Proceedings of The National Conference
On Undergraduate Research (NCUR) 2014
University of Kentucky, Lexington, KY
April 3-5, 2014

Parametric Modeling of Electric Vehicle Charging Profiles

Nicole Ng
Electrical and Computer Engineering
Seattle University
901 12th Avenue
Seattle, Washington 98122 USA

Faculty Advisor: Dr. Henry Louie

Abstract

The power used by electric vehicles (EVs) has a significant impact on the electric grid. It is important to know the probabilistic and statistical characteristics of charging demand, as they can improve load forecast accuracy and electric grid operation. In this research, we identified and evaluated parametric models of the electrical power used by EV charging stations. The models are based on charging station data in the Seattle, Washington and San Diego, California areas. To generate these models, we organized the data in both 15-minutely and hourly intervals. We then created and visually inspected the histograms of the data to select parametric distributions that are the most promising. The best candidates were the Weibull, General Extreme Value, Gamma, Inverse Gaussian, Log Logistic, Normal and Lognormal. We estimated the parameters for each distribution using the maximum likelihood estimation procedure. To evaluate the parametric models, we used the chi-squared test, which evaluates the goodness-of-fit of a considered parametric distribution to the data set. A criticism of the chi-squared test is that the number of bins to use is subjective. To mitigate this, we selected three different numbers of bins to evaluate. First we used $Log_2(N)$ to determine the number of bins—Sturges' rule of thumb—where N is the number of samples for either the 15-minutely or hourly interval, and a plus/minus 3 bin sensitivity. The results of the analysis indicate that the Weibull and Generalized Extreme Value distributions are good candidates for EV charging station load modeling. However, the distribution that was the best fit for Seattle was different than for San Diego. This is likely due to time-of-use pricing that San Diego utilizes. Time-of-use pricing encourages energy consumption to be concentrated into a smaller interval by charging people less money for electricity during non-peak hours. Hence the energy draw is more concentrated in San Diego than Seattle. The results can be used to help analyze the impact of increased energy demand caused by EVs, and enable load forecasters to better understand how to plan and operate the grid.

Keywords: Electric Vehicles, Load Profiles, and Parametric Model

1. Introduction

The existing transportation system relies primarily on fossil fuels in the form of petroleum. In the last 100 years, internal combustion engines (ICEs) have been the dominant source of vehicle propulsion. However, there are concerns about using ICEs such as energy independence, inefficiency, noise pollution and harmful emissions. As society's dependence on fossil fuels grows, these negative consequences become more daunting every year. As a more practical sustainable solution for transportation, plug-in electric vehicles (PEVs) enable environmentally-benign and domestically-produced electricity to replace petroleum fossil fuels. Electric vehicles (EVs) consume approximately 300 Wh/mi, which translates into approximately 10 kWh per average commute. ICEs on the other hand are less efficient, consuming 36.6 kWh per average commute[1]. Although EVs are more efficient, they rely on the power grid to recharge their batteries. There are open questions about how this increased consumption of

This work is supported by the Seattle University Center for Environmental Justice and Sustainability, and the Clare Boothe Luce Foundation. ECOtality is supported by the Department of Energy Award Number DE-EE0002194.

Electricity will affect the power grid, particularly at the lower voltage distribution level. In future scenarios, EV recharging will account for an appreciable amount of energy[2]. The effects of this increase in energy demand needs to be examined because it will enable us to understand how to plan and operate the grid with high levels of EV penetration. Researchers have been unable to create parametric models of EV charging station demand because of a lack of data availability. Instead, they rely on indirect data such as driver surveys and vehicle GPS data and then develop mathematical models approximating drive behavior, vehicle weight, vehicle aerodynamic properties, charger specifications and others[2]. The complexity of these models is no longer needed if sufficient direct measurements are available. Recently, ECOtality accumulated enough data to allow researchers to analyze and model EV charging station demand[3].

