2021 IEEE Sustainable Power and Energy Conference (iSPEC) | 978-1-6654-1439-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/iSPEC53008.2021.9735739

Regulation Capacity Evaluation of Large-scale Heterogeneous Residential Air Conditioning Loads

Hongxun Hui[1], Peipei Yu[1], Hongcai Zhang[1], Ningyi Dai[1], Wei Jiang[2], Yonghua Song[1]

1. State Key Laboratory of Internet of Things for Smart City, University of Macau, Macao, China;

2. School of Electrical Engineering, Southeast University, Nanjing, China.

***Abstract*—The rapidly increasing residential air conditionings (RACs) have been widely considered as good regulation resources for improving the power system flexibility. However, due to the heterogeneity of different consumers and the difficulty of data acquisition in reality, the evaluation of large-scale RACs' regulation capacity is a tricky problem. To address this issue, this paper proposes an evaluation method based on Gaussian Mixture Model (GMM). First, the regulation framework of large-scale RACs is developed based on partial observable data. Then, a calculation model of RACs' regulation capacity is proposed under the premise of guaranteeing heterogeneous consumers' comfort requirements. On this basis, we develop the GMM of RACs to evaluate their regulation capacities in the condition of insufficient data acquisition. The expectation maximization algorithm and Bayesian information criterion are utilized to optimize the multidimensional parameters of the GMM. Finally, we verify the proposed methods based on the realistic data in a demonstration project in China. The results show that the proposed method can evaluate the RACs' regulation capacity with more than 98% accuracy by measuring 1% RACs' parameters.**

***Index Terms*—Residential air conditioning, regulation capacity evaluation, demand response, Gaussian mixture model.**

I. Introduction

The increasing residential appliances have become the major growth source in power consumption and played a more important role in the power system [1]. Among different kinds of appliances, residential air conditionings (RACs) are one of the most important loads because they consume more than half of the energy at home [2]. However, the operating power of RACs is significantly affected by the changeable weather, which is resulting in sharper peak power and threatening the stable operation of the power system with the frequent occurrences of extreme hot weather around the world [3]. To maintain the balance between supply-side and demand-side, the power system needs larger regulation capacity [4]. Nowadays, the regulation capacity is mainly provided by generating units, e.g., thermal and gas generating units. However, with the increase of RACs in demand-side and uncontrollable renewable resources in supply-side, the regulation capacity in the near future may be insufficient [5].

The progressed Internet of Things technologies make the remote and smart control become easier to regulate large-scale demand-side resources for providing regulation services, which is called demand response [6]. Considering the principle that whoever started the trouble should end it, RACs account for a substantial part of residential power consumption and can provide huge regulation capacity to reduce the peak power [7]. Furthermore, RACs can be regulated within the consumers' comfortable indoor temperature requirements by utilizing the buildings' thermal inertia [8]. Hence, RACs have been widely considered as the most potential regulation resources in demand-side and become a research hotspot around the world. For example, Song et al. [9] model RACs as thermal batteries to work with lithium-ion batteries for providing regulation services for the power system. The RACs are also equivalent to traditional thermal generating units to participate in the frequency regulation services in [10]. Jiang et al. [11] develop a combined optimization method to control RACs for improving the power system flexibility.

Generally, RACs are regulated by a control center, which is held by the distribution system operator, the load serving entity, or the aggregator [12]. For convenience, the control center is uniformly called the aggregator of RACs in this paper. Before each round of dispatch, the aggregator should evaluate the available regulation capacity of RACs and sumbit this data to the system operator, so that the system operator can globally optimize different kinds of regulation resources and reserve sufficient regulation capacity for the system's stable operation [13]. However, compared with the regulation capacity provided by generating units, the regulation capacity provided by RACs is a complex problem. There are mainly three difficulties:

1) *Insufficient data acquisition*: One RAC's regulation capacity is influenced by lots of factors [14], including the building's thermal capacity and thermal resistance, the RAC's rated power and energy efficiency ratio (EER), the outdoor temperature, the real-time indoor temperature, and the comfortable temperature requirements. Some data can be easily detected while some data are difficult to be obtained by the aggregator. For example, the RAC's operating power and the building's indoor temperature can be detected by widely used smart meters and temperature sensors, respectively. However, the building's thermal capacity and resistance are related to such as the house area, volume, materials, windows and cracks, which are not easy to be comprehensively detected. Hence, aggregator has to evaluate the regulation capacity of RACs based on partial observable data.

This paper is funded in part by the Science and Technology Development Fund, Macau SAR (File no. SKL-IOTSC-2021-2023, and File no. 0137/2019/A3) and in part by University of Macau (File no. SRG2019-00162-IOTSC, and File no. CPG2021-00019-IOTSC). (*Corresponding author: Hongcai Zhang*, email: hczhang@um.edu.mo)

2) *Large-scale number of RACs*: The capacity of one thermal generating unit is around 60∼100MW or even 1000MW, while the rated power of one RAC is only 1∼3kW. That is to say, if RACs want to provide significant regulation capacities for the power system as generating units, around a number of 100,000 RACs have to be aggregated and regulated at the same time. The aggregator has to detect and process millions of parameters to evaluate the regulation capacity, which increases the computational complexity.
3) *Heterogeneity*: The consumers, RACs and buildings are all highly heterogeneous. For example, consumers have different comfortable requirements on their indoor temperature. RACs have different brands, rated power and EERs. Buildings have different areas, heights and materials. The heterogeneity increases the evaluation difficulty of large-scale RACs' regulation capacity, especially in the condition of insufficient data acquisition.

