

# Optimizing Electric Vehicle Clustering through Improved MVABC-GMM Algorithm Based on Gaussian Mixture Model and Monte Carlo Prediction

Bingcheng Li, Guifang Guo\*

School of Information Engineering, Xizang Minzu University, Shaanxi, China

\*Corresponding author

**Abstract:** In the practical application of electric vehicle (EV) charging, significant differences in charging behaviors across different time periods lead to scheduling challenges, especially when dealing with a large number of vehicles, potentially causing the "curse of dimensionality" in scheduling. To address this issue, an improved MVABC-GMM algorithm based on the Gaussian Mixture Model (GMM) is proposed for clustering large-scale EVs, aiming to optimize EV scheduling strategies and enhance the utilization efficiency of grid resources. The charging data is fitted using a Gaussian distribution, and a Monte Carlo prediction model is built to simulate the charging behavior patterns of EVs. Clustering evaluation metrics are introduced to determine the optimal number of clusters. Comparisons show that the improved algorithm significantly enhances convergence speed and clustering performance. The clustering center at the position (17.79, 7.62) in the cross-shaped region visually reflects the charging patterns of EV users.

**Keywords:** Electric Vehicles; Clustering Algorithm; MVABC-GMM; Monte Carlo; Clustering Evaluation Metrics.

## 1. Introduction

The rapid development of electric vehicles (EVs) in recent years has provided a promising pathway for addressing the challenges of renewable energy integration. However, it has also revealed several issues. When the number of EVs reaches a certain threshold, a large number of vehicles may connect to the grid during specific periods, leading to increased grid load, reduced stability, line congestion, and load imbalances [1]. The clustered charging behavior of EVs inevitably imposes significant negative impacts on the power grid, including heightened instability, line congestion, and severe peak-valley differences [2]. Against this backdrop, adopting a more precise and efficient clustering method to group EVs based on similar charging periods can help optimize load-shifting strategies and improve the utilization efficiency of grid resources. Therefore, this paper combines probabilistic statistics with clustering algorithms to propose a more accurate and efficient method for simulating and clustering EV charging behaviors.

## 2. Literature Review

In data analysis and pattern recognition, the proper selection of clustering algorithms is critical, as it directly impacts the accuracy and effectiveness of the analysis results. Different clustering algorithms exhibit significant differences in theoretical foundations, suitability for data characteristics, and processing capabilities. An inappropriate choice may lead to distorted clustering results, loss of patterns, or an inability to effectively interpret data trends. Therefore, the correct selection of clustering algorithms is of paramount importance. Different clustering algorithms have been designed and studied by many scholars.

The Gaussian Mixture Model (GMM) is a probability distribution-based clustering algorithm that achieves "soft classification" by calculating the probability distribution of

samples belonging to each cluster [3]. Compared to partition-based clustering and hierarchical clustering, GMM demonstrates advantages in terms of time complexity and the local optima problem, enabling more efficient handling of large-scale datasets. GMM provides soft assignments, where each data point can belong to multiple clusters with certain probabilities. This feature makes GMM perform better in cases where data distributions overlap or exhibit uncertainty. Comparative analyses in the literature [4] suggest that GMM outperforms hierarchical clustering, K-means, K-medoids, and self-organizing maps (SOM) in clustering performance and demonstrates significant flexibility in accommodating cluster shapes.

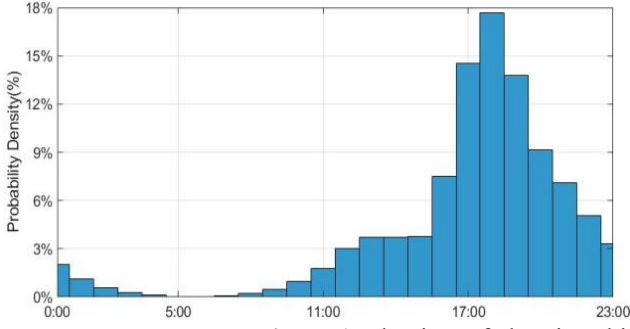
Although GMM is well-suited for clustering EV connection and disconnection time data, its computational time complexity increases significantly as the number of EVs or the volume of charging behavior data grows within a region. Moreover, the EM algorithm in GMM is highly sensitive to initial parameters (means, covariances, weights), and inappropriate initial values may result in convergence to local optima.