In this work, we analyze two data sets provided by ECOtality and investigate the appropriateness of several parametric distributions to model the data. We applied a chi-squared goodness-of-fit to data sets covering the Seattle and San Diego areas over a one-year period. The test allowed us to make recommendations for model use and also comment on the statistical significance of the fit of the model.

This paper is organized as follows. In Section II, the charging data sets of Seattle and San Diego are discussed. In Section III, the methodology of how we applied the chi-squared test to fit the distributions is explained. The results of the analysis are specified and discussed in Section IV, and conclusions follow in Section V.

2. Data Set Description

The "EV Project" was launched by ECOtality to collect EV charging station data for certain Nissan LEAF and Chevrolet Volt vehicles. The information that the EV Project collects includes the vehicle model, energy used and time and duration of charger use. We focused the research on the data sets corresponding to Washington State and California for the year 2012. These areas have similar amounts of registered EVs in the EV Project data set and a similar number of charging stations. This allows us to meaningfully compare the data sets and models. Within Washington State (Seattle), and California (San Diego), 1,200 and 1,259 EVs are currently being analyzed under the EV Project, respectively[3].

Although the data sets are similar in many regards, there is one important difference. San Diego utilizes Time-of-Use (TOU) pricing, and Seattle does not. This is represented in the load demand profiles. Under a TOU pricing scheme, the price of electricity is varied in a predetermined manner throughout the day. This allows utilities to operate the grid more economically. In San Diego, the lowest rates occur for the time interval from 0:00-5:00. There is economic incentive therefore, for drivers to program the electric vehicle supply equipment (EVSE) or PEVs to begin charging during this interval.

The data sets contain aggregated EV charging station demand values sampled at 15-minute intervals during 2012 (35,136 total samples). Our work considered the 15-minutely samples as well as the hourly-averages of these samples.

3. General Observations

Figure 1 shows the hourly average consumption load profile for Seattle in the year 2012. The daily average is 186.52 kW. The load is above the average in the evening, beginning at 16:00 continuing to the end of 2:00. At all other times, the load profile is below the average.

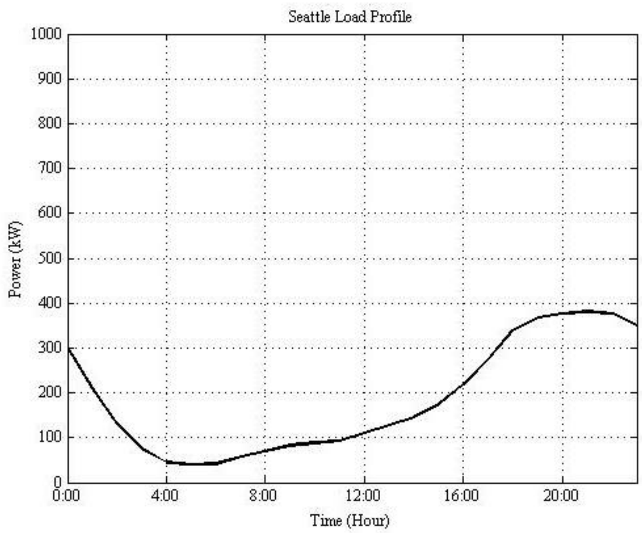

Figure 1: Seattle EV Charging Load Profile

The load profile smoothly transitions throughout the day, peaking around 23:00. The nighttime peak is expected because EV drivers usually come home in the early evening to charge their EVs. As more drivers return, the power consumption continues to increase until 23:00—by this time many EVs have become fully charged. The load steadily decreases until 4:00 then increases as commuters drive to work, some of them charging their EVs at their workplace.

Figure 2 shows the hourly average consumption load profile for San Diego in the year 2012. The daily average is 244.92 kW. The load is above the average in the early morning, beginning at 0:00 to the ending of 3:00. The concentration during this time is likely due to TOU pricing that encourages drivers to charge their EVs during "non-peak hours" (0:00-5:00). The load consumption tends to be smaller outside this period when TOU prices are more expensive.