Some studies attempt to solve this problem. For example, Xie et al. [15] propose a probability density estimation method to calculate the RACs' regulation capacity. Lu [16] develops a regulation capacity calculation model of RACs by considering he consumers' diversity to provide intra-hour balancing services. However, these evaluation methods require to develop the accurate model of each RAC and building, which can not be used in the condition of insufficient data acquisition. Cai et al. [17] and Javed et al. [18] propose to utilize the most sophisticated artificial neural network for evaluating the demand response, while the complexity and computation efficiency of these methods increase the implementation difficulty in practical power systems.

To address aforementioned issues, this paper proposes a regulation capacity evaluation method of large-scale heterogeneous RACs based on Gaussian mixture model (GMM). The main contributions are summarized as follows:

1) We develop the regulation framework of large-scale RACs based on partial observable data. The calculation method of RACs' regulation capacity is proposed under the premise of guaranteeing heterogeneous consumers' comfort requirements.
2) We develop the GMM of RACs for evaluating regulation capacities in the condition of insufficient data acquisition. The expectation maximization (EM) algorithm and Bayesian information criterion (BIC) are utilized to optimize multidimensional parameters of the GMM.
3) The proposed models and methods are verified based on the realistic data in a demand response demonstration project in China. The results show that the evaluation accuracy of RACs' regulation capacity can reach 98% by measuring 1% RACs' parameters.

The remainder of this paper is organized as follows. Section II presents the framework of the RACs' regulation model. The modelling and optimization methodologies of the GMM are formulated in Section III. Numerical studies are presented in Section IV. Finally, Section V concludes this paper.

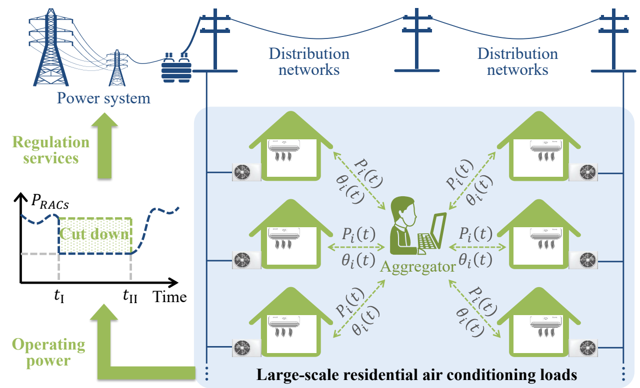

Fig. 1. The regulation framework of large-scale heterogeneous RACs.

II. Framework and Regulation Model of RACs

A. Regulation Framework of RACs

Fig. 1 shows the power system model and regulation framework of large-scale RACs. Each RAC is connected with the power system to get energy for generating cooling capacity. The consumers are at liberty to use RACs based on their temperature requirements. When RACs cause sharp peak power and threaten the system stable operation on extreme hot weather conditions, some RACs will be controlled to provide regulation services for the power system. As shown in Fig. 1, the operating power of RACs is cut down between t_{I} and t_{II} to decrease the peak power.

In order to increase the willingness of consumers for participating in regulation services, the comfortable requirements on their indoor temperatures are guaranteed all the time. Hence, each RAC's regulation power and duration time should be constrained in some ranges to avoid the uncomfortable indoor temperature. For example, if a building's indoor temperature wants to be maintained under 25°C, the corresponding RAC's regulation power and duration time may be constrained to be less than 1kW and 10min, respectively. Under the same indoor temperature constraints and ambient environment, this RAC's available regulation power may be smaller with the increase of duration time. Therefore, the RACs' available regulation capacities should consider the control instruction on duration time, which will be illustrated in detail in the next two subsections II-B and II-C.

Moreover, as the descriptions in Section I, the RACs' regulation capacity is related to lots of factors while some of data are difficult to be obtained. As shown in Fig. 1, here we choose two kinds of most common observable data to evaluate the regulation capacity: the RACs' operating power and the indoor temperature. The specific regulation capacity evaluation method will be shown in section III.

B. Thermal-electrical Model of RACs

The thermal model of buildings installed with RACs can be expressed as [16]:

$$C_i \frac{\partial \theta_i(t)}{\partial t} = \frac{\theta_o(t) - \theta_i(t)}{R_i} - Q_i(t), \quad \forall i \in \mathcal{I}, \forall t \in \mathcal{T}, \quad (1)$$

where $\theta_i(t)$ and $\theta_o(t)$ are the i-th building's indoor temperature and the outdoor ambient temperature at time t, respectively. Symbols C_i and R_i are the thermal capacity (in kJ/°C) and the thermal resistance (in °C/kW) of the i-th building, respectively. The prameter $\mathcal{I}$ is the set of RACs. The cooling capacity from RACs Q_i can be calculated as:

$$Q_i(t) = \eta_i P_i(t), \quad \forall i \in \mathcal{I}, \forall t \in \mathcal{T}, \tag{2}$$

where η_i and $P_i(t)$ are the EER and operating power of the i-th RAC, respectively. Generally, the values of EER distribute among 2.6∼3.6 [10]. With the incresae of EER, more cooling capacity can be generated from the consumed power energy .

