

KUCUKSARI, S. and ERDOGAN, N. 2022. Modeling and data analysis of electric vehicle fleet charging. In *Proceedings of 2022 IEEE (Institute of Electrical and Electronics Engineers)/AIAA (American Institute of Aeronautics and Astronautics) Transportation electrification conference and Electric aircraft technologies symposium (ITEC+EATS)*, 15-17 June 2022, Anaheim, USA. Piscataway: IEEE [online], pages 1139-1143. Available from: <https://doi.org/10.1109/itec53557.2022.9814047>

Modeling and data analysis of electric vehicle fleet charging.

KUCUKSARI, S. and ERDOGAN, N.

2022

© 2022 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Modeling and Data Analysis of Electric Vehicle Fleet Charging

Sadik Kucuksari

*Dept. of Applied Eng. and Technical Management
University of Northern Iowa
Cedar Falls, IA, USA
sadik.kucuksari@uni.edu*

Nuh Erdogan

*School of Engineering
Robert Gordon University
Aberdeen AB10 7GJ, UK
n.erdogan@rgu.ac.uk*

Abstract—In the transition to electric fleets around the world, electricity demand from electric vehicle (EV) fleets is expected to become significant in the future. Since fleet cars can display different charging characteristics than individual EVs, analyzing the charging behavior patterns of fleet cars is essential. To do so, this study first examines real EV fleet data from 724 charging events using data analytics methods. Based on this analysis, a charging behavior model is then developed to predict the realistic charging demand of an EV fleet with any number of EVs. In order to overcome the limitations of traditional probability density functions, this study utilizes Gaussian Mixture Models and Kernel distribution in developing charging behaviour models, i.e., charging start and end times, and total charging energy. The models' behaviours are then compared in terms of goodness-of-fit (GoF) to determine the best match for the original data, in which normalised root mean squared error serving as the fitness criteria.

Index Terms—Data analytics, electrified fleets, fleet Charging, Gaussian mixture model, kernel distribution, plug-in electric vehicles, probability density functions.

I. INTRODUCTION

Fleet electrification is expanding around the world in respond to global zero-emissions mandates. In addition to government commitments to convert public transport fleets to electrify, major companies have declared the transition to electric mobility by shifting their conventional fleets to electric vehicles (EVs) [1]. Due to this transition, the charging demand from EV fleets is expected to drive increments in peak power generation and transmission capacity [2]. The increased demand associated with the electrification of the transport sector can present challenges and opportunities through flexible electric vehicle grid integration schemes [3]. Therefore, analysis of charging behavior patterns of EV fleets and hence their realistic representations are needed to better analyze future energy systems. As a result, a charging behavior model may assist fleet owners in planning optimal sizing of the charging infrastructure required as well as optimal management of the charging demand with efficient use of the grid assets.

Traditional probability density functions (PDFs) were initially used to model charging behavior of EVs [4]. Each mobility metric at workplaces can be represented by the PDF of a normal distribution. As more EVs penetrate on the roads, real charging data from EV field tests have been used to model EV charging behaviors [5], [6]. However, the traditional PDFs

are limited to representing the charging profiles comprehensively. Therefore, Gaussian Mixture Models (GMMs) have been proposed [7]. A GMM is a weighted sum of m different Gaussian distributions that allows fitting various PDFs. As such, based on a histogram of collected data, realistic EV charging profiles are produced. In [8], the GMM is used to calculate the charging probability of EVs thereby generating EV charging profiles. Several studies have applied GMMs to model EV behavior for better planning and operation of future electricity networks. In [9], each charging metric, such as start charging time, initial and final state-of-charge (SOC), is represented by a GMM. The populated EV profiles by the GMMs are then validated against real data from an EV trial test. It is shown that the GMM model avoids under or overestimations, thereby producing more realistic EV profiles as compared to traditional PDFs, which are based on travel surveys. In [10], GMM is applied to predict EV user behavior at workplaces based on learned distribution from actual EV data. As a result, the GMM has shown to be effective to model EV behaviors realistically.