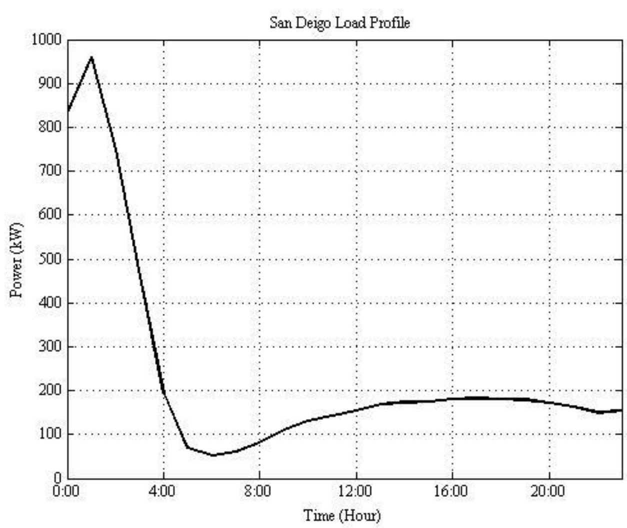

Figure 2: San Diego EV Charging Load Profile

Figures 3 and 4 show the histograms of the Seattle data at 15-minutely and hourly intervals. It is important to note that most of the power consumption in Figures 3 and 4 is around 100 to 200 kW and the maximum power consumed is 875 kW. Due to the relatively smooth load profile, the majority of the time the power consumption is within a narrow range. In fact, 80% of the time the consumption is below 281 kW.

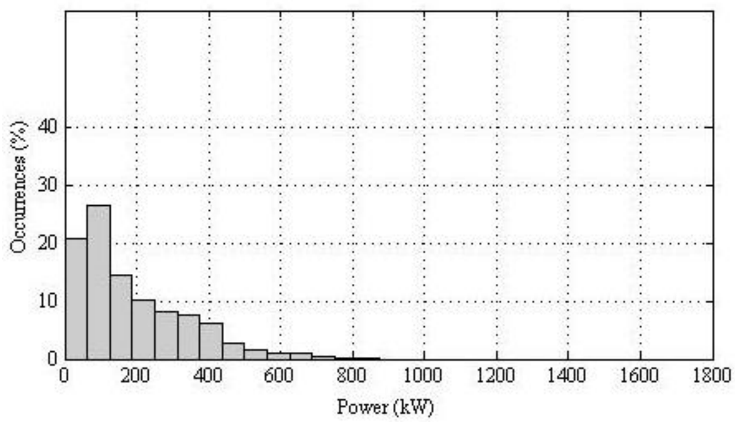

Figure 3: Seattle 15-Minutely Charging Data

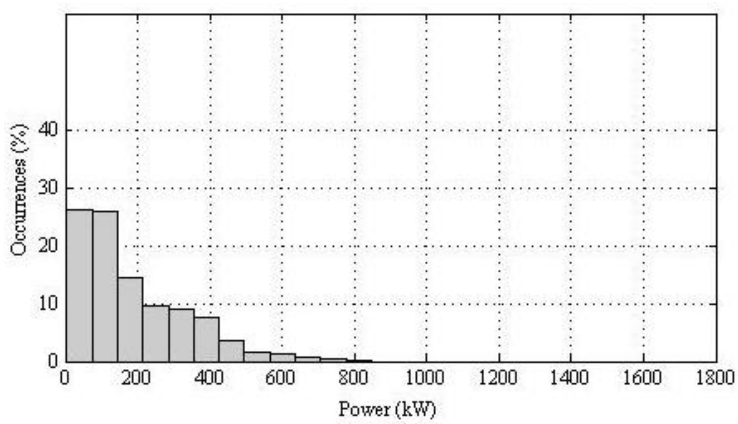

Figure 4: Seattle Hourly Charging Data

Figures 5 and 6 show the histograms of the San Diego data at 15-minutely and hourly intervals. The maximum power consumed is approximately 1500 kW. TOU pricing is likely responsible for the occurrences of extreme consumption shown in Figures 5 and 6. Nearly 12% of the time, consumption is above 484 kW. During the other hours of the day when prices are higher, the demand rarely exceeds 200 kW; this explains the large concentration of occurrences at low power for the San Diego data.

