GOODNESS OF FIT OF PROBABILISTIC MODELS FOR ELECTRIC VEHICLE CHARGING BEHAVIOUR

Letetia M. Addison[1], Govinda A. Hosein[2] and Sanjay Bahadoorsingh[3*]

[1]Faculty of Science and Technology, The University of the West Indies, Trinidad
[2,3]Faculty of Engineering, The University of the West Indies, Trinidad
[1]Email:letetia.addison@sta.uwi.edu
[2]Email: govinda.hosein@gmail.com
[3]Email: sanjay.bahadoorsingh@sta.uwi.edu *(Corresponding author)

Abstract*:* Electric vehicles (EVs) have a number of environmental benefits in an era where fossil fuels have dominated. As such, the upgrade of electricity distribution grids to suit the needs of the modern world where the use of EVs can be accommodated is essential. Management of EV penetration is necessary, since uncoordinated charging can produce load imbalances and sharp variations in current, voltages and power. In order to assess the needs of such a system, estimates of random variables reflecting charging behaviour are necessary, particularly in cases where real data is insufficient. An attempt is made to assess some probabilistic models based on weekday load curves derived from the charging process. Level 1 EV charging profiles for uncoordinated charging schemes over one year for a data set consisting of 348 vehicles corresponding to 200 households are analysed and compared. Charging characteristics are reviewed and probability models are validated by goodness of fit statistics. Probability distribution functions (PDFs) which provide the best fit for these weekday load profiles are identified among the Johnson SB, Generalised Gamma and Dagum functions. This can provide an insight into estimation of PDFs based on EV charging behaviours, in order to build and assess models associated with transportation mobility data in other regions.

Keywords: *Electric Vehicles, Probability distribution function, Goodness of fit test, Uncoordinated charging*

1. Introduction

Globally, the impact of fossil fuel consumption on climate change has been an important topic of discussion within recent times. According to the Intergovernmental Panel on Climate Change (IPCC) [1], the transportation sector accounted for 14% of the global greenhouse emissions in 2014. In 2018, the Energy Information Administration (EIA) reported that 28% of the United States energy consumption was used for transportation, with 92% of that being supplied by petroleum [2]. As a result, the use of electric vehicles (EVs) in communities is seen as a viable alternative.

There are different types of EVs, classified by the amount of electricity used as the energy source. Battery Electric Vehicles (BEVs or most commonly EVs) and Plug-in Hybrid Electric Vehicles (PHEVs) are two main classifications of plug-in electric vehicles (PEVs). In general, PEVs have been identified as the main future alternative to conventional automobiles powered by Internal Combustion Engines (ICEs) [3]. This

is due to the fact that they are environmentally friendly and economically efficient, since they facilitate the use of Renewable Energy (RE).

Despite these advantages, electrical power systems constitute a significant area that could be directly impacted by the integration of such vehicles on the grid [4]. Power drawn by these vehicles is variable and depends on certain characteristics associated with driving behaviour. In order to analyse the impact of the EVs into the grid, models of charging curves are integral to the process. It is important to consider the aggregate effects of charging electric vehicles on electric power system infrastructure [5]. As such, integration of EVs within the smart grid framework requires careful consideration.

Use of probabilistic models of charging station load can assist greatly in this process and subsequently impact the longevity of the grids and overall success of the system. According to [6], factors such as driving patterns, charging characteristics (vehicle demand profiles), charging timing (the magnitude and duration of charging cycle) and vehicle penetration, impact the electric network. Further to this, [3] expressed that there exists a need to conduct a statistical evaluation among a wide range of available theoretical probability density functions (PDFs) in order to find the best model, to reflect the random characteristics of each driver behaviour variable.

Uncoordinated charging refers to the scenario where there is no control over when EV loads charge. This scenario significantly increases the peak demand and will require upgrades to the distribution grid. It can also lead to an increase in load imbalances, possible outages, as well as current and voltage variations. The research in [7] highlights how uncontrolled distribution of single-phase charging could be responsible for local voltage disturbances. These negative effects will increase if there are no Demand Side Management (DSM) schemes [8] to reduce or shift energy consumption from peak hours to leaner demand periods.

On the other hand, a properly designed and well-coordinated charging scheme can provide better flexibility and reliability to the entire electrical system [9]. To demonstrate the capabilities of a coordination scheme, real data and driver behaviours need to be available to perform simulations. However, the scarcity of available real data regarding EVs and charging stations has forced researchers to first develop probability distributions for a number of variables [9]. As such, [5] and [9] use real transaction data, with probabilistic and statistical ideas, in order to assess the impact on the grids.