To address these shortcomings and achieve better clustering results for EV connection and disconnection time data, this paper incorporates the Minimum Variance (MV) theory and combines it with the Artificial Bee Colony (ABC) algorithm to improve the GMM algorithm. This enhancement aims to improve clustering performance and convergence efficiency, enabling more accurate clustering of EV connection and disconnection time data.

## 3. Experimental Analysis

### 3.1. Probability Density Function Fitting

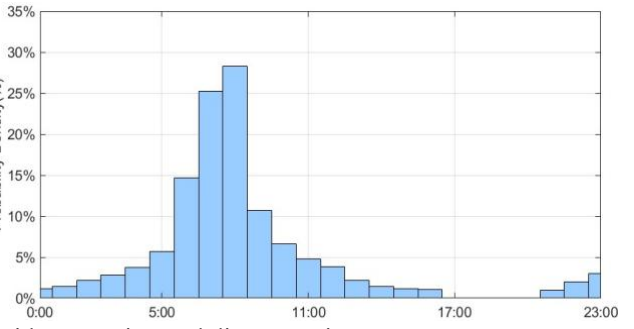
Fig. 1 show the statistics of the times when electric vehicles (EVs) return home and leave home, respectively, based on the "Big Data Report on the Travel of Small Pure Electric Passenger Cars in China" [5].



**Figure 1.** The time of electric vehicle grid connection and disconnection.

Typically, EV users connect their vehicles to the grid for charging after returning home. Therefore, in this study, the return home time of EV users is regarded as the connection time to the grid (entry time). Similarly, the departure time the following day is considered the disconnection time from the grid (exit time).

By inputting the statistical data Fig. 1 into MATLAB and processing it through programming with the fit function, the probability distribution functions for EV connection and disconnection times were derived, as shown below:



**Figure 2.** Probability density curve of electric vehicles

Based on the fitted curves, we obtained the parameters of the probability density functions. The parameter details are

$$f_{ST}(x) = a_1 \times \exp(-((x - b_1) / c_1)^2) + a_2 \times \exp(-((x - b_2) / c_2)^2) \quad (1)$$

$$f_{ST}(x) = a_3 \times \exp(-((x - b_3) / c_3)^2) + a_4 \times \exp(-((x - b_4) / c_4)^2) \quad (2)$$

Eq.(1) represents the probability density function (PDF) for the start time of EV charging, and Eq.(2) represents the PDF for the end time of EV charging. The probability density function curves derived from Eq.(1) and (2) are shown as Fig. 2:

listed in Table 1 below.

**Table 1.** Table of probability density function parameters

Parameter	Value	Parameter	Value	Parameter	Value
$a_1$	0.1732	$b_1$	17.79	$c_1$	1.358
$a_2$	0.0221	$b_2$	0.346	$c_2$	14.25
$a_3$	0.2908	$b_3$	7.62	$c_3$	0.97
$a_4$	0.0276	$b_4$	22.76	$c_4$	9.352

From Table 1, we can conclude that the primary peak of the connection probability is around 17:50, where the proportion of EVs charging is approximately 17.32%. The secondary peak occurs around 0:20, with a charging proportion of approximately 2.21%. For the disconnection probability, the primary peak is around 7:30, with a charging proportion of approximately 29.08%, while the secondary peak is around 22:50, with a charging proportion of approximately 2.76%. The probability density function curves for the connection and disconnection times of EVs align perfectly with the statistical patterns reported in the "Big Data Report on the Travel of Small Pure Electric Passenger Cars in China."

### 3.2. Simulation of Electric Vehicle Charging Behavior

The Monte Carlo algorithm is widely used in research fields such as power prediction for electric vehicles [6]. In the previous section, the probability density functions for EV charging connection and disconnection times, which exhibit general regularities, were fitted. Using these PDFs, the Monte Carlo simulation method can generate time data for EV connection and disconnection that satisfy the given probability distributions. Using this method, 10,000 sets of EV connection and disconnection time data were simulated, and the distribution results are shown in Fig. 3.