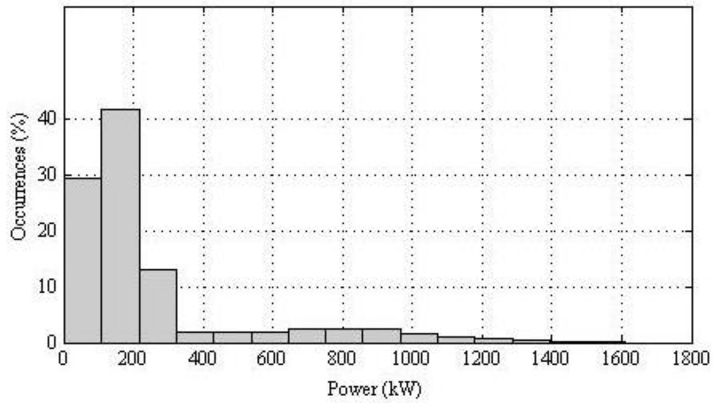

Figure 5: San Diego 15-Minutely Charging Data

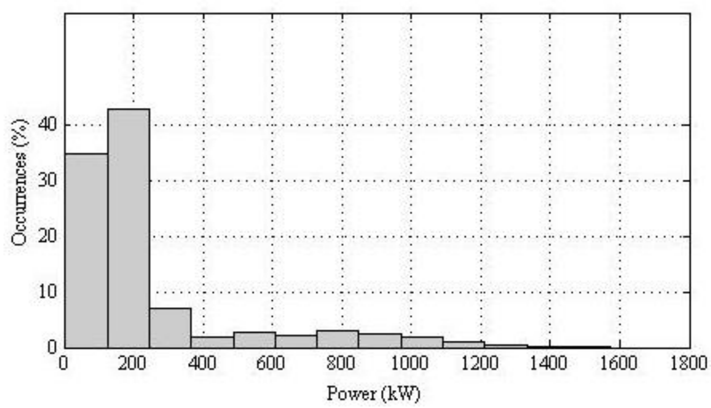

Figure 6: San Diego Hourly Interval Charging Data

Overall, Figures 3-6 show the histograms to be unimodal, having one peak at the 100 or 200 kW mark and being positively skewed. Occurrences of extreme power consumption are possible—particularly in San Diego—but are otherwise rare.

4. Methodology

With the general characteristics of the Seattle and San Diego data sets examined, the next step is to evaluate the parametric distributions. Table 1 lists the different parametric distributions we considered as models for the EV charging station data. We first visually inspected the histograms of the data to select distributions that were the most promising. From the list of distributions in Table 1, the best candidates were the Weibull, General Extreme Value, Gamma, Inverse Gaussian, Log Logistic, Normal and Lognormal.

Table 1: parametric distributions considered

Distribution Name
Beta
Binomial
Birnbaum-Saunders
Burr Type XII
Exponential
Extreme Value
Gamma
Generalized Extreme Value
Generalized Pareto
Inverse Gaussian
Logistic
Log-logistic
Lognormal
Nakagami
Negative Binomial
Normal
Poisson
Rayleigh
Rician
T Location-Scale
Weibull

Next, the parameters for these eight distributions were fit to each data set. The maximum likelihood estimation procedure was used to determine the parameters for each distribution except for the Normal and Lognormal distributions[4]. For the Normal distribution, the estimated value of the standard deviation is the square root of the unbiased estimate of the variance[5]. For the Lognormal distribution, the square root of the unbiased estimate of the variance of the log of the data is the value of the standard deviation parameter[6].