According to [10], addition of EVs would affect the overall load pattern of distribution networks, leading to power quality concerns such as voltage imbalances, depending on EV charging pattern over a day. The research in [10] discusses the complications in attempting to provide a deterministic quantification of the number of EV charging events per day and the associated load on the grid. The complete mobility pattern of the EV driver is an important factor which is not always known. Hence, [11] expresses that there is a need to develop a probabilistic model of EV charging to estimate an expected load in the system, leading to a power index through which utilities can upgrade their infrastructure to support large penetration of EVs.

However, estimates of random variables related to driver behaviour (arrival, departure times, daily mileage etc.) to characterise the PEV charging process is a challenge. Currently, there is a lack of sufficient real data for this purpose. Therefore, samples from transportation mobility data are used to estimate a PDF. The research in [3] explains that the purpose is to not only preserve the characteristics of each variable but to generate simulated data for further comparison. Hence, the use of PDFs is a useful way to reflect driver behaviour.

In addition, [10] develops a probabilistic model of the charging pattern for EVs associated with residential load profiles. The probabilistic model provides the activity for the residential load profiles and EV charging

patterns over a period of twenty-four (24) hours. Other studies ([12], [13]) use a theoretical PDF to provide a fit for the sample data. This gives a snapshot of the intrinsic randomness of driver behaviour variables, which is then used to generate synthetic data from the fitted PDFs. Thus, [12] assumes Normal and Lognormal PDFs for the variables arrival time and daily travel mileage respectively. Alternatively, [13] assumes Chi-square and Power law PDF for the aforementioned variables.

While many papers have fit PDFs for individual driver behaviour variables, most have not considered weekday load profiles. This work attempts to assess which theoretical PDFs provide the best fit for weekday load curves based on aggregate residential and EV charging profiles. Following this overview of the current research, section 2 describes a brief methodology for such a scheme. Next, section 3 outlines suitable goodness of fit (GoF) statistics and PDFs based on each of the load curves by day of the week. Section 4 discusses the importance of such analysis for charging behaviour, followed by preliminary conclusions in section 5. Overall, this research serves as a benchmark for future models which can use these PDFS to assist in implementation of control strategies for different regions.

2. Methodology

This paper is a PDF fitting case study using in-home plug-in electric vehicle recharging profiles for 348 vehicles. These are associated with 200 households randomly selected among the ones available in the 2009 Residential Consumption Survey (RECS) [14] data for the Midwest region of the United States. This probability sample survey enables statistical selection of households to collect energy-related data. The publicly available data set accounts for user activity and appliance usage statistics from more than 12000 households across the United States. These are statistically chosen to represent over 110 million household units, where physical properties of each residence are included.

Power profiles which depict energy consumption patterns can fluctuate and due to this randomness the prediction of energy demand can be difficult. For these reasons, according to the work of Muratori ([15], [16]), residential demand profiles are variable since individual household behaviour is stochastic in nature. Each PEV is charged as soon as connected to the grid, until the battery is fully charged. Therefore, there is no coordination scheme applied to the data set. Simulation of the amount of electricity required to fully charge the battery depends on the previous trips and charging events, which are simulated using a personal energy consumption model constructed by Muratori [16].

The profiles proposed by Muratori et al. ([15], [16]) create realistic patterns of residential power consumption. These are validated using metered data, with a resolution of 10 minutes. Muratori [15] also simulates various scenarios considering different PEV market shares. In this data set, vehicles are assumed to be 60% BEVs with 200-mile range and 40% plug-in hybrid electric vehicles PHEVs with 40-mile all-electric range based on market trends. Both Level 1 charging (1920 W) and Level 2 charging (6600W) are assumed in the aforementioned dataset and the profiles represent total plug-in electric vehicle (PEV) charging demand.

In this paper, the Level I charging data set (PEV-L1) [15] is used, since it is expected most that most households would charge at this level. This data represents an uncoordinated charging scheme where no control strategy is applied to deal with daily demand. The electricity demand profiles for the 200 households (without EV charging) are mapped to the total EV demand within the corresponding households (since some residences have more than one EV). Then, the aggregate household load is computed using the sum of the base load and the total EV load per household. The equation for this computation is expressed as:

$$PW_{Total\ Load,i} = PW_{Resid,i} + PW_{PEV,i} \qquad (1)$$

where $PW_{Total\ Load,i}$ represents the total power demand for residential load, i, and $PW_{Resid,i}$ represents the power demand associated with the residential load, i and $PW_{PEV,i}$ is the corresponding total PEV-L1 load, for that household ($i = 1,2, ...,200$). The models in this paper ignore seasonal variations, considering only weekday and weekend profiles.