Fig. 3 includes the connection and disconnection time data of 10,000 EVs. The data points are most densely distributed in the time intervals of 16:00 to 21:00 for connection and 6:00 to 9:00 for disconnection, indicating that most vehicles are online during these periods. The sparse data points in the upper-left corner of the figure suggest that few vehicles

connect to the grid between 0:00 and 12:00 and disconnect between 16:00 and 24:00, which aligns with the statistical patterns reported in the "Big Data Report on the Travel of Small Pure Electric Passenger Cars in China". The data generated by this model adheres to the general rules of vehicle charging behavior.

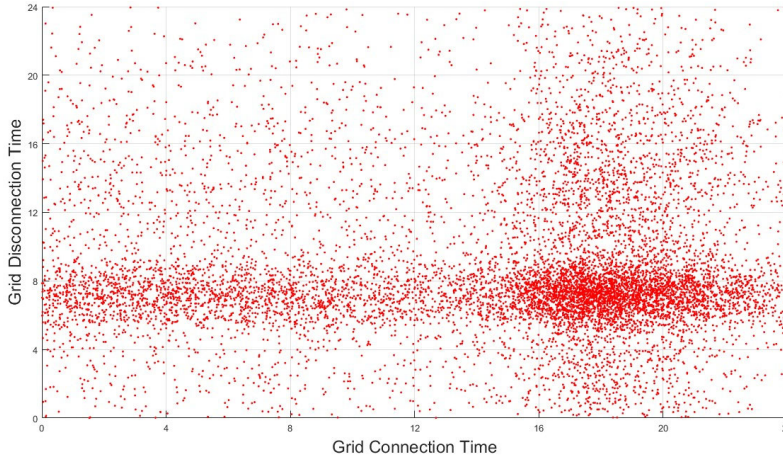


Figure 3. Monte Carlo simulation for 10,000 EVs

### 3.3. Clustering Process of Electric Vehicle Groups Based on the MVABC-GMM Algorithm

Based on the simulation using the Monte Carlo method, data reflecting the general patterns of electric vehicle (EV) charging behavior were obtained. This data consists of a two-dimensional dataset of EV connection and disconnection times. To perform a more precise clustering of EVs with

similar grid connection periods, the MVABC-GMM algorithm is used for clustering the EV charging behavior. This algorithm combines the minimum variance theory with the global search capability of the Artificial Bee Colony (ABC) algorithm, addressing the issues in the Expectation-Maximization (EM) algorithm of Gaussian Mixture Models (GMM), such as sensitivity to initial parameters and susceptibility to local optima. This ultimately enables accurate clustering of EV charging behavior.

The algorithm flow is shown in Fig. 4.

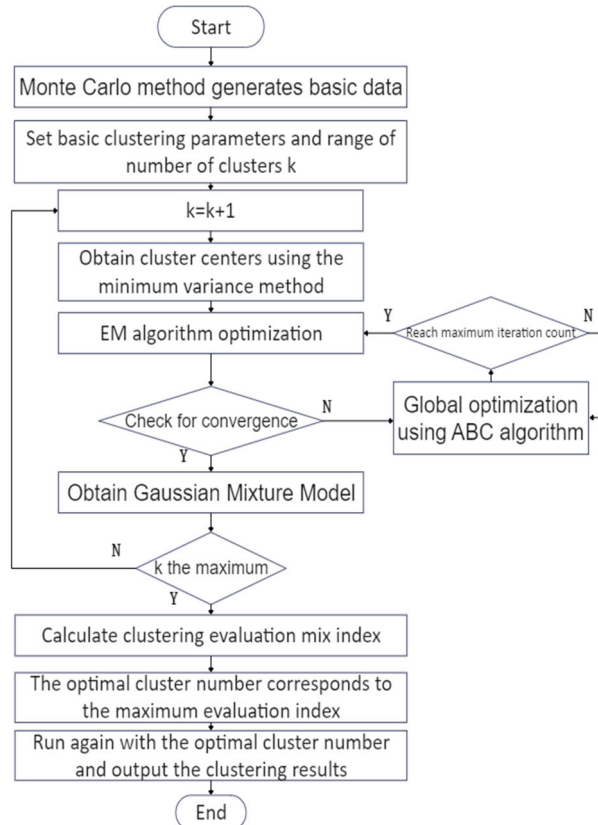


Figure 4. Flowchart of clustering based on the improved Gaussian Mixture Model

### 3.4. Experimental Analysis

The MVABC-GMM algorithm is used to perform clustering analysis on the 10,000 sets of electric vehicle (EV) charging connections and disconnection time data generated through the Monte Carlo method. The control parameters, including the number of scout bees, follower bees, and worker bees, are all set to 14. The bee colony algorithm performs 5 iterations, with the convergence condition set to a maximum of 100 iterations. The number of clusters is also defined. The clustering and iteration processes are shown in Fig. 5(a) and Fig. 6(a).