The fit of the distributions to each data set was then evaluated using the chi-squared goodness-of-fit test. The chi-squared goodness-of-fit test is an established, although somewhat subjective, method of statistical hypothesis testing. In the context of this work, the distribution of samples for each data set is compared to each of the eight candidate distributions. A single metric—the so-called χ^2 test statistic—is used to quantify the fit. The χ^2 test statistic has two uses: first, it allows for a comparative analysis of the goodness-of-fit of the candidate distributions, that is, we can rank the candidate distributions based on how well each fits the sampled data; second, it allows us to judge whether or not the fits are statistically significant[7]. It is rare for the true distribution of the sampled data to be the same as the hypothesized (fitted) distribution function. In many goodness-of-fit tests using sampled data, we tend to find that the hypothesized distribution does not fit the data at a level of statistical significance if the sample size is large enough. In our work, the sample size is large. For this reason, we primarily use the χ^2 test statistic to compare the fit of the distributions against each other.

The chi-square test is a goodness-of-fit test that compares the distribution of observed data with a parametric distribution[8]. The null hypothesis is that the data are from the considered parametric distribution. The chi-square (χ^2) test statistic is computed as

$$\chi^2 = \sum_{i=1}^{M} \frac{(O_i - E_i)^2}{E_i}, \tag{1}$$

where i is the bin number, M is the number of bins, O_i are the observed occurrences and E_i are the occurrences expected from a given distribution[5]. An important consideration in the chi-squared test is the number of bins to use. We selected three different numbers of bins to evaluate: $\mathrm{Log}_2(N)$—Sturges' rule of thumb[9]—where N is the number of samples for either the 15-minutely or hourly interval, and a ±3 bin sensitivity to this number of bins.

Values with expected frequencies less than or equal to five are ignored because the chi-squared approximation may not be reliable.

The 15-minutely data sets contain 35,136 samples. Following Sturges' Rule of Thumb, there should be 15 bins. We calculated the fit of the distribution using 15 bins, and then again using 12 and 18 bins to consider the ±3 bin sensitivity. The hourly data sets contain 8,784 samples. Again following Sturgs' Rule of Thumb, there should be 13 bins. We calculated the fit of the distribution using 13 bins, and then again using 10 and 16 bins to consider the ±3 bin sensitivity. We then ranked the distributions in order according to the chi-square test statistic. A lower chi-square value indicates a better fit to the data. Finally, the χ^2 values were tested for their statistical significance, which is based on the sample size of each data set[10].

5. Results

Tables 2 and 3 show the chi-square test statistics for Seattle and San Diego, respectively. The distribution with the lowest χ^2 value is in bold. Table 2 shows that the Weibull distribution is the best-fit distribution for the Seattle. The hourly result in the 13-bin calculation had the lowest chi-square value. Table 3 shows that the Generalized Extreme Value distribution is the best-fit distribution for San Diego. However, there is one exception. The lowest chi-squared value for 10 bins, of the hourly data is the Inverse Gaussian distribution. In general, the results were consistent regardless of the number of bins chosen.

Table 2: χ^2 test statistics for Seattle

	Bins	Weibull	Gamma	Log-normal	GEV
15-minutely	12	**1050**	1276	2266	3173
	15	**1476**	1554	2397	3306
	18	**1855**	1825	2501	3487
Hourly	10	**364**	415	605	850
	13	**340**	383	607	856
	16	**445**	474	636	887

Table 3: χ^2 test statistics for San Diego

	Bins	GEV	Inverse Gaussian	Log-normal	Log Logistic
15-minutely	12	**6216**	6380	7137	7700
	15	**5889**	6354	7195	7594
	18	**6101**	6459	7280	7754
Hourly	10	1344	**1222**	1507	1770
	13	**1611**	1754	1953	2035
	16	**1571**	1664	1900	2034

Figures 7-10 are the histograms of the 15-minutely and hourly intervals of Seattle and San Diego shown with the best fitting distribution for each case. The PDFs for the data are as expected, with a unimodal peak that is positively skewed. Visually, they appear to be reasonable fits to the data.