These load profiles are used to construct weekday demand curves by time of day. Then, daily load curves are fitted using polynomial curves for each day of the week. In each case, a non-linear equation is obtained. The form of this equation is:

$$f(x)_{fitted} = a_n x^n + \cdots + a_2 x^2 + a_1 x + a_0 \qquad (2)$$

where a_k represents the fitted model co-efficient of x^k, the variable 'time of day (hrs.)', for $k = 0,1, ... , n$. In this work, a 5th order polynomial is used ($k = 5$) for the best fit.

For each day of the week, GoF statistics are generated to describe the overall fit of the model to the original data based on differences between observed and predicted values. The coefficient of determination, R^2, provides a measure of how well the fit explains the variation in the data. RMSE provides a relative measure of the fit, that is, how concentrated the data is around the line or curve of best fit. Let y_i be the true response for the ith observation and $\hat{y}_i$ be the corresponding predicted response. The RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \qquad (3)$$

Lower values of the RMSE closer to 0 indicate a better fit.

Then, aggregate load curves categorised by weekday and weekend are examined in a probabilistic sense. Polynomial curve fitting is used to generate an estimated PDFs for each load curve. XLSTAT 2019, the add-in package for Microsoft Excel, which uses maximum likelihood estimation (MLE) to derive the parameters of the probability distribution, provides the best fitting theoretical PDF for each scenario. It should be noted that these distributions are the chosen from a pre-set database of PDFs, but there may be others which can provide a better fit.

It is necessary to conduct a GoF test to check whether the observed data follow the specified distribution. The Kolmogorov-Smirnov (K-S) test is a nonparametric test which compares the empirical (actual) cumulative distribution function (CDF), say $G(x)$ to the theoretical CDF, say $F(x)$. The K-S Statistic (D) computes the largest difference between the two functions for all sample values $x_i, i = 1,2, ... , n$. using the equation:

$$D = max[G(x) - F(x)] \qquad (4)$$

This statistic is used to determine if the hypothesis of interest "the observed data follow a specified distribution" is true when compared to a critical value. The significance level or p-value of this test is set to 0.05 or 5% and the original assertion is rejected if $p < 0.05$.

3. Results and Analysis

R Statistical Software [17] is used to match EVs to households to obtain the total load. Descriptive statistics, goodness of fit and probability plots are generated to reflect the load profiles by day of the week. These provide developers with tools to provide aggregate comparisons and to visualise the differences between datasets respectively. Figure 1 shows the residential load without EV charging, classified by time of day and day of the week. Ignoring seasonal variations, the general trends for the days of the week are similar. Subsequently, load curves are generated to include EV charging and the goodness of fit of these curves are analysed.
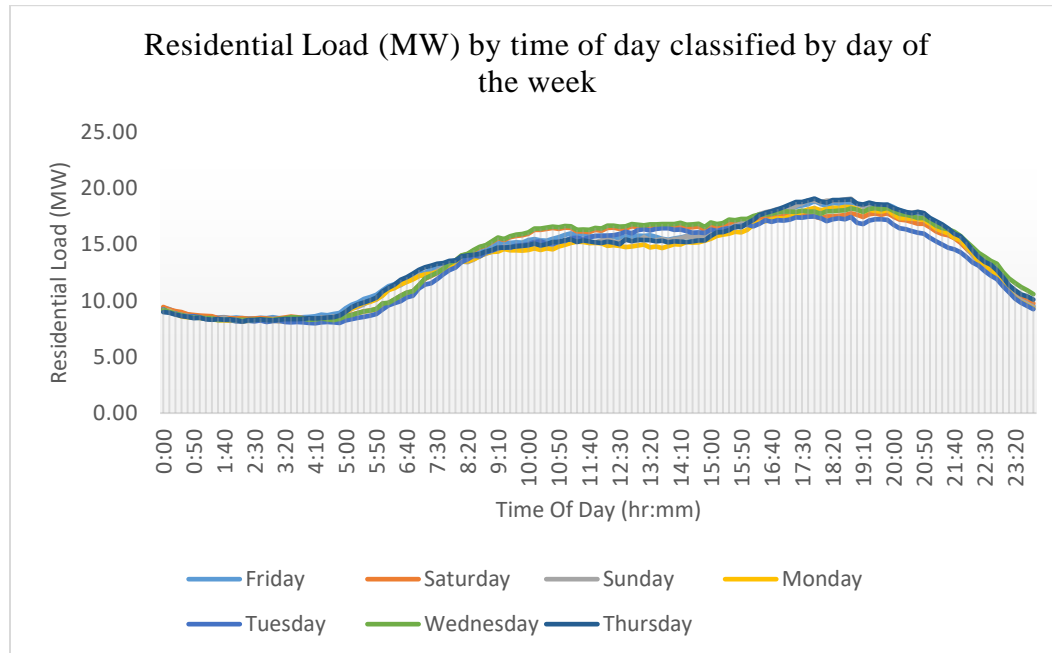

