling: A stochastic affinely adjustable robust approach, IEEE Systems Journal 15 (1) (2021) 893–904.

[15] Y. Li, Y. Zhao, L. Wu, Z. Zhigang, Artificial intelligence enabled computational methods for smart grid forecast and dispatch, Springer (2023).

[16] Y. Gao, X. Xu, Z. Yan, M. Shahidehpour, Gaussian mixture model for multivariate wind power based on kernel density estimation and component number reduction, IEEE Transactions on Sustainable Energy 13 (3) (2022) 1853–1856.

[17] W. Sun, M. Zamani, M. R. Hesamzadeh, H.-T. Zhang, Data-driven probabilistic optimal power flow with nonparametric bayesian modeling and inference, IEEE Transactions on Smart Grid 11 (2) (2020) 1077–1090.

[18] X. Zheng, H. Chen, Data-driven distributionally robust unit commitment with wasserstein metric: Tractable formulation and efficient solution method, IEEE Transactions on Power Systems 35 (6) (2020) 4940–4943.

[19] C. Ning, F. You, Deep learning based distributionally robust joint chance constrained economic dispatch under wind power uncertainty, IEEE Transactions on Power Systems 37 (1) (2022) 191–203.

[20] C. Duan, L. Jiang, W. Fang, J. Liu, Data-driven affinely adjustable distributionally robust unit commitment, IEEE Transactions on Power Systems 33 (2) (2018) 1385–1398.

[21] Y. Chen, Q. Guo, H. Sun, Z. Li, W. Wu, Z. Li, A distributionally robust optimization model for unit commitment based on kullback–leibler divergence, IEEE Transactions on Power Systems 33 (5) (2018) 5147–5160.

[22] C. Wang, Z. Gong, C. He, H. Gao, T. Bi, Data-driven adjustable robust unit commitment of integrated electric-heat systems, IEEE Transactions on Power Systems 36 (2) (2021) 1385–1398.

[23] H. Qiu, W. Gu, X. Xu, G. Pan, P. Liu, Z. Wu, L. Wang, A historical-correlation-driven robust optimization approach for microgrid dispatch, IEEE Transactions on Smart Grid 12 (2) (2021) 1135–1148.

[24] Z. Zhao, J. Guo, X. Luo, C. S. Lai, P. Yang, L. L. Lai, P. Li, J. M. Guerrero, M. Shahidehpour, Distributed robust model predictive control-based energy management strategy for islanded multi-microgrids considering uncertainty, IEEE Transactions on Smart Grid 13 (3) (2022) 2107–2120.

[25] J. Li, M. E. Khodayar, J. Wang, B. Zhou, Data-driven distributionally robust co-optimization of p2p energy trading and network operation for interconnected microgrids, IEEE Transactions on Smart Grid 12 (6) (2021) 5172–5184.

[26] F. Fang, S. Yu, X. Xin, Data-driven-based stochastic robust optimization for a virtual power plant with multiple uncertainties, IEEE Transactions on Power Systems 37 (1) (2022) 456–466.

[27] H. Saberi, C. Zhang, Z. Y. Dong, Data-driven distributionally robust hierarchical coordination for home energy management, IEEE Transactions on Smart Grid 12 (5) (2021) 4090–4101.

[28] Z. Yuan, P. Li, Z. Li, J. Xia, Data-driven risk-adjusted robust energy management for microgrids integrating demand response aggregator and renewable energies, IEEE Transactions on Smart Grid 14 (1) (2023) 365–377.

[29] X. Xu, Z. Yan, M. Shahidehpour, Z. Li, M. Yan, X. Kong, Data-driven risk-averse two-stage optimal stochastic scheduling of energy and reserve with correlated wind power, IEEE Transactions on Sustainable Energy 11 (1) (2020) 436–447.

[30] W. Sun, M. Zamani, H. Zhang, Y. Li, Probabilistic optimal power flow with correlated wind power uncertainty via markov chain quasi-monte-carlo sampling, IEEE Transactions on Industrial Informatics 15 (11) (2019) 6058–6069.