The models in the literature are based on collected data from individual cars. However, the charging behavior of EV fleet cars can differ from that of individual EVs. While workplace electric vehicle supply equipment (EVSEs) can accommodate a portion of the charging demand, public EVSEs are required to meet the demand for increased travel range [11]. It is therefore essential to develop a mobility model for an EV fleet as opposed to EVs at residential and workplace charging that typically rely on a predictable mobility pattern. To be able to truly assess the evolution and implications of the EV fleet, there is a need to analyze their charging behaviors. As such, based on this analysis, a charging behavior model can be developed to populate realistic EV profiles for a fleet with any number of EVs. The motivation in this study is to use data analytics methods in order to understand the charging patterns of an EV fleet. To do so, real EV fleet data from Leeds City Council [12] is analyzed in detail. GMMs and Kernel distribution functions are used to model three charging characteristics, i.e., charging start and end times, and total charging energy. The behaviors of models are then compared in terms of goodness-of-fit (GoF) to find the best fit for the original data, in which normalized root mean squared error is

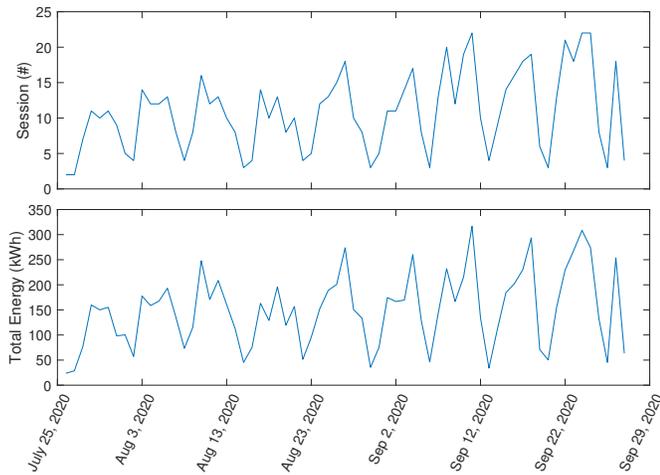


Fig. 1: Session numbers and total energy of charging events

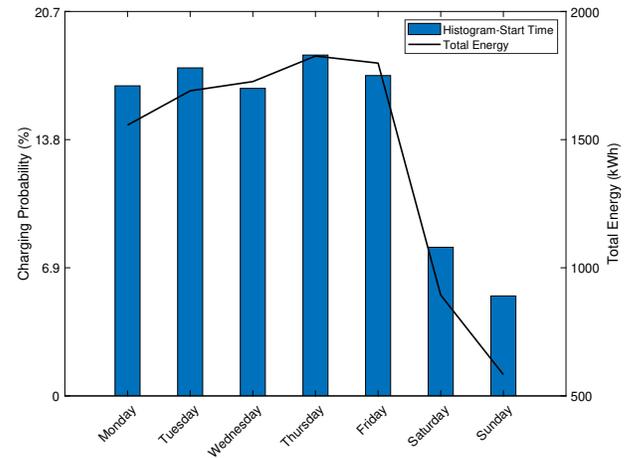
used as the fitness criterion.

II. DESCRIPTION OF EV FLEET CHARGING DATA

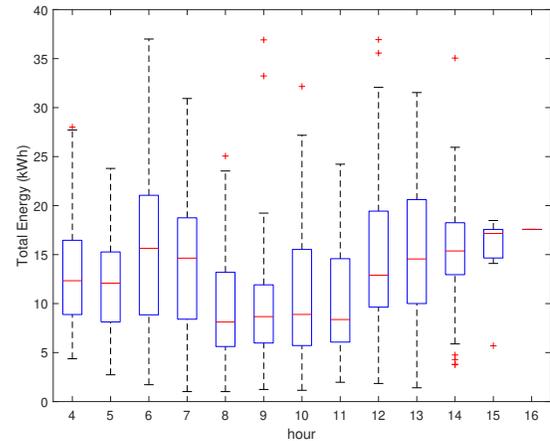
This study utilizes EVSE data set [12] used only by Leeds Council EV fleet. The fleet consists of 339 EVs whose configurations are as follows: (i) 3 different Renault Kangoo models with 33 kWh battery capacity and 7.4 kW on-board charger ratings, (ii) 2 different Nissan models with 24/40 kWh battery capacities and 3.3/6.6 kW on-board charger ratings. The workplace charging station includes, 120 chargers which are single phase, L2/Mode-3 type EVSEs at 7.36 kW. As shown in Fig. 1, the EVSE data set contains 724 charging events from July 25, 2020 to September 29, 2020. While the data includes charging start and end times, total charging energy, and plug-in time, EV identities are not known.