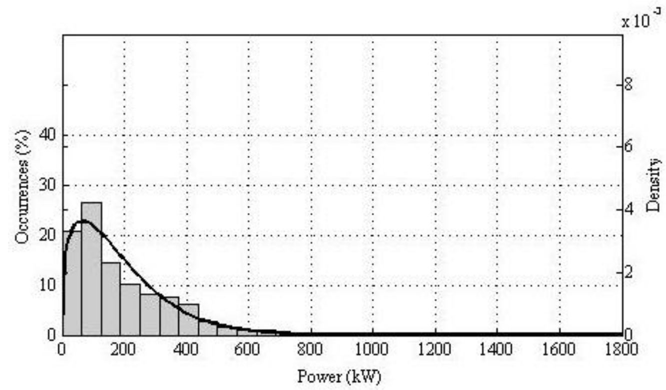

Figure 7: Seattle 15-Minutely Charging Data with a Weibull Distribution

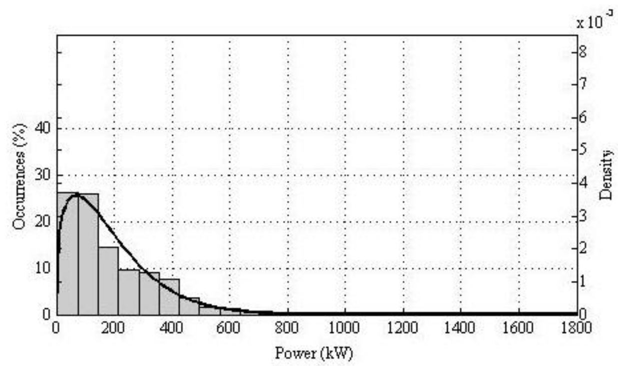

Figure 8: Seattle's Hourly Charging Data with a Weibull Distribution

It is important to note that the Weibull distribution is a reasonably good fit for the Seattle Data.

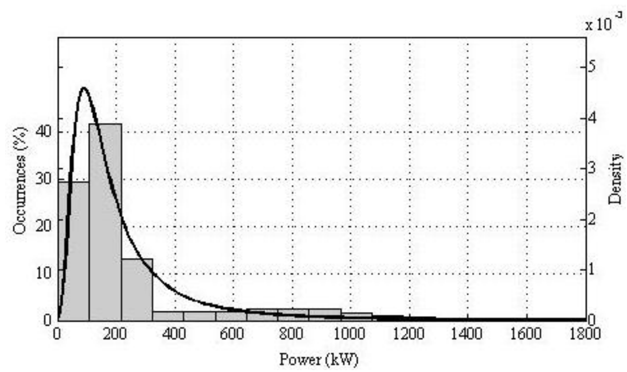

Figure 9: San Diego 15-Minutely Charging Data with a General Extreme Value Distribution

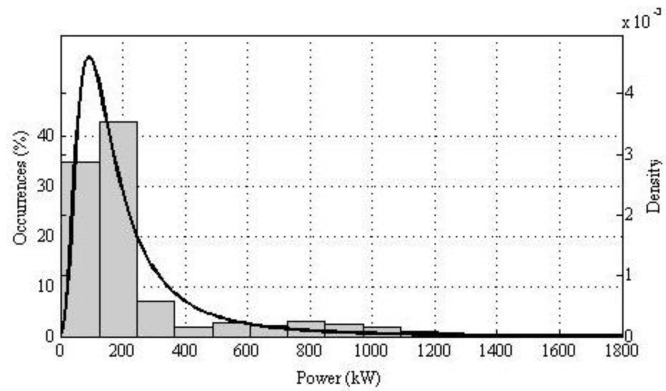

Figure 10: San Diego Hourly Charging Data with a General Extreme Value Distribution

Next, the statistical significance of the results was evaluated. This involved computing critical χ^2 values. The values are a function of the number of bins and number of fit parameters of the considered distributions. Table 4 and Table 5 show the critical values of the best-fitting distributions for Seattle and San Diego. Tables 4 and 5 show that the critical values are rejected, thus rejecting the null hypothesis. This is as expected because the chi-squared values of the San Diego and Seattle were much larger than the critical values, in part due to the large number of samples.