[31] Y. Yang, W. Wu, B. Wang, M. Li, Analytical reformulation for stochastic unit commitment considering wind power uncertainty with gaussian mixture model, IEEE Transactions on Power Systems 35 (4) (2020) 2769–2782.

[32] J. Wang, C. Wang, Y. Liang, T. Bi, M. Shafie-khah, J. P. S. Catalão, Data-driven chance-constrained optimal gas-power flow calculation: A bayesian nonparametric approach, IEEE Transactions on Power Systems 36 (5) (2021) 4683–4698.

[33] Y. Li, J. Huang, Y. Liu, T. Zhao, Y. Zhou, Y. Zhao, Day-ahead risk averse market clearing considering demand response with data-driven load uncertainty representation: A singapore electricity market study, Energy 254 (2022) 123923.

[34] Z. Ma, K. R. Pravin, T. Jalil, F. Markus, L. Arne, Bayesian estimation of dirichlet mixture model with variational inference, Pattern Recognition 47 (9) (2014) 3143–3157.

[35] A. Kheirandish, A. Fatehi, M. S. Gheibi, Identification of slow-rate integrated measurement systems using expectation–maximization algorithm, IEEE Transactions on Instrumentation and Measurement 69 (12) (2020) 9477–9484.

[36] C. Ning, F. You, Data-driven stochastic robust optimization: A general computational framework and algorithm for optimization under uncertainty in the big data era, Computers & Chemical Engineering 111 (2018) 115–133.

[37] T. A. T. Nguyen, S. Chou, T. H. K. Yu, Developing an exhaustive optimal maintenance schedule for offshore wind turbines based on risk-assessment, technical factors and cost-effective evaluation, Energy 249 (2022) 123613.

[38] R. Jiang, J. Wang, Y. Guan, Robust unit commitment with wind power and pumped storage hydro, IEEE Transactions on Power Systems 27 (2) (2012) 800–810.

[39] P. M. Subcommittee, IEEE Reliability Test System, IEEE Transactions on Power Apparatus and Systems PAS-98 (6) (1979) 2047–2054.

[40] Z. Li, X. Han, M. Yang, Y. Lei, K. Kang, Power system dispatch considering wind power grid integration, Automation of Electric Power Systems 34 (19) (2010) 15–19.

Appendix A Variational distributions of $q(\boldsymbol{\pi})$ and $q(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

In TCDPMM, $q(\boldsymbol{\pi})$ is a Dirichlet distribution that can be represented by:

$$q(\boldsymbol{\pi}) = Dir(\boldsymbol{\pi}|\boldsymbol{\alpha}) \tag{A.1}$$

where $\boldsymbol{\alpha}$ has components α_k updated by:

$$\alpha_k = \alpha_0 + N_k \tag{A.2}$$

Since TCDPMM consists of multiple Gaussian distributions, the variational distribution $q(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ can be reformulated as:

$$q(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \prod_{k=1}^{K} q(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{A.3}$$

The posterior of $q(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ is a Gaussian-Wishart distribution which can be given by

$$q^*(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \mathcal{N}(\boldsymbol{\mu}_k|\boldsymbol{m}_k, (\beta_k\boldsymbol{\Sigma}_k)^{-1})W(\boldsymbol{\Sigma}_k|\boldsymbol{\Psi}_k, \upsilon_k) \tag{A.4}$$

where $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ are the mean vector and precision matrix in TCDPMM, respectively. Distinctively, $\boldsymbol{m}_k$ and $\boldsymbol{\Psi}_k$ are the mean vector and precision matrix in the Gaussian-Wishart distribution. β_k and υ_k are parameters in the Gaussian-Wishart distribution. Moreover, the related parameters can be calculated by :

$$\beta_k = \beta_0 + N_k \tag{A.5}$$

$$\boldsymbol{m}_k = \frac{1}{\beta_k}(\beta_0\boldsymbol{m}_0 + N_k\overline{\boldsymbol{x}}_k) \tag{A.6}$$

