Development of Methodology to Forecast Charging Demand for Electric Vehicles Using Location Service Data

J. Newman,[1, a)] A. Teyber,[2] and D. Chassin[2]
[1)]*SULI DOE Intern*
[2)]*SLAC National Accelerator Laboratory*

(Dated: 21 August 2020)

Abstract: To reduce the impact of tail-pipe emissions and provide clean energy for all, accurate electric power grid modeling is paramount. Mitigating the effects of tail-pipe emissions can be accomplished via the adoption of electric vehicles (EVs), which can cause challenges due to additional electrical load on the grid. Deployment of new technologies such as EVs coupled with grid modernization through integration of distributed energy resources, demand response, and sophisticated controllers, require updated and increasingly complex grid models that can advise electrical grid stakeholders regarding infrastructure investment and operation. This work explores methodologies to obtain a scalable load model for a fully electric personal vehicle fleet using cell phone location data. Due to computational complexity, the original data set was reduced to an hourly timescale, and the location data was grouped by the first 3-digits of a location's zip code for calculating zip code to zip code driving distances. Using this scaled down data set, an estimation of user driving behavior for the 940 zip code region was found. Furthermore, a linear regression model based on day of the week was used to develop a model for regional EV consumption. The driving behavior provides insight into the energy load required for an EV and information on duration at location to help optimize the placement and type of EV chargers. The reduced model then can be optimized and scaled geographically and eventually be used develop and train a machine learning model to forecast future EV demand at 100% electrification. Ultimately, the aggregation of the driving travel behavior and duration at location data provides an invaluable resource to utilities companies and grid management to understand and plan for future challenges with introducing large scale EV adoptions and can complement the development of more accurate grid models.

I. INTRODUCTION:

The transportation sector is a major contributor of greenhouse gas (GHG) emissions. In 2018, the U.S. transportation sector accounted for the largest percentage of total U.S. emissions at 28.2%[1]. The large scale use of renewable energy resources to provide electrical energy, is a promising solution to mitigate GHG emissions[2,3]. As a result, electric vehicles (EVs) are becoming a popular solution to minimize GHG emissions and reduce reliance on fossil fuels if integrated strategically[4]. Fully electric plug-in vehicles provide a transportation alternative that emits no tail pipe emissions[3,5]. While there are many types of non-traditionally powered vehicles such as electric and gas hybrids and fuel cell powered cars, this study assumes a fully electric plug-in vehicle when referencing EVs. Unlike internal combustion engines, EVs consume electricity as their fuel source. However, the electric grid that provides energy to the EV may not utilize renewable energy sources. Without strategic integration of EVs, power generation using fossil fuels may be used to meet demand. Thus, there is strong motivation for developing accurate and robust grid models to provide electricity sector stakeholders a comprehensive understanding of EV load and opportunities to integrate clean energy.

This study focuses on exploring the impact of U.S. transportation electrification, which is the transition to a fully electric vehicle fleet. Transportation electrification assumes a dramatic increase in EV prevalence. While there are many design and technological challenges associated with EV development[4,6], this study concentrates on overall power grid impact. This transition introduces a number of challenges for electricity sector stakeholders regarding planning and operation of the distribution system given additional load[7]. Thus, there is strong motivation to have extensive and rigorous load models to anticipate these challenges[5,8].

Furthermore, adoption of EVs can lead to regional even nationwide power system issues and grid disruption. Risk such as an increased rate of infrastructure replacement and rapid ramping up of power production could occur[4,9]. To mitigate loss of load to critical customers such as hospitals, community centers, and emergency responders and to ensure successful integration and effective adoption of EVs, it is critical that the load characteristics for EVs are well understood. An understanding of load characteristics allows utilities stakeholders and power grid management to predict future demand and infrastructure development or upgrades. Better understanding of characteristics will help with the development of accurate load models by providing more details for the diverse EV load components[8]. The methodology developed provides a framework for the use of cell phone location service data to infer EV load characteristics. The characteristics then can be used in conjunction with a power flow solver or other grid model to provide more accurate and detailed load models.