Figure 1: Residential Load Curves by time of day and day of the week (no EV charging)

3.1 Hourly Load Curves by Day of the Week

Load curves are generated for the original uncoordinated charging scheme by day of the week, in contrast to seasonal load variations, which are most commonly examined in the literature. In order to determine the charging behaviour on a weekly basis, total load curves based on the residential consumption and additional load from daily EV charging are constructed graphically.

These are shown in Fig. 2 where, upon initial inspection, the peak demand occurred mainly between the hours of 4 p.m. (16:00 hrs) and 10 p.m. (21:00 hrs) daily. Descriptive statistics are generated as shown in Table 2. The mean (average) and standard deviation (spread of data points about the mean) can later be used as estimates of shape and scale parameters in certain probability density functions (PDFs). Tuesdays experienced the highest average load (19.47 MW) while Thursdays experienced the lowest (18.35 MW). The standard deviations are similar for each day, ranging from 5.32 MW to 6.07 MW.

Each load curve is best fit to a polynomial to the 5th degree. The overall fit for each curve is quite good based on the R^2 values (all close to 1) while the RMSE value is lowest for Thursday (0.65), indicating this particular day had the best relative fit. Since the demand by day of the week is fairly similar, aggregate

weekday and weekend load curves are generated and validated. The mean weekday load is 19.04 MW, slightly less than that of the weekend (19.12 MW). These load curves are examined further in Section 3.2 to find suitable theoretical PDFs to represent them.
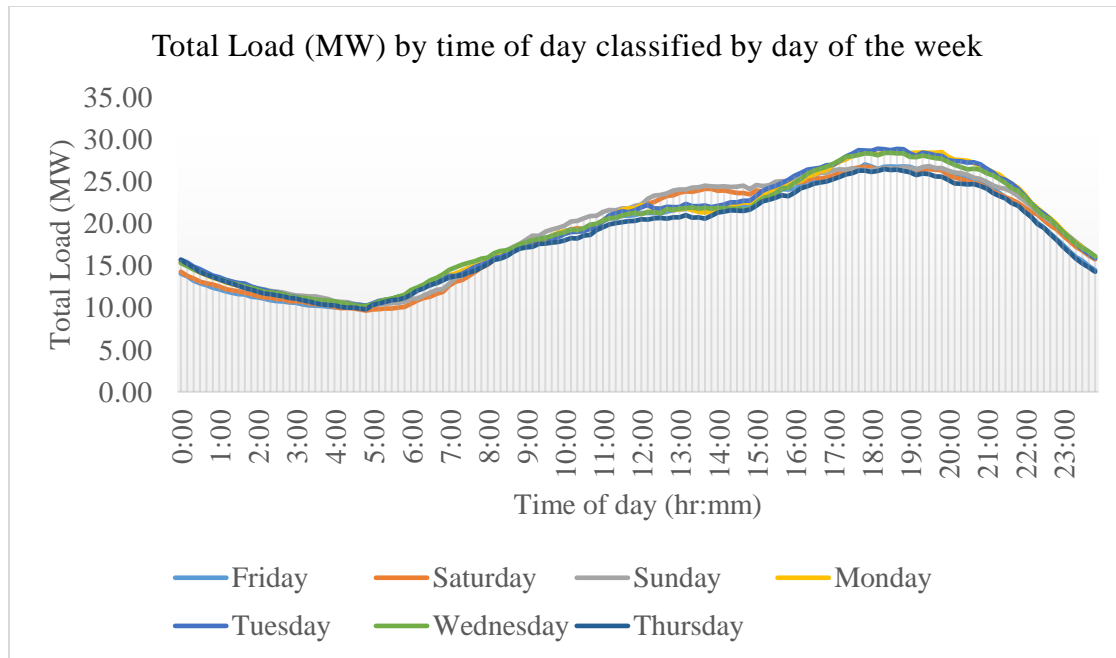