$$\boldsymbol{W}_k^{-1} = \boldsymbol{W}_0^{-1} + N_k\boldsymbol{S}_k + \frac{\beta_0 N_k}{\beta_0 + N_k}(\overline{\boldsymbol{x}}_k - \boldsymbol{m}_0)(\overline{\boldsymbol{x}}_k - \boldsymbol{m}_0)^T \tag{A.7}$$

$$\upsilon_k = \upsilon_0 + N_k \tag{A.8}$$

$$N_k = \sum_{i=1}^{I} r_k'^i \tag{A.9}$$

$$\overline{\boldsymbol{x}}_k = \frac{1}{N_k}\sum_{i=1}^{I} r_k'^i \boldsymbol{x}_i \tag{A.10}$$

$$\boldsymbol{S}_k = \frac{1}{N_k}\sum_{i=1}^{I} r_k'^i(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_k)(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_k)^T \tag{A.11}$$

where β_0 and υ_0 are the initial values for β_k and υ_k. m_0 is the initial vector for m_k.

On this basis, ρ_k^i is defined as below:

$$\ln\rho_k^i = E[\ln\pi_k] + \frac{1}{2}E[\ln|\boldsymbol{\Sigma}_k|] - \frac{D}{2}\ln(2\pi) - \frac{1}{2}E_{\mu_k,\boldsymbol{\Sigma}_k}[(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k(\boldsymbol{x}_i - \boldsymbol{\mu}_k)] \tag{A.12}$$

$$E[\ln\pi_k] = \psi(\alpha_k) - \psi(\widehat{\alpha}) \tag{A.13}$$

$$E[\ln|\boldsymbol{\Sigma}_k|] = \sum_{n=1}^{D}\psi\left(\frac{\upsilon_k + 1 - n}{2}\right) + D\ln 2 + \ln|\boldsymbol{W}_k| \tag{A.14}$$

$$E_{\mu_k,\boldsymbol{\Sigma}_k}[(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k(\boldsymbol{x}_i - \boldsymbol{\mu}_k)] = D\beta_k^{-1} + \upsilon_k(\boldsymbol{x}_i - \boldsymbol{m}_k)^T\boldsymbol{W}_k(\boldsymbol{x}_i - \boldsymbol{m}_k) \tag{A.15}$$

where D is the dimensionality of the observation $\boldsymbol{x}$, and $\psi(\cdot)$ is the digamma function.

Appendix B EM algorithm for TCDPMM

In general, EM algorithm is divided into Expectation-step and Maximization step. In Expectation-step, the weight $r_k'^i$ can be calculated when the parameters $\boldsymbol{\pi}$ and $\boldsymbol{\theta}$ are fixed. In other words, the variational distribution $q(\boldsymbol{z})$ can be determined. Furthermore, keeping the weight $r_k'^i$ fixed, the parameters $\boldsymbol{\pi}$ and $\boldsymbol{\theta}$ are updated in Maximization step. It means that the variational distributions $q(\boldsymbol{\pi})$ and $\boldsymbol{\Sigma}$ are obtained. The optimal posterior distributions $q^*(\boldsymbol{z})$, $q^*(\boldsymbol{\pi})$ and $q^*(\boldsymbol{\theta})$ are determined when convergence is reached at the Expectation and Maximization steps. That is to say, the posterior distribution $p(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{\theta})$ is approximated by $q^*(\boldsymbol{z})q^*(\boldsymbol{\pi})q^*(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The EM algorithm for TCDPMM is shown in **Algorithm** 2.