When the indoor temperature is maintained to be equal to the set value, i.e., $\theta_i(t) = \theta_i^{\text{set}}(t)$, the corresponding RAC's operating power can be calculated from Eqs. (1)-(2) as:

$$P_i(t) = \frac{\theta_o(t) - \theta_i^{\text{set}}(t)}{\eta_i R_i}, \quad \forall i \in \mathcal{I}, \forall t \in \mathcal{T}. \tag{3}$$

C. Regulation Capacity of RACs Under Comfort Constraints

When RACs are controlled to provide regulation services for the power system, the most straight forward approach is to shut off the RACs directly. Then the regulation capacity can be easily obtained by summarizing all the RACs' operating power, as follows:

$$P_{\text{RACs}}(t) = \sum\nolimits_{i=1}^{N} P_i(t), \quad \forall i \in \mathcal{I}, \forall t \in \mathcal{T}. \tag{4}$$

However, this ON-OFF method can probably impact consumers' comforts significantly. Most research and practical projects have shifted to regulating RACs under the consumers' comfort constraints. In this paper, it is assumed that all the consumers can set their comfortable ranges of the indoor temperature, i.e., $\theta_i(t) \in [\theta_i^{\text{set}}(t) - \theta_i^{\text{dev}}(t),\ \theta_i^{\text{set}}(t) + \theta_i^{\text{dev}}(t)]$. Hence our problem in this paper is how to evaluate the available regulation capacity of RACs under the comfort constraints.

Based on Eqs. (1)-(2), the indoor temperature deviation during the regulation process can be obtained as:

$$\int_{t_{\text{I}}}^{t_{\text{II}}} C_i \mathrm{d}\theta_i(t) = \int_{t_{\text{I}}}^{t_{\text{II}}} \frac{\theta_o(t) - \theta_i(t)}{R_i} \mathrm{d}t - \int_{t_{\text{I}}}^{t_{\text{II}}} \eta_i P_i^{\text{II}}(t) \mathrm{d}t, \quad \forall i \in \mathcal{I}, \forall t \in \mathcal{T}, \tag{5}$$

where $P_i^{\text{II}}(t)$ is the i-th RAC's operating power during the regulation period.

It is assumed that the operating power $P_i^{\text{II}}(t)$ and the ambient temperature $\theta_o(t)$ keep constants during the regulation process. Then the Eq. (5) can be calculated as:

$$P_i^{\text{II}} = \frac{\theta_o^{\text{reg}}}{\eta_i R_i} - \frac{\theta_i^{\text{II}} + \theta_i^{\text{I}}}{2\eta_i R_i} - \frac{C_i(\theta_i^{\text{II}} - \theta_i^{\text{I}})}{\eta_i T_D}, \quad \forall i \in \mathcal{I}, \tag{6}$$

where T_D is the regulation duration time ($T_D = t_{\text{II}} - t_{\text{I}}$); θ_o^{reg} is the ambient temperature during the regulation process; θ_i^{I} and θ_i^{II} are the indoor temperature at the beginning and ending time of the regulation, respectively. The indoor temperature θ_i^{II} should be within the i-th consumer's comfort constraints, i.e., $\theta_i^{\text{II}} \in [\theta_i^{\text{set}}(t_{\text{II}}) - \theta_i^{\text{dev}}(t_{\text{II}}),\ \theta_i^{\text{set}}(t_{\text{II}}) + \theta_i^{\text{dev}}(t_{\text{II}})]$. In the up regulation services, the i-th RAC's maximum regulation capacity can be obtained by maximizing the indoor temperature, i.e., $\theta_i^{\text{II}} = \theta_i^{\text{set}}(t_{\text{II}}) + \theta_i^{\text{dev}}(t_{\text{II}})$. Then we can get the i-th RAC's maximum regulation capacity as:

$$P_i^{\text{reg}} = P_i^{\text{I}} - P_i^{\text{II}}, \quad \forall i \in \mathcal{I}, \tag{7}$$

where P_i^{I} is the i-th RAC's operating power at time t_{I}; P_i^{reg} is the regulation capacity provided by the i-th RAC. Hence, the total regulation capacity of RACs can be calculated by:

$$P_{\text{RACs}}^{\text{reg}} = \sum\nolimits_{i=1}^{N} P_i^{\text{reg}}, \quad \forall i \in \mathcal{I}. \tag{8}$$

III. Regulation Capacity Evaluation Method of Large-scale Heterogeneous RACs

A. Regulation Capacity Evaluation Based on GMM

As shown in Eqs. (6)-(8), the RAC's regulation capacity depends on the indoor temperature θ_i^{I} and θ_i^{II}, the building's thermal resistance R_i and thermal capacity C_i, the RAC's EER η_i, the outdoor temperature θ_o^{reg}, and the regulation duration time T_D. It is assumed that the aggregator can monitor RACs' operating power $P_i(t)$ by smart meters and buildings' indoor temperature $\theta_i(t)$ by temperature sensors, i.e., parameters P_i^{I}, θ_i^{I} and θ_o^{reg} can be detected at the beginning of the regulation. The regulation duration time T_D, the comfortable indoor temperature θ_i^{set} and θ_i^{dev} are preset values.