[a)]Colorado School of Mines

TABLE I. Example of Data in places.csv from Gimbal

Property	Description	Value
Safe_Graph_Place_Id	Location Id	sg:00001b8625b64052888b8c2f2e3736bb
Location_Name	Location Name	Golden Corral
Six_Digit_NAICS_Code	Industry Code	722511
Street_Address	Address	8032 International Dr
City	City	Orlando
State	State	FL
Five_Digit_ZIP_Code	Zip Code	32819
Safe_Graph_Brand_Ids	Brand Id	SG_BRAND_37b906124bc3dbe4910fb6222a8b37eb

TABLE II. Example of Cell Phone Location Data from Gimbal

Property	Description	Value
Timestamp_With_Second_Precision	YYYY-MM-DD T HH:MM:SS Z	2020-02-01T00:00:00Z
Advertising_User_Id	User Id	0140D62A-F6DA-4951-BB02-1AAAB0CA359C
Advertising_User_Id_Type	ADID or IDFA	ADID
Safe_Graph_Place_Id	Location Id	sg:6dc8445ee8d548d68772048b90960026

II. METHODOLOGY:

A. Description of Data Set

Cell phone location service data has not been extensively documented, developed and validated as a method to predict EV load. Previously used methodologies using location service data[10,11] present areas for improvement. The methodology shown outlines how cell phone location service data can define EV load profiles to further understand EV load characteristics. The cell phone location data was obtained via the third party company called PaeDae (dba Gimbal), a mobile advertising, location solutions, and data company. Three sets of data were obtained, location service data for February 2020 (02/2020) along with reference information for locations and businesses. The first set of data contained over 1 million locations in the U.S. or U.S. territory (Table I) along with the full address; a location reference id called safe_graph_place_id; and business reference id called safe_graph_brand_id, and a list of business information containing the safe_graph_brand_id, brand name, North American Industry Classification System (NAICS) code, and stock symbol if applicable. The second data set contained the list of NAICS codes and the connected National Industry. The final data set contained 29 files of cell phone location data corresponding to each day in February. The files contained location data from 00:00:00 to 23:59:59 each day. The files provided the timestamp with second precision, the specific user id called advertising_user_id, phone OS type called advertising_user_id_type, and the location reference or safe_graph_place_id (Table II). Additionally, a third party data set was introduced, the U.S. Census Zip Code Tabulation Database, to provide latitude and longitude coordinates for a majority of the U.S zip codes.

TABLE III. Geodesic versus Euclidean Distances

Zip Code Pair	Geodesic Distance (km)	Euclidean Distance (km)
(94002,94005)	21.67	21.67
(94002,94025)	12.81	12.81
(94010,94014)	15.20	15.20

B. Preprocessing Data

A majority of the work done was spent preprocessing the cell phone location data into a feasible representation of driving behavior. Additionally, due to the size of the original data set and the computational complexity of preprocessing it, data reduction algorithms were implemented. The first improvement introduced was reducing the timestamp precision to hourly, which significantly reduced the amount of data and better represented driving behavior by removing insignificant travel. Then the region of focus was reduced by grouping the location service data by the first three digits of the location's zip code, to highlight specific areas within the U.S. Finally, to reduce computation time but still maintain relatively large regions of focus, the geospatial scale was set to zip code to zip code distance estimations.

To further speed up computation time, a cached database was created of the ratio (α) between the Euclidean distance (E) and the driving distance (D) calculated using the Google Distance Matrix API (GDM API) (Eq 1). Note that α can equal 0 as travel within the same zip code is currently represented as 0. Also, note that the geodesic distance algorithm[12] which assumes the WGS-84 ellipsoidal model of the Earth, was used to get an estimation of the Euclidean distance. But, at this scale and because no altitude geocoordinates for zip codes were provided it did not significantly affect the results (Table III).

TABLE IV. Current EV Consumption Rates

Make/Model	Year	Energy Use Combined City/Hwy (kWh/100mi)
Tesla/Model 3 Standard Range	2020	26
Tesla/Model X P100D	2019	40
BMW/I3	2020	30
Nissan/Leaf	2018	30

$$\alpha = E/D, \quad 0 \leq \alpha < 1 \tag{1}$$

For each combination of two zip codes in the region, the driving and Euclidean distances and α's were cached in a database. Additionally, estimated driving travel times provided by the GDM API were cached.

This study focused on the analysis of the 940 zip code region which encompasses the San Mateo and Santa Clara counties in California. The location services data was filtered to only include zip codes within this region which resulted in a data set of 379,527 data points for 60,991 users. Note that 60,991 users is the total amount of unique users with location data in the 940 region, but not all users had significant travel data or multiple days of data. Using the cached α database, the data for the 940 region was processed. The travel behavior was used to obtain estimates for the distance traveled for all users within the region, which can be used to forecast EV consumption.