Algorithm 2 The EM Algorithm for TCDPMM

1: Input observation dataset $X = \{x_i\}_{i=1}^{I}$;
2: Initialize the mean vector $\boldsymbol{m}_0$, the precision matrix $\boldsymbol{W}_0$, the parameters α_0, β_0 and υ_0;
3: **while** $\left|\frac{\theta_t - \theta_{t-1}}{\theta_{t-1}}\right| \geq \varsigma$ **do**
4: **for** $k = 1$ to M **do**
5: Evaluate the weight r_k^i by the expression in (8);
6: Modify the weight $r_k'^i$ by the expression in (13);
7: **end for**
8: **for** $i = 1$ to I **do**
9: Calculate N_k, $\overline{\boldsymbol{x}}_k$ and $\boldsymbol{S}_k$ by the expression in (A.9), (A.10) and (A.11);
10: **end for**
11: **for** $k = 1$ to M **do**
12: Update β_k, $\boldsymbol{m}_k$, $\boldsymbol{W}_k^{-1}$ and υ_k by the expression in (A.5), (A.6), (A.7) and (A.8);
13: **end for**
14: **end while**
15: Output $p(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{\theta}) = \prod_{i=1}^{I}\sum_{k=1}^{\infty} N_k(x_i|\mu_k, \Sigma_k)^{z_k^{[i]}}$

Note that lines 4 to 7 represent the Expectation-step, and the lines 8 to 13 stand for the Maximization step in **Algorithm** 2. ς is the gap, which can affect the number of iterations for EM algorithm and the accuracy of the optimal results. The values of $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ is equal to the value of $\boldsymbol{m}_k$ and $\boldsymbol{W}_k$ when the EM algorithm reaches convergence.

Appendix C Conventional two-stage robust optimization and deterministic model for GMS

The conventional two-stage robust optimization (CRO) model for is shown as below.

$$\min_{x,g} \sum_{i=1}^{n_G}\sum_{t=1}^{T} \overbrace{(c_{M,i}x_{i,t} + c_{R,i}(1-g_{i,t}))}^{\textbf{The first term}} + \max_{P_{wt},P_{load}\in U} \min_{P_G} \sum_{i=1}^{n_G}\sum_{t=1}^{T} \underbrace{(c_{O,Gi}+c_{F,Gi})P_{Gi}}_{\textbf{The second term}} \tag{C.1}$$

$$\textbf{s.t.}\ (19)-(30) \tag{C.2}$$

$$u_{wt}(t) = P_{wt}(t) - 0.2 * B_{wt}(t) * P_{wt}(t) \tag{C.3}$$

$$u_{Load}(t) = P_{Load}(t) + 0.2 * B_{Load}(t) * P_{Load}(t) \tag{C.4}$$

$$\sum_{t=1}^{T} B_{wt}(t) \le \Phi_{wt} \tag{C.5}$$

$$\sum_{t=1}^{T} B_{Load}(t) \le \Phi_{Load} \tag{C.6}$$

$$U = u_{wt} \cup u_{Load} \tag{C.7}$$

where B_{wt} and B_{Load} are binary variables. Φ_{wt} and Φ_{Load} represent the uncertainty adjustment parameters of the wind power and load.

Unlike the two-stage NRO model for GMS, the wind power generation and load demand are determined in the deterministic GMS model, which means that the uncertainties of wind power and load is not considered. The specific expression of the deterministic model for GMS is shown below.

$$\min_{x,y} \sum_{i=1}^{n_G}\sum_{t=1}^{T} \{c_{M,i}x_{i,t} + c_{R,i}(1-g_{i,t}) + (c_{O,Gi}+c_{F,Gi})P_{Gi}\} \tag{C.8}$$

$$\textbf{s.t.}\ (19)-(30) \tag{C.9}$$

Appendix D DPMM and VBI for DPMM

D.1 DPMM

An I-dimensional vector $\boldsymbol{x} = \{x_i\}_{i=1}^{I}$can represent the observation data, and the probability density function of $\boldsymbol{x}$ is expressed as follows.

$$p(\boldsymbol{x}|\boldsymbol{\pi},\boldsymbol{\theta}) = \sum_{k=1}^{\infty} \pi_k p_k(\boldsymbol{x}|\theta_k) \tag{D.1}$$

where π_k, θ_k, and $p_k(x|\theta_k)$ represent the weight, parameter and probability density function of the component k. Here, the prior of the component k is set to be a Gaussian distribution. Therefore, θ_k stands for the mean and covariance of the Gaussian distribution, and (D.1) is reformulated as below.