Compared with the above observable parameters and preset parameters, other parameters of the buildings (i.e., R_i, C_i), and RACs (i.e., η_i) are hard to be obtained by the aggregator, especially when the number of RACs reaches million level. Here we propose using the GMM to evaluate the total regulation capcity of RACs (i.e., $P_{\text{RACs}}^{\text{reg}}$) by sampling $\alpha\%$ parameters of RACs and corresponding buildings. Generally, the sampling share $\alpha\%$ is around 1% or even less. In this manner, the difficulty of data acquisition can be substantially reduced.

For convenience, this paper labels the sampled $\alpha\%$ of RACs as full-info RACs with the set $\mathcal{I}_\alpha \subset \mathcal{I}$, and labels the other RACs as semi-info RACs with the set $\mathcal{I}_\beta \subset \mathcal{I}$. The total set of RACs $\mathcal{I} = \mathcal{I}_\alpha + \mathcal{I}_\beta$. Therefore, the regulation capacities of RACs in the set $\mathcal{I}_\alpha$ can be calculated based on Eqs. (6)-(8), i.e., $P_{i,\alpha}^{\text{reg}}$ are known values. The key problem is evaluating the regulation capacity of RACs in the set $\mathcal{I}_\beta$, i.e., $P_{i,\beta}^{\text{reg}}$ are unknown values to be evaluated based on the following data:

1) The RACs' total number N, outdoor temperature θ_o^{reg} and regulation duration time T_D;
2) Parameters P_i^{I}, θ_i^{I}, θ_i^{set} and θ_i^{dev} of all the RACs ($\forall i \in \mathcal{I}$);
3) Parameters R_i, C_i and η_i of $\alpha\%$ RACs ($\forall i \in \mathcal{I}_\alpha \subset \mathcal{I}$).

B. Developing GMM for RACs

Firstly, the multidimensional vector of the full-info RACs' parameters in set $\mathcal{I}_\alpha$ can be formed as:

$$\boldsymbol{x}_i = [P_i^{\text{I}}, \theta_i^{\text{I}}, \theta_i^{\text{set}}, \theta_i^{\text{dev}}]^T, \quad \forall i \in \mathcal{I}_\alpha. \tag{9}$$

The total number of RACs in set $\mathcal{I}_\alpha$ is assumed to be N_α. Then the sample set of full-info RACs can be expressed as:

$$\mathcal{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_{N_\alpha}]. \tag{10}$$

As for an arbitrary vector $\boldsymbol{x}$ in $\mathcal{X}$, the joint probability density function (PDF) of the multidimensional vector $\boldsymbol{x}$ can be expressed as:

$$\mathcal{N}(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}|D|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{u})^T D^{-1}(\boldsymbol{x}-\boldsymbol{u})\right], \quad \forall \boldsymbol{x} \in \mathcal{X}, \tag{11}$$

where d is the dimension of the vector $\boldsymbol{x}$. Symbolos $\boldsymbol{u}$ and D are the mean value and covariance matrix of the vector $\boldsymbol{x}$, which can be expressed as:

$$\boldsymbol{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_d \end{bmatrix}, D = \begin{bmatrix} \delta_{11} & \delta_{12} & \cdots & \delta_{1d} \\ \delta_{21} & \delta_{22} & \cdots & \delta_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{d1} & \delta_{d2} & \cdots & \delta_{dd} \end{bmatrix}. \tag{12}$$

The GMM is generally composed of multiple PDFs, which can be calculated by:

$$f(\boldsymbol{x}) = \sum\nolimits_{k=1}^{K} \pi_k \cdot \mathcal{N}(\boldsymbol{x}|\boldsymbol{u}_k, D_k), \tag{13}$$

$$\sum\nolimits_{k=1}^{K} \pi_k = 1, \pi_k > 0, \quad \forall \boldsymbol{x} \in \mathcal{X}, \forall k \in \mathcal{K}, \tag{14}$$

where K is the total number of components in GMM. Symbol $\mathcal{N}(\boldsymbol{x}|\boldsymbol{u}_k, D_k)$ indicates the k-th PDF. Symbol π_k is the weight of the k-th PDF, which is also called the prior probability of choosing the k-th PDF. The summation of all the weights should be equal to 1.