Figure 2: Total Load Curves by time of day and day of the week (EV charging included)

Table 2: Descriptive and Goodness of Fit Statistics for Total Load Curves (MW.) by day of the Week

	Mon	Tues	Wed	Thurs	Fri	Sat	Sun	Weekday	Weekend
Mean	19.38	**19.47**	19.33	**18.35**	18.59	18.82	19.4	19.04	19.12
Standard Deviation	6.03	6.07	5.84	5.32	5.67	5.94	5.83	5.76	5.87
Minimum	10.01	10.2	10.17	9.78	9.94	9.63	10.23	10.03	9.93
Maximum	28.5	28.86	28.37	26.44	27	26.7	26.8	27.75	26.62
R²	0.98	0.98	0.98	0.99	0.99	0.99	0.98	0.98	0.98
RMSE	0.83	0.76	0.86	0.65	0.69	0.72	0.75	0.74	0.74

In addition, Fig. 3 shows a stacked bar chart of the total load by day of the week. This particular artefact is useful for charging schemes that use Time of Use (TOU) tariffs to coordinate. Peak demand times are between 6 p.m. and 8 p.m. where the total weekly load is just under 1200 MW. The off-peak demand times occur between 4 a.m. and 6 a.m., where the total weekly load is at its minimum at 5 a.m. (just over 400 MW). These times are most crucial for coordination schemes to address. Examination of these attributes can assist in the determination of the most appropriate charging schemes for a particular region.
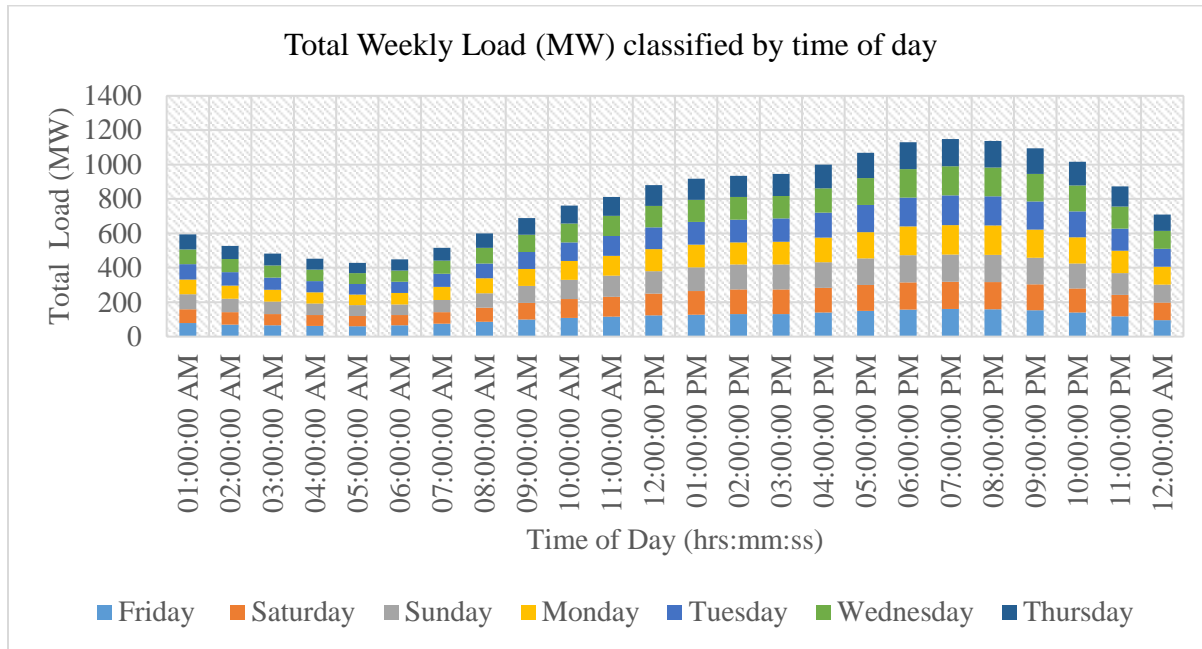
Figure 3: Stacked Bar Chart of Total Weekly Load with EV charging by time of day