C. Duration at Location and Machine Learning

Additionally, two more uses for the location service data were explored. The travel behavior collected to build EV consumption profiles also contained information about frequently visited locations and time spent at location. Rudimentary exploration of duration information was used to estimate a user's charging behavior throughout the day. Rates were assumed using level two chargers with a charging rate of 6 kWh and an average of current EV battery discharge rates of 30 kW / 100 mi (Table IV). The creation of time sensitive estimates of EV charging profiles assumed that if a user spent at least 1 hour at the same location they could recharge their vehicle until fully recharged or until the user left the location. This methodology needs to be optimized further before applying to multiple users and larger geospatial regions or timescales.

The second methodology explored was using machine learning (ML) techniques to use the EV consumption estimate from the collected data to forecast future loads. The sk-learn python package was used to apply a linear regression based on day of the week for the 940 region EV consumption profile. The collected profile was used to train a model to predict EV load. The linear regression can easily be further refined by including more parameters that influence driving behavior to train the data such as weather, holidays, or other phenomenon.

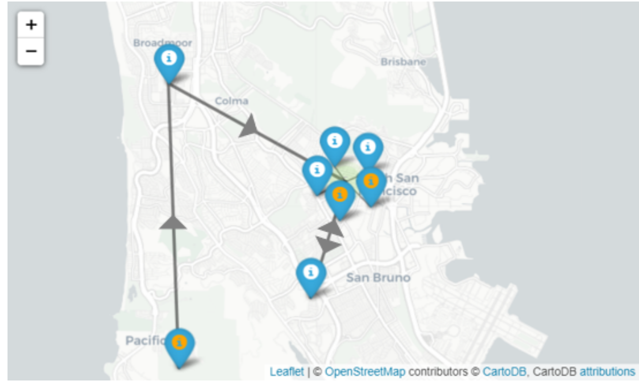

FIG. 1. Map of 1 user's travel for 1 day. Orange markers signify start and end locations. Cluster of locations signify repeated visits.

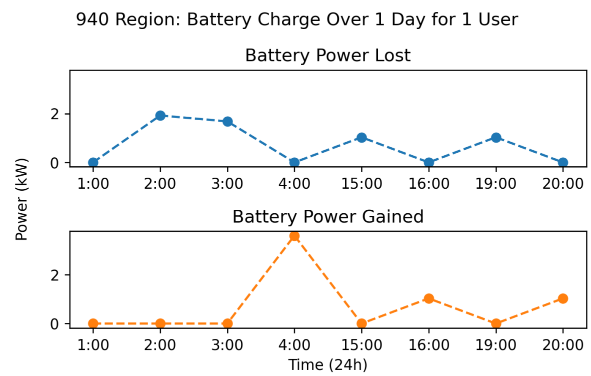

FIG. 2. Estimated power lost traveling and gained via recharging for 1 user's travel for 1 day.

III. RESULTS

A. One User and One Day Analysis

Working with the 940 zip code region, 33 unique zip codes were identified and listed in the U.S. Census Zip Code Tabulation Database. 538 combinations of zip code pairs with their associated α and estimated driving time required to travel between the locations were cached. The cached database was able to be used to estimate user travel for a zip code to zip code resolution. For each user a driving travel path (Fig 1) was estimated along with an estimate for time spent at each location.

The user travel profile (Fig 1) correlates to the user charging profile (Fig 2). For zip code to zip code travel the assumed battery power lost is accumulated over time and decreases once a user is at a location long enough to charge. The battery power gained represents a EV load profile over time. This time specific methodology was not further developed for all users, but could be developed to provide information about timing of loads and load flexibility for EVs.

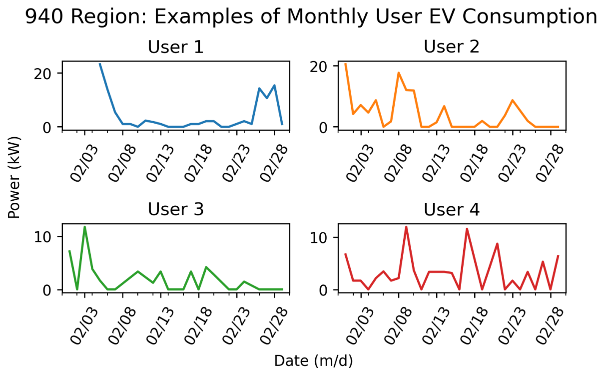

FIG. 3. Estimate of EV load for 4 users and 1 month of data.