III. DATA ANALYSIS

The number of occurrences and the amount of total charging energy are both investigated in more depth. Fig. 1 indicates that the fleet's energy consumption and number of charging sessions vary by day. However, a similar pattern exists as weekly energy needs. The total charging energy is mainly higher during the week days and low in weekends. This is more readily seen in Fig. 2a. The weekly distribution of charging events and average energy are reported. The figure shows that a number of events during the weekdays form 86.6% of the total events with an average of 1719.8 kWh energy while being 13.4% and 738.3 kWh on weekends, respectively. The figure also indicates that when the energy need changes, so does the number of events. The data is then further examined to determine the hourly energy consumption requirements. The hourly energy need per day is shown in Fig. 2b. The charging events occur in typical work hours of between 4 am and 4 pm with an hourly average energy of 13.91 kWh. The charging events' hourly energy requirement varies by hour and mainly higher in morning hours. Each fleet may have a varied hourly requirement based on the work shifts and service provided.



(a)



(b)

Fig. 2: EV fleet charging energy demand distribution, a) weekly, b) daily.

Data on charging events for each EVSE unit was examined further. The average charging energy and number of events for each EVSE are calculated and the correlation is presented in Fig. 3. The marginal distributions given in Fig. 3a show that most of the EVSE units have fewer charging events while the average energy supplied per session is varying in a wide range. It is also seen that the EVSE units hosting a higher number of charging events usually supply higher average energy in each session. This may be due to the variations of the group of EVs in each departments. The correlation is also clustered and *k-means* classification is presented in Fig. 3b. Results depict that charging events can be classified mainly as two groups, each having 15.62 kWh and 88 sessions, and 11.84 kWh and 19.5 sessions as the center points that are shown in the figure with a cross symbol.

IV. STATISTICAL MODELING AND ANALYSIS

Mathematical representations of EV fleet charging behavior are first developed in Matlab using GMM and Kernel Distribution separately and then compared to show the best

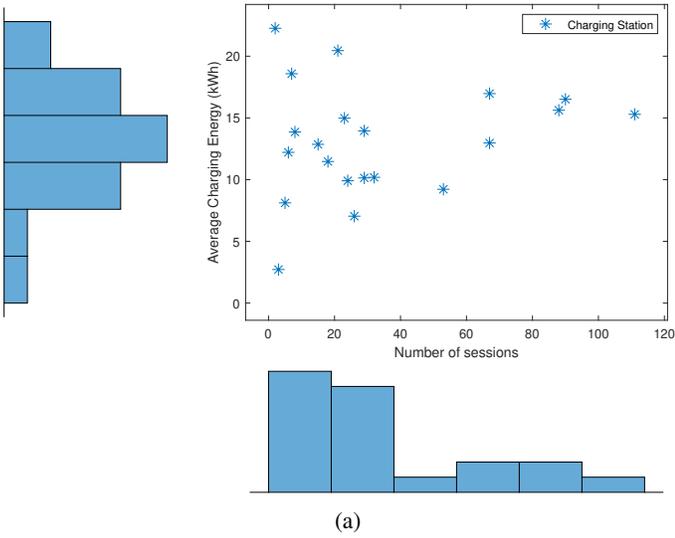


Fig. 3: EVSE units a) correlation and marginal distributions b) k-means classification.

