

Poster: A review of electric vehicle charging session open data

Yvenn Amara-Ouali, Pascal Massart,
Jean-Michel Poggi
LMO - Université Paris-Saclay
Orsay, FRANCE

Yannig Goude, Hui Yan
OSIRIS - EDF R&D
Palaiseau, FRANCE

ABSTRACT

The need for reliable and accessible electric vehicle (EV) charging data is becoming increasingly important as governments and industries aim to create low-carbon transport systems. Without careful grid management, the security of supply could be compromised. In this work, the results of an open data search are presented with 8 open charging session datasets highlighted and discussed from a practical perspective. These datasets cover a range of charging options (residential, workplace and public) and it is shown that they are all relevant for developing EV load models. To aid with this future modelling work, a distributional analysis of the main influential factors of EV charging load is provided.

CCS CONCEPTS

• **Hardware** → *Smart grid*; • **Mathematics of computing** → *Maximum likelihood estimation*; • **General and reference** → *Empirical studies*.

KEYWORDS

Electric Vehicles, Smart Grid, Load, Open Data, Statistical Modelling

ACM Reference Format:

Yvenn Amara-Ouali, Pascal Massart, Jean-Michel Poggi and Yannig Goude, Hui Yan. 2021. Poster: A review of electric vehicle charging session open data. In *The Twelfth ACM International Conference on Future Energy Systems (e-Energy '21)*, June 28-July 2, 2021, Virtual Event, Italy. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3447555.3466568>

1 INTRODUCTION

Numerous governments are investing in electric mobility and electric vehicles (EV) [3]. Combined with a low-carbon energy mix, they represent a low-carbon alternative to internal combustion engine vehicles. However, the increasing number of EVs will add more constraints onto the grid. Therefore, anticipating the wide adoption of EVs requires an understanding of EV charging behaviours. A variety of EV load models have been proposed in the literature [2]. One obstacle to the production of such models is the lack of reliable data. In particular, the deployment of electric vehicle supply equipment (EVSE) is still in its early stages and charging points do not always communicate relevant data for modelling EV load.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

e-Energy '21, June 28-July 2, 2021, Virtual Event, Italy

© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-8333-2/21/06...\$15.00
<https://doi.org/10.1145/3447555.3466568>

Furthermore, as the market is still in its infancy, data is a strategic asset that EVSE owners are not always willing to distribute openly.

In this work, an open data search focused on finding EV charging session information is presented. This endeavour was undertaken in order to foster reproducible work in the field and is an extension of [2] which reviews EV load models. The rest of the poster is structured as follows: Section 2 details the method and results of the open data search. Section 3 presents 8 open charging session datasets found in this research and their quality is assessed. Finally, statistical characterisations of variables commonly used to reconstruct the EV load are compared in Section 4.

2 OPEN DATA SEARCH

The scope of this study comprises the top 14 countries in terms of EV market share associated with the Electric Vehicle Initiative between 2018 and 2019 [3]. Out of the 860 repositories explored, more than 60 datasets have been found that can be used to model EV load. In particular, three categories of data emerged: historical charging sessions (HCS), real-time charging sessions (RTCS) or traffic (T). HCS are past transactions data, the most relevant information to model EV load. RTCS refers to EVSE occupation information which is updated on short time frames and not stored. It requires regular scraping to be transformed into a HCS dataset. T data refers to traffic counts or travel surveys.

Figure 1 highlights the fact that countries with the highest market share are not necessarily the ones where HCS data was found. This finding demonstrates the existing gap between EV penetration and the availability of open HCS data. This open data search highlighted 8 HCS datasets which altogether cover residential, workplace and public charging.

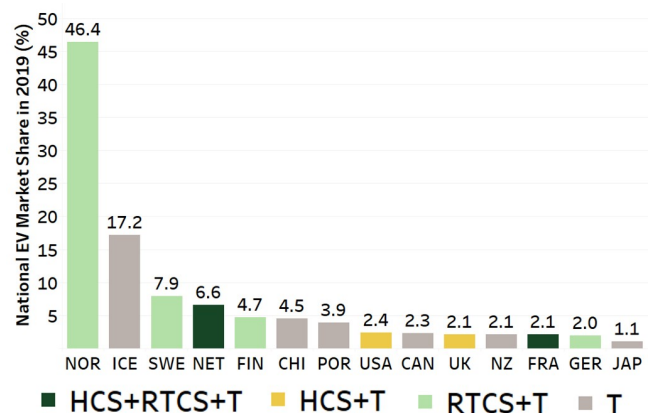


Figure 1: National EV market share [3] coloured by the type of data available for each country in scope.