Based on the GMM in Eq. (13), the regulation capacities of the i-th RAC in set $\mathcal{I}_\beta$ can be obtained by:

$$P_{i,\beta}^{\text{reg}} = \sum\nolimits_{k=1}^{K} p_k(\boldsymbol{x}_i) \cdot \overline{P_{k,\alpha}^{\text{reg}}}, \quad \forall i \in \mathcal{I}_\beta, \forall k \in \mathcal{K}, \tag{15}$$

where $p_k(\boldsymbol{x}_i)$ is the probability of the vector $\boldsymbol{x}_i$ belonging to the k-th PDF in GMM; $\overline{P_{k,\alpha}^{\text{reg}}}$ features the expected regulation power of RACs in set $\mathcal{I}_\alpha$ in the k-th component of the GMM. These two parameters can be calculated as follows:

$$p_k(\boldsymbol{x}_i) = \frac{\pi_k \cdot \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{u}_k, D_k)}{f(\boldsymbol{x}_i)}, \quad \forall i \in \mathcal{I}_\beta, \forall k \in \mathcal{K}, \tag{16}$$

$$\overline{P_{k,\alpha}^{\text{reg}}} = \frac{\sum_{i=1}^{N_\alpha} P_{i,\alpha}^{\text{reg}} \cdot p_k(\boldsymbol{x}_i)}{\sum_{i=1}^{N_\alpha} p_k(\boldsymbol{x}_i)}, \quad \forall i \in \mathcal{I}_\alpha, \forall k \in \mathcal{K}. \tag{17}$$

To sum up, the regulation power of RACs in the set $\mathcal{I}_\beta$ can be evaluated by Eq. (15) according to the probability function in Eq. (16) and the expected regulation power in Eq. (17).

C. EM Algorithm for Optimizing the Parameters in GMM

The regulation power evaluation algorithm in Eqs. (15)-(17) is based on the GMM in Eqs. (11)-(14), while the parameters (K, π_k, $\boldsymbol{u}_k$, D_k) in Eqs. (11)-(14) are unknown. Because the sample set $\mathcal{X}$ of full-info RACs is an incomplete data set. We do not know the vector $\boldsymbol{x}_i$ belongs to which component k in GMM and cannot use maximum likelihood estimation. To address this issue, the EM algorithm is utilized here to obtain the parameters in GMM. Specifically, there are two steps: expectation-step (E-step) and maximization-step (M-step).

In the E-step, a latent variable vector $\boldsymbol{z}$ with K dimensions is introduced to indicate that the vector $\boldsymbol{x}_i$ is from the k-th component. Hence, the data set $\mathcal{X}$ can be extended to a complete data set $\mathcal{Y}$:

$$\mathcal{Y} = [\boldsymbol{x}_i, \boldsymbol{z}_i] = [\boldsymbol{x}_i, z_{i,1}, z_{i,2}, ..., z_{i,K}], \quad \forall i \in \mathcal{I}_\alpha, \tag{18}$$

where $z_{i,k}$ can only be equal to 0 or 1. The summation is equal to 1, i.e., $\sum_{k=1}^{K} z_{i,k} = 1$. For example, when the data set $\mathcal{X}$ has three components, the latent variable $\boldsymbol{z}_i = (1, 0, 0)$ indicates the vector $\boldsymbol{x}_i$ is generated by the first components in the GMM. Therefore, the prior probability of the vector $\boldsymbol{x}_i$ generated by the k-th component in the GMM can be expressed as $p_k(\boldsymbol{z}_k = 1|\boldsymbol{x}_i)$. For simplification, the $p_k(\boldsymbol{z}_k = 1|\boldsymbol{x}_i)$ is denoted by $\gamma_{i,k}$. Based on the Bayesian formula, the $\gamma_{i,k}$ can be calculated by:

$$\begin{aligned} \gamma_{i,k} &\triangleq p_k(\boldsymbol{z}_k = 1|\boldsymbol{x}_i) \\ &= \frac{\pi_k^j \cdot \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{u}_k^j, D_k^j)}{\sum_{k=1}^{K} \pi_k^j \cdot \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{u}_k^j, D_k^j)}, \\ &\forall i \in \mathcal{I}_\alpha, \forall j \in \mathcal{J}, \forall k \in \mathcal{K}, \end{aligned} \tag{19}$$

where j indicates the iteration times of the optimization for calculating π_k^j, $\boldsymbol{u}_k^j$ and D_k^j, i.e., $j = [0, 1, ..., J]$.

Based on the Eq. (19), the log-likelihood function of the GMM can be calculated by:

$$\mathcal{L} = \sum_{k=1}^{K} \left(\sum_{i=1}^{N_\alpha} \gamma_{i,k} \ln \pi_k^j + \sum_{i=1}^{N_\alpha} \gamma_{i,k} \ln \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{u}_k^j, D_k^j) \right), \quad \forall i \in \mathcal{I}_\alpha, \forall j \in \mathcal{J}, \forall k \in \mathcal{K}. \tag{20}$$

In the M-step, the iteration model in the next time is calculated by maximizing the log-likelihood function, which can be expressed as:

$$\left[\pi_k^{j+1}, \boldsymbol{u}_k^{j+1}, D_k^{j+1}\right] = \arg\max(\mathcal{L}), \forall j \in \mathcal{J}, \forall k \in \mathcal{K}. \tag{21}$$