In article [7], a simple and efficient K-means algorithm is used to optimize the initial clustering centers for GMM clustering, which provides a good comparative basis. Based on the Kmeans-GMM algorithm, clustering analysis is also performed on the 10,000 sets of EV charging connection and disconnection time behavior data generated through the Monte Carlo method. The number of clusters is set, with a maximum of 100 iterations. The initial clustering centers are randomly selected from 6 data points in the dataset. The final clustering results and iteration processes are shown in Fig. 5(b) and Fig. 6(b).

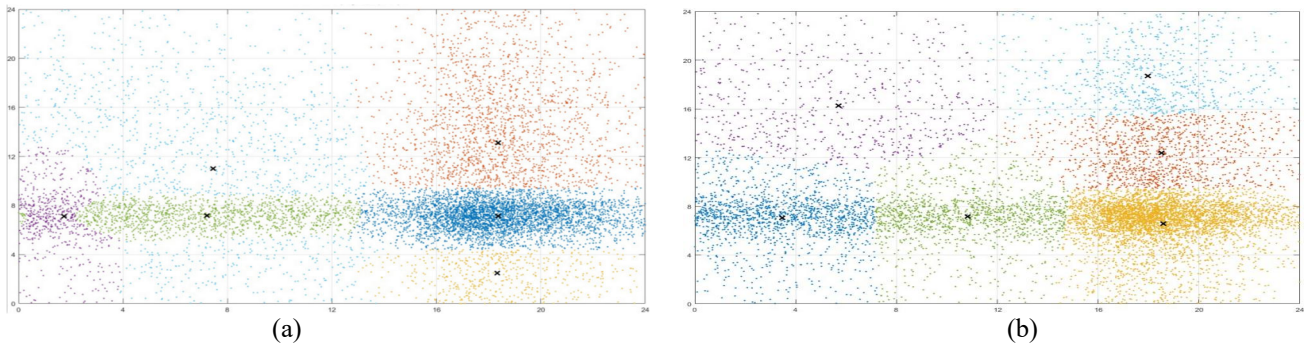


Figure 5. Clustering effect comparison

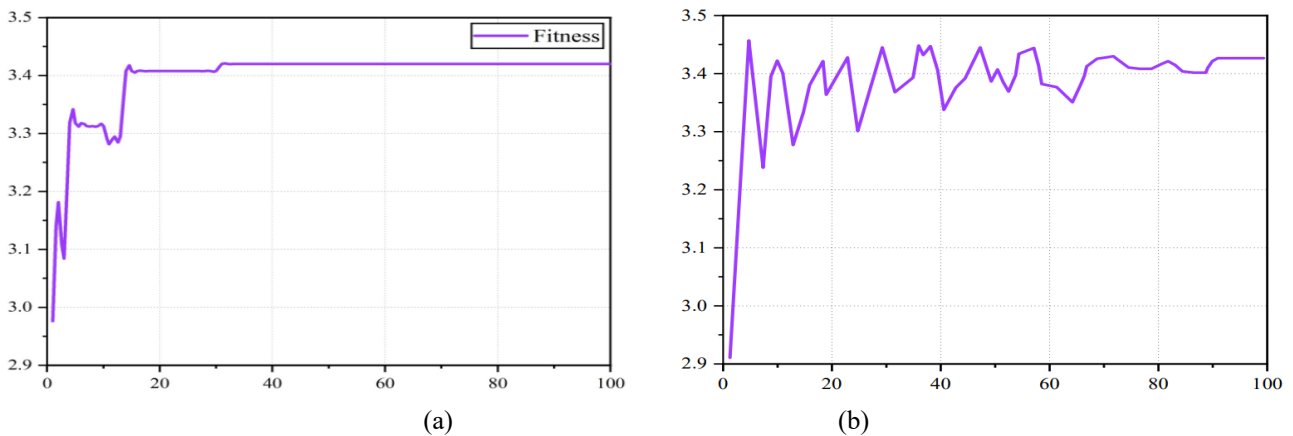


Figure 6. Iteration Process Comparison

#### Comparison of the Two Algorithms:

1. Convergence of Fitness Comparison: The clustering process of the Kmeans-GMM algorithm relies heavily on the randomly selected initial clustering centers, which causes significant oscillation in the fitness value during the initial phase, slower convergence, and a higher likelihood of getting stuck in local optima. As shown in Fig. 6(b), the Kmeans-GMM algorithm only stabilizes after 93 iterations, with the fitness value gradually increasing through fluctuations and ultimately reaching a suboptimal solution. In contrast, the MVABC-GMM algorithm selects initial clustering centers based on the minimum variance theory and incorporates the Artificial Bee Colony (ABC) algorithm for global search and local optimization, significantly improving both convergence speed and stability. As shown in Fig. 6(a), the MVABC-GMM algorithm escapes local optima after just 32 iterations and quickly converges to a stable state, demonstrating better global search capability and optimization efficiency.

2. Clustering Results Comparison: From the clustering

results, the Kmeans-GMM algorithm's cluster center distribution is significantly influenced by the initial values, leading to uneven distribution of certain categories of electric vehicles. As shown in Fig. 5(b). This causes some dense points and sparse points that are far apart to be grouped into the same cluster, failing to fully reflect the charging patterns and user habits of electric vehicles. In contrast, the MVABC-GMM algorithm, through optimization of the initial parameters and global search, not only ensures a more even distribution of the cluster centers in the data space but also effectively groups the points that are sparsely distributed due to the crossing region into the same category. The data points within different categories are more regularly distributed, allowing for a more accurate reflection of the charging behavior characteristics of electric vehicles.