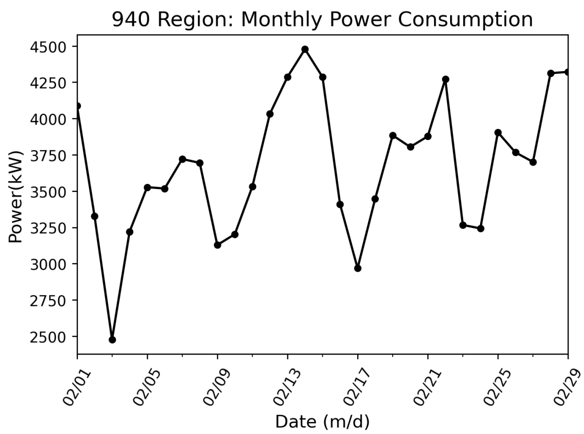

FIG. 4. Estimate of EV load for all users in the 940 region

B. Multi-User and Monthly Analysis

The aggregation of methodology used for one user was then scaled to encompass the monthly data set (Fig 3). Four user EV consumption profiles are shown. The power consumed varied significantly in value and in profile shape. Peak loads for User 1, User 2, User 3, and User 4 of 23.21 kW, 20.45 kW, 11.74 kW, and 11.95 kW occurred on 02/05, 02/01, 02/03, and 02/09, respectively. Each user also had multiple 0 consumption days due to either no data, no zip code to zip code travel, or removal of travel behavior when using a hourly timestamp. The 0 consumption days occurred 6 of 25 days, 10 of 25 days, 13 of 24 days, and 8 of 28 days for User 1, User 2, User 3, and User 4, respectively.

All user travel behavior for the month within the 940 region was aggregated (Fig 4). The result gave a peak load of 4477.82 kW on 02/14.

C. Linear Regression

The final step in developing the methodology was to implement ML techniques to use the data set to predict future EV loads. A linear regression model using day of the week as the parameter was used to train the model (Fig 5).

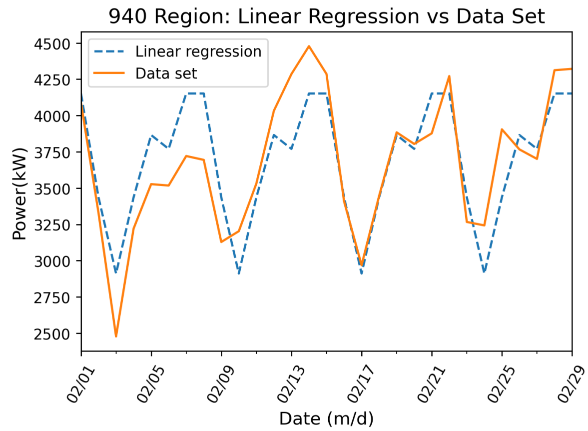

FIG. 5. Estimate of EV load using linear regression based on day of the week

IV. DISCUSSION

The methodology developed provides an estimate of EV charging profiles using location service data. This study focused on the development of new methodology to use cell phone location service data to aid in the creation of accurate EV load profiles. These profiles add insight regarding how much additional demand the electrical grid must be able to support without disruptions. This process can be optimized and scaled to larger regions including the original nationwide data set. Additionally, the further development of an α database for address mapping could become an invaluable tool for future models. While further refinement and validation of the α values collected is necessary, a robust and complete set of nationwide α values could be used to estimate travel distances without calls to an external API like GDM API.

The exploration of new methodology relied on many assumptions and estimations, which introduced many project limitations. Most significantly, no methodology to validate the use of location service data as an accurate representation of driving behavior was implemented. Therefore, the data may not be a valid estimate of travel behavior or provide enough information about specific regions to get a representative set of the population. Further testing and validation is required to prove the durability and reliability of this methodology. Furthermore, many details were lost due to size and computational limits. As the next steps, it is invaluable to quantify how data size reductions have affected the accuracy of the model output.

The use of ML to forecast future EV loads was limited by the data set and the diversity of parameters that influence driving behavior. The limited timescale of the data along with the lack of expected data to compare a model to, made training a ML model difficult. However, addition of parameters such as weather data can be easily implemented to refine and improve the use of ML.