fit. GMM is used to represent almost any distribution as a convex combination of Normal Distributions (NDs) with their means (μ) and variances (σ) [9]. The parameters are estimated from the original data using an iterative process known as Expectation-Maximization (EM), which maximizes the log-likelihood expectation depending on the number of GMM components that have been chosen as the target [13], [14]. A GMM, $f(x)$, can be formulated as the sum of weighted k number of NDs [9], [13], [14] as

$$f(x) = \sum_{i=1}^k \omega_i f_{ND(\mu, \sigma^2)}(x), \quad (1)$$

where, ω_i , μ_i , and σ_i^2 are the weight, mean, and variance of each Normal Distribution. Each GMM component's density

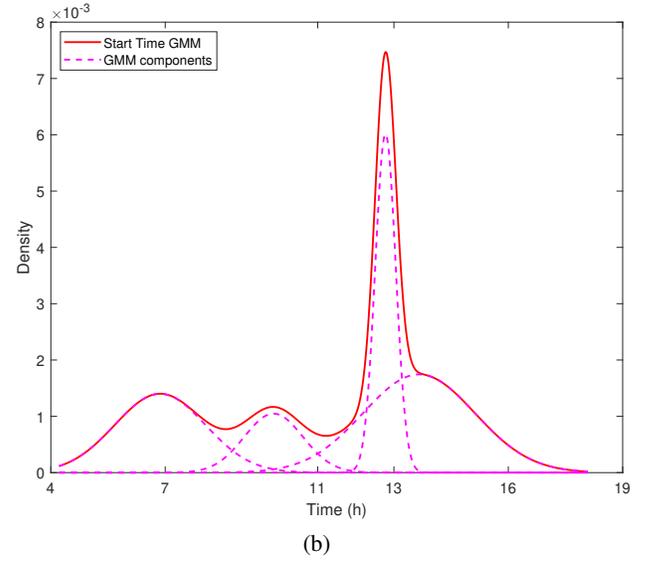
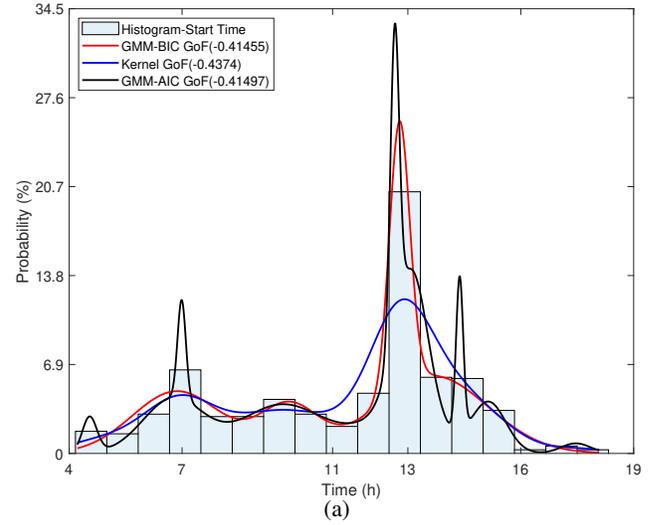


Fig. 4: Charging start time histogram and PDFs, a) GMM and Kernel PDF comparison, b) GMM with 4 components.

function is a normal distribution with the following:

$$f_{ND(\mu, \sigma^2)}(x) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}}. \quad (2)$$

The weight of each distribution, ω_i , ranges between 0 and 1 with $\sum \omega_i = 1$. In order not to overfit the GMM and find the best number of normal distribution components, Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) that are commonly used as model selection criteria [15] are compared for a wide range of possible component numbers (1~20). Among each model with different number of components, GMM with minimum BIC and AIC values are selected as the best fit models. In addition, a new set of data with the same number of charging events is generated using the GMM models and the GoF with the original data is compared. Normalized root mean squared error is used as the

fitness criterion where the closer to zero, the better fit for the original data. As the second method of modeling the charging events, Kernel Distribution is selected as a nonparametric representation of PDF [16]–[20]. Kernel density estimator can be used as the estimated PDF of a random variable and formulated for any real values of x as follows:

$$f(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right), \quad (3)$$

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad (4)$$

where, x_1, x_2, \dots, x_n are random samples, n is the sample size, $K(\cdot)$ is the kernel smoothing function, and h is the bandwidth. New set of data is generated with the Kernel Distribution model and the GoF is compared with that of GMM for performance comparison.