3.2 Goodness of Fit to Polynomial Functions and Known PDFs

Non-linear polynomial equations are used to fit aggregate weekday and weekend load curves for this data set. These can be used to estimate the probability of daily demand at a particular time of day. The weekend and weekday scenarios are both displayed on Fig. 4. The PDFs of the load curves are best fit to polynomials to the 5th degree. The coefficients of the two equations are close, which indicates that the demand for weekday and weekend, even with inclusion of the use of EVs on the grid, is quite similar for this particular region. The differences in demand is reflective of their daily residential and EV charging behaviours.

Subsequently, standard PDFs are also used to determine the distributions with the best fit. EasyFit 5.6 Professional, an add-in package in Microsoft Excel, performs GoF tests evaluate the most appropriate distribution based on the smallest K-S value compared to the critical value, which is the best fit for the data. Figure 5 shows the best three PDFs for the weekday load curves. The histogram represents the spread of daily load over 24 hours. The PDF known as Johnson SB distribution is found to be the best fit (K-S = 0.0162, p > 0.05), followed by the Generalised Gamma (K-S =0.0277, p > 0.05) and then the Dagum distribution (K-S = 0.0282, p > 0.05), as shown in Table 3.

In Fig. 5, the values of the parameters are $\gamma = -0.441, \delta = 0.758, \lambda = 26.639$ and $\xi = -2.564$ for the Johnson SB distribution. The corresponding parameters for the Generalised Gamma are $k = 146.71, \alpha = 0.00917$ and $\beta = 23.824$, while those for the Dagum distribution are $k = 0.0068, \alpha = 198.62$ and $\beta = 23.766$. The shape parameters in the aforementioned distributions allow for flexibility which allows them to fit various real-life scenarios. This enables the realistic stochastic dynamics of EV charging profiles to be showcased.
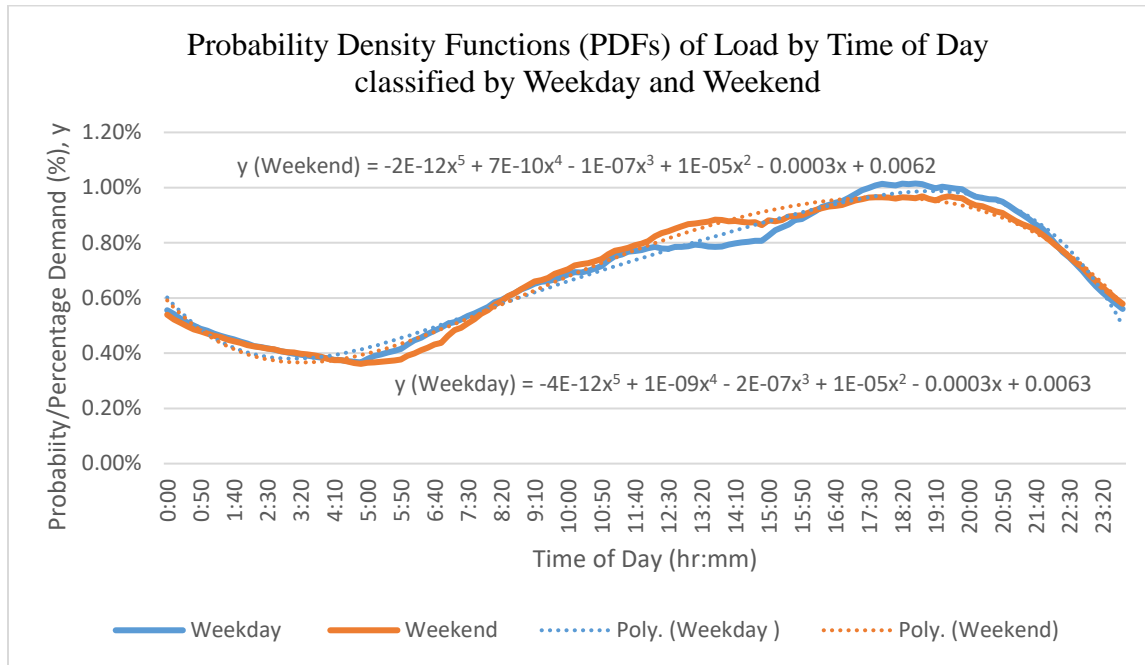
Figure 4: Probability Density Functions for power demand curves by weekday and weekend.