Additionally, further research and methodology refinement will improve model results and the utility of the model. EV consumption and battery capacity is unpredictable and varied depending on car type. To account for the diversity of EVs, additional parameters based on car technology can be added. The generated load profiles can also be feed into a grid distribution model to forecast the effects of additional load. Finally, the travel behavior analysis can be optimized to project different travel patterns given a variety of human behavior patterns.

V. CONCLUSION

Increased EV usage presents a possible solution to reduce tail pipe emissions, yet safe adoption of EVs requires a comprehensive understanding of EV load characteristics. A methodology to go from cell phone location data to zip code resolution driving profiles and EV load estimates was created. The novel methodology detailed in this project provides a resource for creating EV load profiles to further define EV load characteristics to aid the development of accurate grid forecasts. Electrical load for transportation electrification of the 940 zip code region was explored. A majority of the work was focused in preprocessing the data to obtain driving profile information from cell phone location data, creating python scripts, and developing a cached database to speed up computation. A load profile for the users in the 940 region was estimated and a linear regression based on day of the week was applied. Implementation of ML provided a framework for further refinement and development of ML techniques with the introduction of more parameters.

The development and detailing of this methodology provides a fundamental basis for further application of characterising an EV load with future projections in mind. The usage of cell phone location data to infer driving behavior to develop EV load predictions is not a well explored area. This methodology is scalable to apply to larger portions of the cell phone location data used in this study or similar data sets. Several areas of improvement are identified to further refine the algorithm: development of methodology to validate the load model, addition of different car parameters, or integration with grid distribution model. The understanding of EV load characteristics is critical for safe adoption into the grid and for maximizing renewable energy use, thus the further development of the methodology is needed.

VI. ACKNOWLEDGMENTS

This work was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships (SULI) program.

I want to thank the GISMo group at SLAC especially Alyona Teyber, David Chassin, Mayank Malik, and Lily Buechler for their contributions and support on this project.

VII. REFERENCES

[1] Vanguard Renewables, “Inventory of U.S. greenhouse gas emissions and sinks: 1990-2009,” Federal Register **76**, 10026 (2011).

[2] T. Morgan, “Smart grids and electric vehicles: Made for each other?” Automotive Industries AI **192** (2012).

[3] K. M. Tan, V. K. Ramachandaramurthy, and J. Y. Yong, “Integration of electric vehicles in smart grid: A review on vehicle to grid technologies and optimization techniques,” Renewable and Sustainable Energy Reviews **53**, 720–732 (2016).

[4] A. M. Haidar, K. M. Muttaqi, and D. Sutanto, “Technical challenges for electric power industries due to grid-integrated electric vehicles in low voltage distributions: A review,” Energy Conversion and Management **86**, 689–700 (2014).

[5] C. B. Harris and M. E. Webber, “An empirically-validated methodology to simulate electricity demand for electric vehicle charging,” Applied Energy **126**, 172–181 (2014).

[6] J. Yano, S. Nishimura, K. Fukunaga, M. Nakajima, H. Yamada, and M. Moriguchi, “Estimation of EV power consumption and route planning using probe data,” SEI Technical Review , 29–34 (2014).

[7] P. B. Jones, J. Levy, J. Bosco, J. Howat, and J. W. V. Alst, “The Future of Transportation Electrification: Utility, Industry and Consumer Perspectives,” Tech. Rep. (2018).

[8] A. Arif, Z. Wang, J. Wang, B. Mather, H. Bashualdo, and D. Zhao, “Load modeling - A review,” IEEE Transactions on Smart Grid **9**, 5986–5999 (2018).

[9] A. D. Hilshey, P. D. Hines, and J. R. Dowds, “Estimating the acceleration of transformer aging due to electric vehicle charging,” IEEE Power and Energy Society General Meeting (2011), 10.1109/PES.2011.6039848.

[10] S. Selvarajoo, M. Schlapfer, and R. Tan, “Urban Electric Load Forecasting with Mobile Phone Location Data,” 2018 Asian Conference on Energy, Power and Transportation Electrification, ACEPT 2018 , 1–6 (2019).

[11] M. M. Vazifeh, H. Zhang, P. Santi, and C. Ratti, “Optimizing the deployment of electric vehicle charging stations using pervasive mobility data,” Transportation Research Part A: Policy and Practice **121**, 75–91 (2019).

[12] C. F. Karney, “Algorithms for geodesics,” Journal of Geodesy **87**, 43–55 (2013), arXiv:1109.4448.