Table 1: Fields available in the charging session datasets

Fields	Bou	Pal	Dun	Per	Par	Cha	Cal	Sap
Session Start and End Times	✓	✓	✓	✓	✓	✓	✓	✓
Energy Consumed	✓	✓	✓	✓	✓	✓	✓	✓
EVSE Location	✓	✓	✓	✓	✓			
Power Level	✓	✓	✓	✓	✓			
User ID		✓	✓		✓	✓	✓	
Driver Postal Code		✓						
Greenhouse Gas & Gasoline Savings	✓	✓						
Initial State of Charge								✓

3 CHARGING SESSION DATA

The 8 HCS datasets found can be divided into three categories: (a) Municipalities: Boulder (Bou) and Palo Alto (Pal) in the USA, Dundee (Dun) and Perth (Per) in the UK and Paris (Par) in France. (b) Universities/Companies: Caltech (Cal) in the USA and SAP Labs (Sap) in France. (c) Residential: UK Chargepoint Analysis (Cha). The fields shared across all these datasets are the session start and end times as well as the energy consumed (Table 1). 5 datasets provide EVSE location which can help with spatial models. Rarer fields such as driver postal code (Pal) and initial state of charge (Sap) are also available.

A data quality check was run on these datasets to evaluate their reliability: 1. Charge and/or park duration have to be positive and less than 24 hours. 2. Energy consumption needs to be positive and less than 100 kWh. The first criterion discards obvious timestamp errors. In addition, as an EV is fully charged in a few hours maximum, charging sessions lasting for more than a day were discarded. Negative records for the energy consumption were discarded as no EV leaves with a battery less charged than when it arrived. The 100 kWh upper bound was chosen as it is the battery capacity of the Tesla Model S, the largest amongst top selling EVs. All datasets retained more than 85% of transactions after discarding irrelevant observations (Table 2). For example, only 63 transactions were discarded from the Palo Alto dataset which covers more than 9 years of data.

4 DISTRIBUTIONAL ANALYSIS

Three quantities are paramount for reconstructing an EV load curve when working with HCS data: arrival time, charge duration and energy consumed. When the charge duration is not available, the park duration can be used instead. Various distributions have been considered in the literature to model these variables. Namely, Gaussian, exponential, lognormal, Weibull and gamma distributions [2]. The variety of datasets found provides an ideal setup to check whether the distributional assumptions made in the literature can be applied to various datasets. The parameters of the candidate distributions were estimated by maximum likelihood. Table 3 shows which distributional fit minimises the Akaike Information Criteria (AIC).

Table 2: Transactions retained after data preprocessing

Dataset	Timespan	Retained	Discarded
Bou	01-2018/03-2021	21,569 (89.57%)	2,512 (10.43%)
Pal	07-2011/12-2020	259,352 (99.98%)	63 (0.02%)
Dun	01-2017/12-2018	47,051 (89.19%)	5,701 (10.81%)
Per	01-2016/12-2019	63,936 (95.91%)	2,728 (4.09%)
Par	04-2017/05-2017	5,780 (85.72%)	963 (14.28%)
Cha	01-2017/12-2017	2,956,198 (93.06%)	220,605 (6.94%)
Cal	04-2018/04-2021	56,976 (94.44%)	3,357 (5.56%)
Sap	06-2017/01-2021	26,434 (98.55%)	389 (1.45%)

Table 3: Distributions fitted with lowest AIC

Dataset	Arrival	Charge Duration	Park Duration	Energy
Bou	Gaussian	Weibull	Lognormal	Gamma
Pal	Weibull	Weibull	Gamma	Gamma
Dun	Gaussian	-	Lognormal	Weibull
Per	Gaussian	-	Lognormal	Weibull
Par	Gaussian	-	Lognormal	Weibull
Cha	Gaussian	-	Weibull	Gamma
Cal	Gamma	Gamma	Weibull	Gamma
Sap	Gamma	Weibull	Weibull	Weibull

The arrival time is more accurately characterised by a Gaussian distribution across the majority of datasets. The charge duration and energy consumed are better characterised by the Weibull and Gamma distributions with AIC values consistently being much closer relative to the other distributions. Finally, the park duration is better characterised by a lognormal distribution for public charging (except for Pal), while a Weibull distribution is more suited for workplace or residential charging.

5 CONCLUSION AND FUTURE WORK

In this work, an open data search focusing on EV charging session data was conducted. This study highlighted 8 openly available HCS datasets which are presented in this poster. The quality of these datasets was found to be satisfactory for in-depth analysis. The codes and preprocessed data outputs were released on a github repository [1]. We believe that this work will foster reproducible research in the field and will enable better comparisons of the various modelling methods presented in the literature.

REFERENCES

- [1] Yvonn Amara-Ouali. 2021. EV load Open Data. <https://github.com/yvonn-amara/EV-Load-Open-Data> Library Catalog: github.com.
- [2] Yvonn Amara-Ouali, Yannig Goude, Pascal Massart, Jean-Michel Poggi, and Hui Yan. 2021. A Review of Electric Vehicle Load Open Data and Models. *Energies* 14, 8 (April 2021), 2233. <https://doi.org/10.3390/en14082233>
- [3] International Energy Agency. 2019. *Global EV Outlook 2019: Scaling-up the transition to electric mobility*. OECD. <https://doi.org/10.1787/35fb60bd-en>