Fig. 4 and Fig. 5 show the GMM and Kernel Distributions, and their GoFs with the original data for charging start time, charging energy, and charging end time, respectively. The charging start time histogram, GMM with BIC criteria, GMM with AIC criteria, and Kernel models are provided and compared in Fig. 4a. The histogram shows that most of the charging events occur in the afternoon with the higher rate around 1 pm. The figure also depicts that each model has distinct fit. Although the GoFs of GMM with AIC and BIC are close to each other, the AIC model over fits with 11 GMM components while it is 4 for BIC. On the other hand, the Kernel distribution has slightly different fit with the worse GoF value when the Kernel bandwidth is found to be 44.60. GMM with BIC outperforms the other two models in terms of GoF. The number of components for the charging start time GMM contains four normal distributions, and each normal distributions containing the GMM is illustrated in Fig. 4b as an example for the analysis. It can be shown that the final GMM (red line) consists of the weighted sum of each components (dashed lines), as shown in the figure.

The charging energy and end time for the data set have been modeled in a similar manner. Fig. 5a compares the three model fits. The distribution of the charging energy shows that majority of the charging energy ranges from 8 to 13 kWh. There is less of a disparity between the fit lines, as may be seen in this figure. The worst fit is obtained by the Kernel distribution, while the greatest fit is achieved by the GMM BIC with two components. The difference in GMM AIC and BIC models is significantly less compared to that of charging start time models since the GMM AIC model has three components. The charging end time models are given in Fig. 5b. The two GMM models have the same number of components as the four-component GMM model. This causes the two charts to overlap. Despite the fact that the lines overlap, there is a little variation in the GoF. While the two GMM models overlap, the Kernel model has the worse GoF among the three models.

The modeling results show that in terms of the GoF value, the GMM BIC achieves the closest the zero as compared to the Kernel Distribution and GMM with AIC. The GoF for