Get derivative with respect to the variables (π_k^j, $\boldsymbol{u}_k^j$, D_k^j) and let the equations be equal to 0:

$$\frac{\partial \mathcal{L}}{\partial \pi_k^j} = 0, \quad \frac{\partial \mathcal{L}}{\partial \boldsymbol{u}_k^j} = 0, \quad \frac{\partial \mathcal{L}}{\partial D_k^j} = 0, \tag{22}$$

we can get the optimized parameters (π_k^{j+1}, $\boldsymbol{u}_k^{j+1}$, D_k^{j+1}) for the $(i+1)$-th iteration:

$$\pi_k^{j+1} = \frac{1}{N_\alpha} \sum\nolimits_{i=1}^{N_\alpha} \gamma_{i,k}, \forall i \in \mathcal{I}_\alpha, \forall j \in \mathcal{J}, \forall k \in \mathcal{K}, \tag{23}$$

$$\boldsymbol{u}_k^{j+1} = \frac{\sum_{i=1}^{N_\alpha} \gamma_{i,k} \boldsymbol{x}_i}{\sum_{i=1}^{N_\alpha} \gamma_{i,k}}, \quad \forall i \in \mathcal{I}_\alpha, \forall j \in \mathcal{J}, \forall k \in \mathcal{K}, \tag{24}$$

$$D_k^{j+1} = \frac{\sum_{i=1}^{N_\alpha} \gamma_{i,k} \|\boldsymbol{x}_i - \boldsymbol{u}_k^j\|^2}{\sum_{i=1}^{N_\alpha} \gamma_{i,k}}, \forall i \in \mathcal{I}_\alpha, \forall j \in \mathcal{J}, \forall k \in \mathcal{K}. \tag{25}$$

D. Determination of the Component Number in GMM

Based on the Eq. (18)-(25), the parameters of the GMM can be optimized. However, the total number of components in the

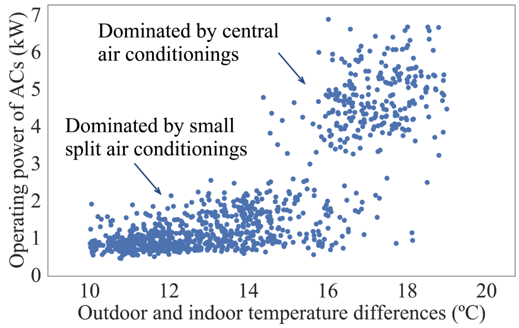

Fig. 2. The distribution of full-info RACs' initial operating power and temperature differences between the outdoor and indoor.

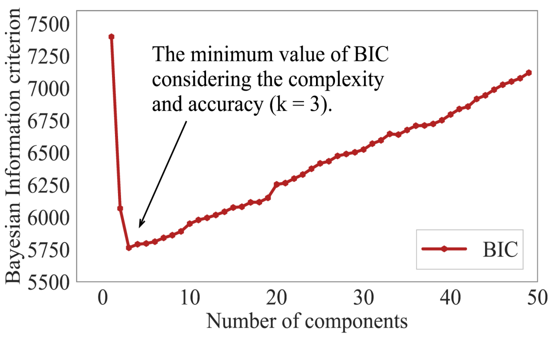

Fig. 3. Bayesian information criterion for determining the component number of GMM.

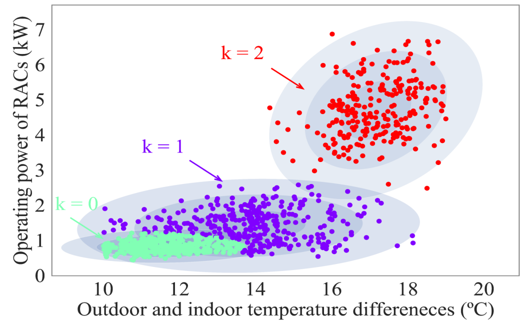

Fig. 4. The clustering results of full-info RACs based on the GMM method with 3 components.

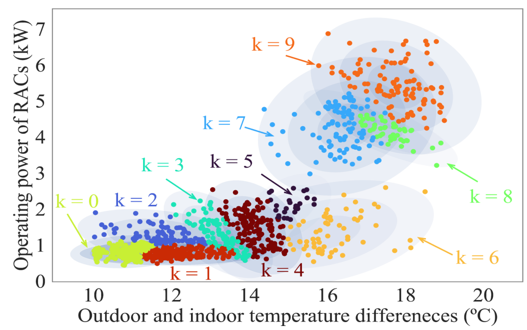

Fig. 5. The clustering results of full-info RACs based on the GMM method with 10 components.

GMM is not optimized. The BIC is utilized in this paper to select the best component number, as follows:

$$\mathrm{BIC} = K \ln N_\alpha - 2\hat{\mathcal{L}}. \tag{26}$$

The BIC model considers the complexity and accuracy of the GMM. The first item in the Eq. (26) features the complexity, which is lower with the decrease of the parameters K and N_α. The second item in the Eq. (26) features the accuracy, which is higher with the increase of the maximum value of the log-likelihood function $\hat{\mathcal{L}}$. Therefore, the GMM is considered to be better with a smaller BIC value to decrease the complexity and increase the accuracy.