Table 4: critical values for Seattle

	Bins	Dist.	Critical Value	χ^2 Value	Reject?
15-minutely	12	Weibull	17	1050	Reject
	15	Weibull	21	1476	Reject
	18	Weibull	25	1855	Reject
Hourly	10	Weibull	14	364	Reject
	13	Weibull	18	340	Reject
	16	Weibull	22	445	Reject

Table 5: critical values for San Diego

	Bins	Dist.	Critical Value	χ^2 Value	Reject?
15-minutely	12	GEV	16	6216	Reject
	15	GEV	20	5889	Reject
	18	GEV	24	6101	Reject
Hourly	10	Inv. Gauss	14	1222	Reject
	13	GEV	17	1611	Reject
	16	GEV	21	1571	Reject

6. Conclusions and Future Work

This paper modeled electric vehicle charging profiles. The chi-squared test can be used to recommend a parametric distribution to model the charging data collected in the San Diego and Seattle areas. San Diego's best-fit distribution is the Generalized Extreme Value distribution. Seattle's best-fit distribution is the Weibull distribution. These distributions are based on the Sturges' Rule of Thumb bin count. We believe that the parametric distributions fit the data reasonably well, based upon the inspection, probability distribution functions and the χ^2 values, even though the fit was not statistically significant. Time of use pricing makes also makes a difference in the shape of the graph because the charging supports smart coordinate charging on the grid. In the Seattle data, the load profile followed a smooth curve, which varied over a narrow range of power consumption. In the San Diego area, the energy consumption was tightly compressed into a smaller time window because of the time of use charging feature.

This research includes analyzing the compilation of the entire hourly data set for each region, but future work includes creating time dependent models. In other words, separating the data sets into morning, afternoon and nighttime and evaluating the parametric distributions for each time set.

7. Acknowledgments

The author wishes to express her appreciation to her professor, Dr. Henry Louie, who helped guide her through this research project, the Clare Boothe Luce Foundation for funding the project, and Seattle University.

8. References

1. E. Ungar and K. Fell, "Plug In, Turn On, and Load Up," *IEEE Power & Energy Mag.*, vol.8, no.3, pp. 30-35, May/June 2010.
2. M. Kintner-Meyer, K. Schneider, and R. Pratt, "Impacts Assessment of Plug-In Hybrid Vehicles on Electric Utilities and Regional U.S. Power Grids: Part 1: Technical Analysis," [Online]. Available: http://www.pnl.gov/publications/abstracts.asp?report=243074
3. EV Project EVSE and Vehicle Usage Report: 4th Quarter 2012 [Online]. Available: http://www.theevproject.com/downloads/documents/Q4%202012%20EVP%20Report.pdf
4. C. Forbes, M. Evans, N. Hastings and B. Peacock, *Statistical Distributions*, 4th ed. Hoboken, NJ, USA: John Wiley & Sons, 2011
5. G. Box, W. Hunter and J. Hunter, *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, NJ, USA: John Wiley & Sons, 1978
6. Variance [Online]. Available: http://mathworld.wolfram.com/Variance.html
7. W. Conover, *Practical Nonparametric Statistics*, NJ, USA: John Wiley & Sons, 1971
8. J. Taylor, *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*, 2nd ed. Sausalito, CA, USA, 1982.
9. Histograms [Online]. Available: http://onlinestatbook.com/2/graphing_distributions/histograms.html
10. Table: Chi-Square Probabilities [Online]. Available: https://people.richland.edu/james/lecture/m170/tbl-chi.html