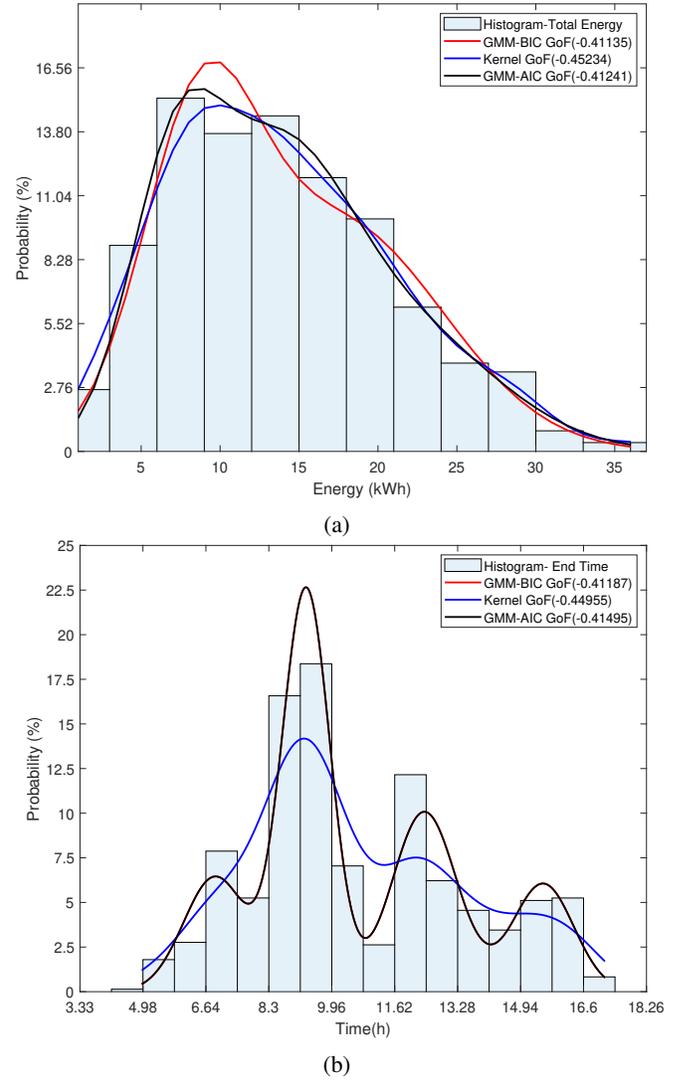


Fig. 5: Histograms and GMM and Kernel PDF comparison for, a) Total charging energy, b) charging end time.

BIC and AIC are close to each other, however, AIC has equal and higher ND components compared to that of GMM with BIC in all three behaviours. This confirms that the AIC model is less smoother than the BIC model. Among the three fits, the GMM with BIC model is found to be the best fit. While parametric modeling, GMM requires the user to determine the number of components an iterative process, Kernel does it directly. In order to get the least AIC or BIC values, it is necessary to manually run the GMM model over a large range of component numbers at a high computational speed. The parameters (weights ω , standard deviation μ , and variance σ_i^2) and number of components for GMM BIC model calculated are reported in Table I.

V. CONCLUSION

In this study, real EV fleet charging data was examined in detail. Since normal distributions lack the ability to represent their charging behavior, two advanced PDFs were developed

TABLE I: GMM with BIC components for EV fleet charging parameters

Component	Start Time			End Time			Total Energy		
	ω_i	μ_i [min]	σ_i^2	ω_i	μ_i [min]	σ_i^2	ω_i	μ_i [kWh]	σ_i^2
$i=1$	0.1214	590.57	2138.6	0.1473	935.74	2360.4	0.5938	17.52	44.03
$i=2$	0.2441	766.70	262.52	0.4305	559.26	1451.4	0.4062	8.64	12.84
$i=3$	0.3846	819.37	7738.3	0.1614	414.76	2497.9			
$i=4$	0.2499	411.97	5072.7	0.2608	747.11	2665.8			

based on GMM and Kernel distributions. A comparative analysis was performed to find the GoF whose criterion is normalized root mean squared error that represents the best charging behavior. The data analysis found that the charging behavior of EV fleet displays different characteristics as compared to residential and typical workplace charging. The charging start time expands throughout the day. Furthermore, EVSE occupancy times per EV is longer. However, the charging energy consumption follows a pattern that is close to normal distributions. It has been observed that in terms of GoF, the performance of the models depends on the distribution characteristics of data. Based on the EV fleet charging pattern considered, the GMM has shown superiority over the Kernel distribution model. As the normal distribution fails to model EV fleet charging behaviors, the analysis confirms that advanced PDFs such as GMM or Kernel distribution can be a good alternative to better represent the data. Fleet charging events, on the other hand, have been observed to occur mostly during the weekdays, with fleet EVs beginning charging operations as soon as they plug in. In this respect, smart charging can be utilized to evenly distribute the charging times. As a result, the grid assets can be efficiently used.