Table 3: Top three best fitting PDFs for weekday load curve

Distribution	PDF	K-S Statistic	p-value
1. Johnson SB	$f(x) = \dfrac{\delta}{\lambda\sqrt{2\pi}z(1-z)}\exp\left(-\dfrac{1}{2}\left(\gamma + \delta\,ln\left(\dfrac{z}{1-z}\right)\right)^2\right),$ $\xi < x < \xi + \lambda; \quad -\infty < \gamma, \xi < \infty; \quad \delta > 0, \lambda > 0$	0.0162	1.0
2. Generalised Gamma	$f(x) = \dfrac{kx^{k\alpha-1}}{\beta^{k\alpha}\Gamma(\alpha)}\exp\left(-\left(\dfrac{x}{\beta}\right)^k\right),$ $x > 0, k > 0, \alpha > 0, \beta > 0, \Gamma(\alpha)$ is the Gamma function	0.0277	0.99
3. Dagum	$f(x) = \dfrac{\alpha k \left(\dfrac{x}{\beta}\right)^{k\alpha-1}}{\beta\left(1 + \left(\dfrac{x}{\beta}\right)^\alpha\right)^{k+1}},$ $x > 0, k > 0, \alpha > 0, \beta > 0$	0.02821	0.99

These PDFs can be used to generate the uncoordinated charging power demand, $f(x)$, at a certain charging time, x. The advantage of these 'best' fitting PDFs is the additional shape parameters which provide a more realistic fit than popular PDFs such as the Normal and Weibull. This can be repeated for the weekend scenario also. In this way, random characteristics of the data can be preserved and EV charging load profiles can be easily estimated using these equations. Although known PDFs can be fitted as shown, polynomial curve fitting also provides a feasible technique to provide a close fit for such analysis. This flexibility is essential for further predictive analysis in terms of demand in different regions.
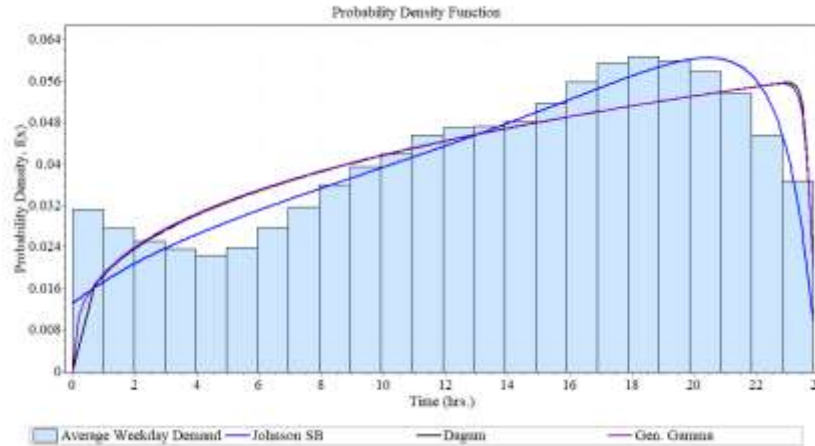
Figure 5: Probability Density Functions for the Weekday Load Curve

4. Discussion

Descriptive statistics and goodness of fit metrics are shown to be useful tools for compressing large datasets without requiring significant amount of time to process. In particular, maximum daily peak values are of significance to distribution grid protection schemes to predict system disturbances. The mean values can be used to demonstrate the percentage increase of daily demand compared to lower EV penetration scenarios, thereby enabling better scheduling.

However, there is room for further work to generate more accurate PDFs for corresponding load curves in order to subsequently predict EV demand, as well as feed into more complexed schemes. A limitation of the theoretical PDFs is the inability to accurately describe the weekday demand at particular hours in the day. This may affect the accuracy of performance indicators and estimates at certain times.

Development of realistic PEV load profiles is an essential for accurate determination of impacts on power system planning and operation applications [3]. The aim is for system operators to have tools to evaluate uncoordinated systems properly in order to build coordination charging schemes based on EV demand. The outputs of this assessment can also be fed into Demand Side Management (DSM) programs to improve power consumption efficiencies. DSM plays an important role in the development of smart grids.

5. Conclusion

This paper has presented an assessment of EV charging using goodness of fit of PDFs. Such evaluations need to be adopted as EV penetration increases and coordination schemes are being developed more intensely. These can then be utilised to build suitable PDFs to estimate country specific metrics and enable forecasting models to be created for EV charging, using data-driven approaches.