$$p(\boldsymbol{x}|\boldsymbol{\pi},\boldsymbol{\theta}) = \sum_{k=1}^{\infty} \pi_k \mathcal{N}_k(\boldsymbol{x}|\mu_k,\Sigma_k) \tag{D.2}$$

where $\mathcal{N}_k(\cdot|\cdot)$ is the Gaussian component k. μ_k and Σ_k are the mean and covariance of the component k, and $\theta_k = (\mu_k,\Sigma_k)$. Note that $\boldsymbol{x} = \{x_i\}_{i=1}^{I}$ contains I observation data which is mutually independent samples. Hence, the DPMM on the observation data is represented as follows.

$$p(\boldsymbol{x}|\boldsymbol{\pi},\boldsymbol{\theta}) = \prod_{i=1}^{I}\left\{\sum_{k=1}^{\infty} \pi_k \mathcal{N}_k(x_i|\mu_k,\Sigma_k)\right\} \tag{D.3}$$

Here, it is not clear which Gaussian component the observation data x_i belongs to. To address this issue, the latent variable vector $\boldsymbol{z} = \{z_k^{[i]}\}_{i=1,k=1}^{i=I,k=\infty}$ is introduced to indicate the association of the observed data with the Gaussian component, and $\boldsymbol{z}$ is the binary variable vector. For instance, $z_k^{[i]} = 1$ shows that the observation data x_i belongs to the Gaussian component k. In this way, it is known that the observation data x_i follows which Gaussian component, which means that the latent variable vector $\boldsymbol{z}$ and the parameter set $\boldsymbol{\theta}$ are given. Therefore, (D.3) is formulated as follows.

$$p(\boldsymbol{x}|\boldsymbol{z},\boldsymbol{\theta}) = \prod_{i=1}^{I}\prod_{k=1}^{\infty} N_k(x_i|\mu_k,\Sigma_k)^{z_k^{[i]}} \tag{D.4}$$

The relationship between $\boldsymbol{z}$ and $\boldsymbol{\pi}$ is as follows.

$$p(\boldsymbol{z}|\boldsymbol{\pi}) = \prod_{i=1}^{I}\prod_{k=1}^{\infty} \pi_k^{z_k^{[i]}} \tag{D.5}$$

D.2 VBI for DPMM

In section D.1, the mathematical model of DPMM is constructed. VBI is introduced to estimate the parameters of the model in this section. Note that we need to find the maximum of the posterior probability $p(\boldsymbol{x})$ when the parameters set $\Omega = \{\boldsymbol{\pi},\boldsymbol{z},\boldsymbol{\theta}\}$ is given. Since the posterior distribution p is difficult to be determined directly, the variational distribution $q(\Omega)$ is introduced. In this way, the logarithm of p is expressed as follows.

$$\begin{aligned}\ln p(\boldsymbol{x}) &= \int \ln\frac{p(\boldsymbol{x},\Omega)}{p(\Omega|\boldsymbol{x})} \times q(\Omega)d\Omega \\ &= \int q(\Omega)\{\ln p(\boldsymbol{x},\Omega) - \ln p(\Omega|\boldsymbol{x})\}d\Omega \\ &= \int q(\Omega)\ln\frac{p(\boldsymbol{x},\Omega)}{q(\Omega)}d\Omega - \int q(\Omega)\ln\frac{p(\Omega|\boldsymbol{x})}{q(\Omega)}d\Omega\end{aligned} \tag{D.6}$$

Here, the evidence lower bound and Kullback-Leibler (KL) divergence are denoted as follows.

$$\text{ELBO} = \int q(\Omega)\ln\frac{p(\boldsymbol{x},\Omega)}{q(\Omega)}d\Omega \tag{D.7}$$

$$\text{KL}(q(\Omega)||p(\Omega|\boldsymbol{x})) = -\int q(\Omega)\ln\frac{p(\Omega|\boldsymbol{x})}{q(\Omega)}d\Omega \tag{D.8}$$

Therefore, (D.6) can be reformulated as

$$\ln p(\boldsymbol{x}) = \text{ELBO} + \text{KL}(q(\Omega)||p(\Omega|\boldsymbol{x})) \tag{D.9}$$