REFERENCES

- [1] N. Erdogan, S. Kucuksari, and U. Cali, "Co-simulation of optimal EVSE and techno-economic system design models for electrified fleets," *IEEE Access*, vol. 10, pp. 18 988–18 997, 2022.
- [2] T. Simolin, K. Rauma, R. Viri, J. Mäkinen, A. Rautiainen, and P. Järventausta, "Charging powers of the electric vehicle fleet: Evolution and implications at commercial charging sites," *Applied Energy*, vol. 303, p. 117651, 2021.
- [3] M. C. Kisacikoglu, F. Erden, and N. Erdogan, "Distributed control of PEV charging based on energy demand forecast," *IEEE Trans. Ind. Informat.*, vol. 14, no. 1, pp. 332–341, Jan 2018.
- [4] N. Erdogan, D. Pamucar, S. Kucuksari, and M. Deveci, "An integrated multi-objective optimization and multi-criteria decision-making model for optimal planning of workplace charging stations," *Applied Energy*, vol. 304, p. 117866, 2021.
- [5] J. Schäuble, T. Kaschub, A. Ensslen, P. Jochem, and W. Fichtner, "Generating electric vehicle load profiles from empirical data of three ev fleets in southwest germany," *Journal of Cleaner Production*, vol. 150, pp. 253–266, 2017.
- [6] T. Simolin, K. Rauma, R. Viri, J. Mäkinen, A. Rautiainen, and P. Järventausta, "Charging powers of the electric vehicle fleet: Evolution and implications at commercial charging sites," *Applied Energy*, vol. 303, p. 117651, 2021.
- [7] R. Singh, B. C. Pal, and R. A. Jabr, "Statistical representation of distribution system loads using gaussian mixture model," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 29–37, 2010.
- [8] M. Gödde, T. Findeisen, T. Sowa, and P. H. Nguyen, "Modelling the charging probability of electric vehicles as a gaussian mixture model for a convolution based power flow analysis," in *2015 IEEE Eindhoven PowerTech*. IEEE, 2015, pp. 1–6.
- [9] J. Quirós-Tortós, A. Navarro-Espinosa, L. F. Ochoa, and T. Butler, "Statistical representation of EV charging: Real data analysis and applications," in *2018 Power Systems Computation Conference (PSCC)*. IEEE, 2018, pp. 1–7.
- [10] Z. J. Lee, T. Li, and S. H. Low, "Acn-data: Analysis and applications of an open ev charging dataset," in *Proceedings of the Tenth ACM International Conference on Future Energy Systems*, 2019, pp. 139–149.
- [11] S. Kucuksari and N. Erdogan, "Ev specific time-of-use rates analysis for workplace charging," in *2021 IEEE Transportation Electrification Conference Expo (ITEC)*, 2021, pp. 783–788.
- [12] L. C. Council. (September 29, 2020) Ev fleet chargepoint use. [Online]. Available: <https://data.gov.uk/dataset/2279b730-bf4e-40c4-b2de-c82d43ae16d2/ev-fleet-chargepoint-use>
- [13] D. Reynolds, *Gaussian Mixture Models*. Boston, MA: Springer US, 2009, pp. 659–663.
- [14] R. Singh, B. C. Pal, and R. A. Jabr, "Statistical representation of distribution system loads using gaussian mixture model," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 29–37, 2010.
- [15] F. J. Fabozzi, S. M. Focardi, S. T. Rachev, and B. G. Arshanapalli, *The basics of financial econometrics: Tools, concepts, and asset management applications*. John Wiley & Sons, 2014.
- [16] N. G. Paterakis and M. Gibescu, "A methodology to generate power profiles of electric vehicle parking lots under different operational strategies," *Applied Energy*, vol. 173, pp. 111–123, Jul. 2016.
- [17] S. Schaeck, T. Karspeck, C. Ott, M. Weckler, and A. Stoermer, "A field operational test on valve-regulated lead-acid absorbent-glass-mat batteries in micro-hybrid electric vehicles. Part I. Results based on kernel density estimation," *Journal of Power Sources*, vol. 196, no. 5, pp. 2924–2932, Mar. 2011.
- [18] C. Ran, Y. Zhang, and Y. Yin, "Demand response to improve the shared electric vehicle planning: Managerial insights, sustainable benefits," *Applied Energy*, vol. 292, p. 116823, Jun. 2021.
- [19] M. Liang, W. Li, J. Yu, and L. Shi, "Kernel-based electric vehicle charging load modeling with improved latin hypercube sampling," in *2015 IEEE Power & Energy Society General Meeting*. Denver, CO, USA: IEEE, Jul. 2015, pp. 1–5.
- [20] Matlab. (R2021b) Kernel distribution. [Online]. Available: <https://www.mathworks.com/help/stats/kernel-distribution.html>