References

[1] "Global Greenhouse Gas Emissions Data - US EPA," *US EPA*, 2019. [Online]. Available: https://www.epa.gov/ghgemissions/global-greenhouse-gas-emissions-data. [Accessed: 24- Dec- 2019].

[2] "Use of energy for transportation - U.S. Energy Information Administration (EIA)," *Eia.gov*, 2019. [Online]. Available: https://www.eia.gov/energyexplained/use-of-energy/transportation.php. [Accessed: 24- Dec-2019].

[3] A. Almutairi, M. Alotaibi and M. M. A. Salama, "Goodness of Fit Statistical Analysis for Different Variables of PEV Driver Behaviour," *2018 IEEE Canadian Conference on Electrical & Computer Engineering (CCECE)*, Quebec City, QC, 2018, pp. 1-4, 2018.

[4] Z. Darabi and M. Ferdowsi. "Aggregated Impact of Plug-in Hybrid Electric Vehicles on Electricity Demand Profile, " *IEEE Transactions on Sustainable Energy*, vol. 2, no. 4, pp. 501-508, Oct. 2011.

[5] H. Louie, "Probabilistic modeling and statistical analysis of aggregated electric vehicle charging station load," *Electric Power Components and Systems*, vol. 43, no. 20, pp. 2311-2324, 2015

[6] R. C. Green, L. Wang and M. Alam, "The impact of plug-in hybrid electric vehicles on distribution networks: a review and outlook," *IEEE PES General Meeting*, Providence, RI, pp. 1-8, 2010.

[7] L. Calearo, A. Thingvad, K. Suzuki and M. Marinelli, "Grid Loading Due to EV Charging Profiles Based on Pseudo-Real Driving Pattern and User Behavior," *IEEE Transactions on Transportation Electrification,* vol. 5, no. 3, pp. 683-694, 2019.

[8] S. Bahadoorsingh, C. Meetoo, C. Sharma and P. Hosein, "A Unique Approach to Demand Side Management of Electric Vehicle Charging for Developing Countries," *IEEE International Smart Cities Conference (ISC2)*, Kansas City, MO, USA, pp. 1-2, 2018

[9] M. Flammini, G. Prettico, A. Julea, G. Fulli, A. Mazza and G. Chicco, "Statistical characterisation of the real transaction data gathered from electric vehicle charging stations," *Electric Power Systems Research*, vol. 166, pp. 136-150, 2019.

[10] Ul-Haq, Azhar, Marium Azhar, Yousef Mahmoud, Aqib Perwaiz, and Essam A. Al-Ammar. "Probabilistic modeling of electric vehicle charging pattern associated with residential load for voltage unbalance assessment," *Energies*, vol. 10, no. 9, 1351, 2017.

[11] M. Yilmaz and P. T. Krein, "Review of Battery Charger Topologies, Charging Power Levels, and Infrastructure for Plug-In Electric and Hybrid Vehicles," *IEEE Transactions on Power Electronics*, vol. 28, no. 5, pp. 2151-2169, 2013.

[12] X. Wang and R. Karki, "Exploiting PHEV to Augment Power System Reliability," *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2100-2108, 2017.

[13] D. Wang, X. Guan, J. Wu, P. Li, P. Zan and H. Xu, "Integrated Energy Exchange Scheduling for Multimicrogrid System with Electric Vehicles," *IEEE Transactions on Smart Grid*, vol. 7, no. 4, pp. 1762-1774, 2016.

[14] Residential Energy Consumption Survey (RECS) - Data - U.S. Energy Information Administration (EIA). *Eia.gov*, 2019. [Online]. Available: https://www.eia.gov/consumption/residential/data/2009/. [Accessed: 31- Dec- 2019]

[15] M. Muratori, "Impact of uncoordinated plug-in electric vehicle charging on residential power demand - supplementary data." National Renewable Energy Laboratory. United States, 2017. doi:10.7799/1363870.

[16] M. Muratori, M. J. Moran, E. Serra, and G. Rizzoni, "Highly-resolved modeling of personal transportation energy consumption in the United States," *Energy*, vol. 58, pp. 168-177, 2013.

[17] R Core Team. "R: A language and environment for statistical computing. R Foundation for Statistical Computing, " Vienna, Austria. 2019. http://www.R-project.org/