IV. Case Study

A. Test System

The test system is based on a realistic demand response demonstration project in China. The total number of residential buildings and corresponding RACs N is 100,000. The buildings' thermal parameters are based on the design standard for residential buildings (JGJ134-2010). The EER of RACs η_i distributes among 2.6∼3.6. The set temperature of each RAC θ_i^{set} distributes among 18∼27°C according to heterogeneous consumers' requirements. The maximum deviation of the indoor temperature θ_i^{dev} is 2°C from the set values. The sample parameter $\alpha\%$ is set as 1%. At the beginning of the regulation, the indoor temperature is generally near the set value in most buildings, i.e., $\theta_i^{\mathrm{I}} \approx \theta_i^{\mathrm{set}}$. To reduce the number of dimensions and highlight the important factors on the regulation capacity, without loss of generality, we can simplify the vector of RACs' parameters $\boldsymbol{x}_i$ in Eq. (9) into a two-dimensional vector $[P_i^{\mathrm{I}}, \theta_i^{\mathrm{I}}]$. Besides, considering the outdoor temperature has significant influence on the regulation capacity of RACs as shown in Eq. (1), here we pay attention to the differences between the outdoor and indoor temperature, i.e., $[P_i^{\mathrm{I}}, \theta_o^{\mathrm{reg}} - \theta_i^{\mathrm{I}}]$.

It is assumed that the regulation beginning time is at 12:30 am, and the regulation duration period T_D is 10min. The outdoor temperature θ_o^{reg} is 37°C. The test system is implemented using Python with an Intel core i7-9700 CPU @3.00 GHz with 16.0GB RAM.

B. Results of Regulation Capacity Evaluation

Fig. 2 shows the distribution of the full-info RACs' initial states. It can be seen that the operating power distributes among 0∼7kW. The temperature differences between the outdoor and indoor distribute among 10∼19°C, i.e., the RACs' set temperatures are among 18∼27°C. On the whole, there are mainly two kinds of RACs: i) small split RACs with the operating power of 1∼2.5kW for cooling some single rooms; ii) central RACs with the operating power of 3.5∼7kW for regulating the whole house temperature.

In order to select the best component number in the GMM, the BIC is calculated based on Eq. (26). As shown in Fig. 3, the BIC can get the minimum value when the component number K is set as 3. That is to say, $K = 3$ can get the best balance between the accuracy and complexity of the GMM. When the component number K is set larger than 3, the accuracy of the GMM can get improved, while the complexity also gets increased. It may also impact the generalization ability and the computational efficiency, especially when there are large-scale number of RACs. To illustrate this problem, the clustering results of RACs based on GMMs with different component numbers are shown in Fig. 4, Fig. 5 and Table I.

TABLE I
RESULTS OF REGULATION CAPACITY EVALUATION BASED ON GMM

K	Regulation capacity/MW	Evaluation accuracy/%	Computation time/s	Relative time/%
3	31.20	98.48%	24.50	100.00%
5	31.19	98.45%	27.63	112.78%
10	31.29	98.76%	32.04	130.77%
15	31.10	98.17%	37.44	152.82%
20	31.02	97.91%	43.03	175.63%

As shown in Fig. 4, there are three components. RACs in component $k = 0$ represent some small split RACs with the operating power of 1∼1.5 kW. This kind of RACs may be mainly used in small bedrooms with a relatively higher set temperature (23∼27°C). RACs in component $k = 1$ represent some small split RACs with the operating power of 1∼3 kW. This kind of RACs are similar with that in the component $k = 0$ while with wider set temperature ranges (20∼26°C), which may be mainly used in living rooms or some bedrooms with lower RACs' set temperature. RACs in component $k = 2$ have larger operating power (3.5∼7kW), which are mainly the central air conditioning systems. Compared with split RACs in the previous two components, the central RACs in component $k = 2$ have relatively lower set temperature and consume more energy. This is because the air tightness and thermal insulation of large spaces are generally lower than the values in small single rooms.

The above three components in Fig. 4 have different expected regulation capacities: $\overline{P_{0,\alpha}^{\text{reg}}} = 0.264$ kW, $\overline{P_{1,\alpha}^{\text{reg}}} = 0.179$ kW, and $\overline{P_{2,\alpha}^{\text{reg}}} = 0.590$ kW. The $\overline{P_{1,\alpha}^{\text{reg}}}$ is less than $\overline{P_{0,\alpha}^{\text{reg}}}$, because the set temperature of most RACs in component $k = 1$ is lower than that in component $k = 0$. To maintain a lower indoor temperature, the RACs have to keep operating in relatively higher power. Besides, the central RACs have the largest expected regulation power because of their large operating power. Compared with 3∼6 times operating power than the RACs in component $k = 0$, the central RACs in component $k = 2$ only have about twice regulation capacities of the RACs in component $k = 0$. This is because the low air tightness and thermal insulation equate to decreasing the buildings' thermal resistance, which is unfavorable for RACs to provide regulation capacities.