As shown in Fig. 5(a).

## 4. Conclusion

This paper focuses on the analysis of electric vehicle (EV)

charging behavior patterns and regional clustering methods. By combining probability statistics and clustering algorithms, a more precise and efficient EV charging behavior simulation and clustering method is proposed. For clustering analysis of charging behavior data, the GMM (Gaussian Mixture Model) algorithm was selected after comparison, as it performed soft clustering on EV connection and disconnection time data. The results demonstrated GMM's advantages in handling overlapping data distributions and uncertainties. However, it was also found that GMM is sensitive to initial parameters, can get trapped in local optima, and has a relatively slow convergence rate. To address these issues, this chapter introduces the MVABC-GMM (Minimum Variance and Artificial Bee Colony optimized GMM) algorithm, combining the minimum variance theory and the Artificial Bee Colony (ABC) algorithm. The comparison with GMM optimized by the K-means algorithm significantly improved convergence speed and clustering performance.

## References

- [1] González L G, Siavichay E, Espinoza J L. Impact of EV fast charging stations on the power distribution network of latin American intermediate city[J]. *Renewable and Sustainable Energy Reviews*, 2019, 144 (107): 309-318.
- [2] Kempton W, Tomi J. Vehicle-to-grid power implementation: from stabilizing the grid to supporting large-scale renewable energy[J]. *Journal of Power Sources*, 2021, 144(1): 280-294.
- [3] LIU Jinpeng, YANG Hao, WU Lan, et al. Evaluation of residential demand response potential under multiple confidence scenarios based on Gaussian mixture model[J]. *Electric Power Engineering Technology*, 2023, 42 (2) : 20-28
- [4] Zhang Meixia, Li Li, Yang Xiu, et al. Load classification method based on Gaussian Mixture Model clustering and multidimensional scaling analysis[J]. *Power System Technology*, 2020, 44(11): 4283-4296.
- [5] Wang Tonghui, Hou Yi, Yan Ying. China New Energy Passenger Car Big Data Research Report [M]. Beijing: China Machine Press, 2021.
- [6] Wang Xin, Ding Yunfei, Lu Hongzhuang. Improved Kernel Extreme Learning Machine for Electric Vehicle Charging Load Prediction. *Journal of Shanghai University of Electric Power*, 2022, 25(1): 1-6.
- [7] Bu Hui. Research on User Information Demand Aggregation and Application in Online Q&A Communities Based on the Integration of GMM and K-means [D]. Qufu Normal University, 2022.