Here, in order to maximize $\ln p(\boldsymbol{x})$, the variational distribution $q(\boldsymbol{\Omega})$ is used to approximate $p(\Omega|\boldsymbol{x})$, which can help minimize $\text{KL}(q(\Omega)||p(\Omega|\boldsymbol{x}))$. Note that the posterior probability $p(\boldsymbol{x})$ in (D.9) is a constant, and the KL divergence is non-negative. Hence, the minimization of the KL divergence

can be transformed into the maximization of the ELBO. That is to say, the maximization of $\ln p(\boldsymbol{x})$ is transformed into the maximization of the ELBO. To solve the maximum value of the ELBO, the parameters in q including $\boldsymbol{\pi}$, $\boldsymbol{z}$ and $\boldsymbol{\theta}$ are assumed to be mutually independent. In this way, $q(\Omega)$ is decomposed as follows.

$$q(\Omega) = q(\boldsymbol{z})q(\boldsymbol{\pi})q(\boldsymbol{\theta}) \tag{D.10}$$

Next, (D.7) can be transformed as

$$\begin{aligned}
\text{ELBO} &= \int q(\Omega) \ln \frac{p(\boldsymbol{x}, \Omega)}{q(\Omega)} d\Omega \\
&= \int q(\boldsymbol{z})q(\boldsymbol{\pi})q(\boldsymbol{\theta}) \left\{\ln p(\boldsymbol{x}, \Omega) - \ln q(\Omega)\right\} d\Omega \\
&= \int q(\boldsymbol{z}) \left\{ \int q(\boldsymbol{\pi})q(\boldsymbol{\theta}) \ln p(\boldsymbol{x}, \Omega) d\boldsymbol{\pi} d\boldsymbol{\theta} \right\} d\boldsymbol{z} \\
&- \int q(\boldsymbol{z}) \ln q(\boldsymbol{z}) d\boldsymbol{z} + const \\
&= \int q(\boldsymbol{z}) \left\{ E_{\boldsymbol{\pi},\boldsymbol{\theta}} \left[\ln p(\boldsymbol{x}, \Omega)\right] - \ln q(\boldsymbol{z}) \right\} d\boldsymbol{z} + const
\end{aligned} \tag{D.11}$$

We calculate the first order partial derivative of (D.11) with respect to $q(z)$, which is shown as follows.

$$\frac{\partial ELBO}{\partial q(\boldsymbol{z})} = E_{\boldsymbol{\pi},\boldsymbol{\theta}} \left[\ln p(\boldsymbol{x}, \Omega)\right] - \ln q(\boldsymbol{z}) - 1 + const = 0 \tag{D.12}$$

Therefore, the optimal distribution of $q^*(\boldsymbol{z})$ can be obtained, which is as below.

$$\begin{aligned}
&\ln q^*(\boldsymbol{z}) = E_{\boldsymbol{\pi},\boldsymbol{\theta}} \left[\ln p(\boldsymbol{x}, \Omega)\right] + const \\
&= E_{\boldsymbol{\pi},\boldsymbol{\theta}} \left[\ln p(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\theta})\right] + const \\
&= E_{\boldsymbol{\pi},\boldsymbol{\theta}} \left[\ln p(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{\theta}) + \ln p(\boldsymbol{z}|\boldsymbol{\pi}) + \ln p(\boldsymbol{\pi}) + \ln p(\boldsymbol{\theta})\right] + const \\
&= E_{\boldsymbol{\pi}}[\ln p(\boldsymbol{z}|\boldsymbol{\pi})] + E_{\boldsymbol{\theta}}[\ln p(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{\theta})] + const
\end{aligned} \tag{D.13}$$

where *const* represents the constant terms not related to $\boldsymbol{z}$. Substituting (D.4) and (D.5) into (D.13), $\ln q^*(\boldsymbol{z})$ is represented by

$$\ln q^*(\boldsymbol{z}) = \sum_{i=1}^{I} \sum_{k=1}^{K} z_k^{[i]} \ln \rho_k^i + const \tag{D.14}$$

where ρ_k^i represents the probability that k-th Gaussian component samples observation data x_i, which is shown in (A.12). Further, taking logarithms on both sides, $q^*(\boldsymbol{z})$ is formulated by.