Fig. 5 shows the clustering results of RACs when the component number is set as 10. It achieves a more precise classification of different RACs, while it may decrease the generalization ability and the computational efficiency. As shown in Table I, based on the optimized GMM with the component number $K = 3$, the evaluation accuracy of the regulation capacity can reach 98.48% utilizing 24.50s computation time. With the increase of component number, the computation time is increased significantly. the evaluation accuracy is almost unchanged or even lower because of insufficient generalization ability in $K = 20$ scenario. This illustrates the effectiveness of the proposed GMM and parameter optimization method for evaluating large-scale RACs' regulation capacity.

V. CONCLUSION

This paper proposes an regulation capacity evaluation method of large-scale heterogeneous RACs based on GMM. First, the regulation framework and calculation model of RACs are developed in the premise of guaranteeing consumers' comfort requirements. Then we propose an evaluation method on RACs' regulation capacity, which can carry out in the condition of insufficient data acquisition. Finally, numerical studies illustrate that the evaluation accuracy can reach 98.48% by measuring 1% RACs' parameters, which can promote the development of demand response in smart grid paradigm.

REFERENCES

[1] S. Xu, X. Chen, J. Xie, S. Rahman, J. Wang, H. Hui, and T. Chen, "Agent-based modeling and simulation for the electricity market with residential demand response," *CSEE Journal of Power and Energy Systems*, vol. 7, no. 2, pp. 368–380, 2020.

[2] P. Siano and D. Sarno, "Assessing the benefits of residential demand response in a real time distribution energy market," *Applied Energy*, vol. 161, pp. 533–551, 2016.

[3] Q. Shi, F. Li, G. Liu, D. Shi, Z. Yi, and Z. Wang, "Thermostatic load control for system frequency regulation considering daily demand profile and progressive recovery," *IEEE Transactions on Smart Grid*, vol. 10, no. 6, pp. 6259–6270, 2019.

[4] G. Chen, H. Zhang, H. Hui, N. Dai, and Y. Song, "Scheduling thermostatically controlled loads to provide regulation capacity based on a learning-based optimal power flow model," *IEEE Transactions on Sustainable Energy*, 2021.

[5] C. Huang, H. Zhang, Y. Song, L. Wang, T. Ahmad, and X. Luo, "Demand response for industrial micro-grid considering photovoltaic power uncertainty and battery operational cost," *IEEE Transactions on Smart Grid*, 2021.

[6] H. Hui, Y. Ding, Q. Shi, F. Li, Y. Song, and J. Yan, "5g network-based internet of things for demand response in smart grid: A survey on application potential," *Applied Energy*, vol. 257, p. 113972, 2020.

[7] X. Zhang, M. Pipattanasomporn, T. Chen, and S. Rahman, "An iot-based thermal model learning framework for smart buildings," *IEEE Internet of Things Journal*, vol. 7, no. 1, pp. 518–527, 2019.

[8] G. Chen, H. Zhang, H. Hui, and Y. Song, "Fast wasserstein-distance-based distributionally robust chance-constrained power dispatch for multi-zone hvac systems," *IEEE Transactions on Smart Grid*, 2021.

[9] M. Song, C. Gao, H. Yan, and J. Yang, "Thermal battery modeling of inverter air conditioning for demand response," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 5522–5534, 2017.

[10] H. Hui, Y. Ding, and M. Zheng, "Equivalent modeling of inverter air conditioners for providing frequency regulation service," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 2, pp. 1413–1423, 2018.

[11] T. Jiang, Z. Li, X. Jin, H. Chen, X. Li, and Y. Mu, "Flexible operation of active distribution network using integrated smart buildings with heating, ventilation and air-conditioning systems," *Applied Energy*, vol. 226, pp. 181–196, 2018.

[12] T. Chen, H. Pourbabak, Z. Liang, and W. Su, "An integrated evoucher mechanism for flexible loads in real-time retail electricity market," *IEEE Access*, vol. 5, pp. 2101–2110, 2017.

[13] Y. Ding, D. Xie, H. Hui, Y. Xu, and P. Siano, "Game-theoretic demand side management of thermostatically controlled loads for smoothing tie-line power of microgrids," *IEEE Transactions on Power Systems*, 2021.

[14] Y.-J. Kim, "Optimal price based demand response of hvac systems in multizone office buildings considering thermal preferences of individual occupants buildings," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 11, pp. 5060–5073, 2018.

[15] D. Xie, H. Hui, Y. Ding, and Z. Lin, "Operating reserve capacity evaluation of aggregated heterogeneous tcls with price signals," *Appl. Energy*, vol. 216, pp. 338–347, 2018.

[16] N. Lu, "An evaluation of the hvac load potential for providing load balancing service," *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1263–1270, 2012.

[17] M. Cai, M. Pipattanasomporn, and S. Rahman, "Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques," *Applied Energy*, vol. 236, pp. 1078–1088, 2019.

[18] F. Javed, N. Arshad, F. Wallin, I. Vassileva, and E. Dahlquist, "Forecasting for demand response in smart grids: An analysis on use of anthropologic and structural data and short term multiple loads forecasting," *Applied Energy*, vol. 96, pp. 150–160, 2012.