$$q^*(\boldsymbol{z}) = \prod_{i=1}^{I} \prod_{k=1}^{K} (r_k^i)^{z_k^{[i]}} \tag{D.15}$$

where r_k^i is the weight that the observation x_i belongs to k-th Gaussian component, and r_k^i can be obtained after normalizing ρ_k^i, which is represented by:

$$r_k^i = p(z_k^{[i]} = 1) = \frac{\rho_k^i}{\sum_{j=1}^{K} \rho_j^i} \tag{D.16}$$

Note that the optimal distributions of $q^*(\pi)$ and $q^*(\theta)$ in DPMM are considered to be the Dirichlet distribution and Gaussian-Wishart distribution, which can be shown as follows.

$$q^*(\boldsymbol{\pi}) = Dir(\boldsymbol{\pi}|\boldsymbol{\alpha}) \tag{D.17}$$

$$q^*(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \mathcal{N}(\boldsymbol{\mu}_k|\boldsymbol{m}_k, (\beta_k \boldsymbol{\Sigma}_k)^{-1}) W(\boldsymbol{\Sigma}_k|\boldsymbol{\Psi}_k, \upsilon_k) \tag{D.18}$$

where the parameter calculations for (D.17) and (D.18) are shown in Appendix A. Therefore, the optimal parameters π^* and θ^* can be obtained.

Importantly, our goal is to find the maximum value of the posterior probability $p(\boldsymbol{x})$, which is divided into Expectation step and Maximization step. In Expectation step, the optimal distribution of $q(\boldsymbol{z})$ is solved by (D.13), and the variational distribution $q^*(\boldsymbol{z})$ is fixed. On this basis, the optimal parameters $\boldsymbol{\pi}^*$ and $\boldsymbol{\theta}^*$ are obtained by (D.17) and (D.18) in Maximization step, which can help maximize $p(\boldsymbol{x})$. When Expectation step and Maximization step converge simultaneously, the maximum value of $p(\boldsymbol{x})$ is obtained. Meanwhile, $\boldsymbol{z}$, $\boldsymbol{\pi}$ and $\boldsymbol{\theta}$ are determined, which indicates that the posterior distribution $p(\boldsymbol{x}|\boldsymbol{\pi}, \boldsymbol{\theta})$ is obtained. The EM algorithm for DPMM is shown in **Algorithm** 3.

Algorithm 3 The EM Algorithm for DPMM

1: Input observation dataset $X = \{x_i\}_{i=1}^{I}$;
2: Initialize the mean vector $\boldsymbol{m}_0$, the precision matrix $\boldsymbol{W}_0$, the parameters α_0, β_0 and υ_0;
3: **while** $\left|\frac{\theta_t - \theta_{t-1}}{\theta_{t-1}}\right| \geq \varsigma$ **do**
4: **for** $k = 1$ to M **do**
5: Evaluate the weight r_k^i by the expression in (8);
6: **end for**
7: **for** $i = 1$ to I **do**
8: Calculate N_k, $\bar{\boldsymbol{x}}_k$ and $\boldsymbol{S}_k$ by the expression in (A.9), (A.10) and (A.11);
9: **end for**
10: **for** $k = 1$ to M **do**
11: Update β_k, $\boldsymbol{m}_k$, $\boldsymbol{W}_k^{-1}$ and υ_k by the expression in (A.5), (A.6), (A.7) and (A.8);
12: **end for**
13: **end while**
14: Output $p(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{\theta}) = \prod_{i=1}^{I} \sum_{k=1}^{\infty} N_k(x_i|\mu_k, \Sigma_k)^{z_k^{[i]}}$

Note that lines 4 to 6 represent the Expectation-step, and the lines 7 to 12 stand for the Maximization step in **Algorithm** 3. ς is the gap, which can affect the number of iterations for EM algorithm and the accuracy of the optimal results. The values of $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ is equal to the value of $\boldsymbol{m}_k$ and $\boldsymbol{W}_k$ when the EM algorithm reaches convergence.