World Electric Vehicle Journal

MDPI

Article

Electric Vehicle Charging Sessions Generator Based on Clustered Driver Behaviors †

Gilles Van Kriekinge *, Cedric De Cauwer, Nikolaos Sapountzoglou, Thierry Coosemans and Maarten Messagie

EVERGi Research Group, MOBI Research Centre & ETEC Department, Vrije Universiteit Brussel (VUB), Pleinlaan 2, 1200 Brussels, Belgium

* Correspondence: gilles.van.kriekinge@vub.be

† This paper is an extended version of our paper published in 35th International Electric Vehicle Symposium and Exhibition (EVS35), Oslo, Norway, 11–15 June 2022.

Abstract: Increasing penetration of electric vehicles brings a set of challenges for the electricity system related to its energy, power and balance adequacy. Research related to this topic often requires estimates of charging demand in various forms to feed various models and simulations. This paper proposes a methodology to simulate charging demand for different driver types in a local energy system in the form of time series of charging sessions. The driver types are extracted from historical charging session data via data mining techniques and then characterized using a kernel density estimation process. The results show that the methodology is able to capture the stochastic nature of the drivers' charging behavior in time, frequency and energy demand for different types of drivers, while respecting aggregated charging demand. This is essential when studying the energy balance of a local energy system and allows for calculating future demand scenarios by compiling driver population based on number of drivers per driver type. The methodology is then tested on a simulator to assess the benefits of smart charging.

Keywords: electric vehicle; smart charging; simulation; clustering; user behavior

Citation: Van Kriekinge, G.; De Cauwer, C.; Sapountzoglou, N.; Coosemans, T.; Messagie, M. Electric Vehicle Charging Sessions Generator Based on Clustered Driver Behaviors. *World Electr. Veh. J.* **2023**, *14*, 37. https://doi.org/10.3390/wevj14020037

Academic Editor: Grzegorz Sierpiński

Received: 8 December 2022
Revised: 13 January 2023
Accepted: 30 January 2023
Published: 2 February 2023

1. Introduction

Governments are pushing towards the electrification of the transport sector to reduce the human-related CO_2 emissions in order to fight climate change. For instance, the European Parliament adopted the proposal stating that CO_2 emissions of new passenger cars should be reduced by 100% by 2035 [1]. Such climate-focused policy pledges and announcements will increase the number of electric vehicles on the market since they are zero tailpipe emission cars. In the International Energy Agency (IEA) Announced Pledges Scenario (APS), the electric vehicle (EV) global market share is expected to grow up to 30% in 2030 [2].

Following the increase in the number of EVs, new charging infrastructures must be installed at different locations (residences, office buildings, hospitals, etc.). Such new installations should be carried out in a controlled way by assessing the possible impacts on electric systems, and potentially, to act accordingly in case of emerging problems. In scientific literature, simulations show that the installation of new charge points will impact the electric grid on different levels, from local energy systems (LES) [3] to distribution and transmission system operators (DSO and TSO) [4]. Typical issues are for instance voltage deviations, congestion issues and/or increase of peak powers [5]. Solutions to these problems are already strongly studied and often presented in simulations. There exist multiple solutions to the problem of installing new charge points, but two are of high interest: (a) adapt the design of electric system by reinforcing it, and (b) charge in an intelligent manner (smart charging or vehicle-to-grid) [6].

While simulations are of great interest to understand, anticipate and solve future problems, they require accurate data and models to have a correct representation of the reality. With respect to the simulations of EV charging sessions, an important stakeholder needs to be modelled which is the EV driver with their EV. The complexity relies on the modeling of individual human behavior (e.g., arrival and departure time) as well as the characteristics of the EV used by the driver (e.g., charging power). In addition, for most of the use cases, a fleet of EVs needs to be modelled which contains a panel of different EV drivers with different charging behaviors. These complexities are the main subject of this paper which proposes an electric vehicle charging sessions data generator with specific new contributions that simplify the modeling of the charging demand of a fleet of EVs.

1.1. Literature Review

The modeling of the mobility demand has already been conducted in previous works and can be divided into two main stochastic methods. The first method uses Markov chain processes. The Markov chain method associates a probability for the transition from a state to another state, where state could mean, for instance, driving state or charging state.

Authors in [7] propose an inhomogeneous Markov chain method to model the use of EV in order to assess the impact of load profiles at different parking locations. The results show a low deviation with the real mobility dataset. In [8], driver's behavior are simulated using a heterogeneous Markov chain process fitted on data collected by the US Bureau of Labor Statistics. The objective is to predict personal vehicle use and to assess the impact on the electricity demand. Authors in [9] use an inhomogeneous Markov chain process to model driving patterns. The model is fitted using data collected directly from an EV. In [10], authors use a cyclic Markov chain model to describe the behavior of charging stations. From this, they derive charge profiles to include in power grid simulation or to forecast electricity demand.

The advantage of the Markov-chain process is that the model takes the previous state into account, and hence assumes a consequential relationship between states. The disadvantage is that it usually requires abundant empirical data that is often not available. To tackle the issue of need of abundant data, a second method is also used in literature and consists often of using probability density functions without using consequential relationships. Such method is less accurate but easier to implement since it requires less data.

Authors in [11] developed a tool that generates time series of vehicle mobility, driving consumption, grid availability and grid electricity demand. The input parameters are numerous, ranging from EV driver information to EV characteristics and charge point information. Such input data are usually unknown for new charging locations, for instance, when a site wants to design or expand its charging location. Authors in [4] adapted a "Remote-Areas Multi-energy systems load Profiles" (RAMP) software engine, from [12], for stochastic EV driver behavior simulation. The simulator uses EV driver data from surveys to classify the behavior into default groups such as student, workers and inactive users, but also into small, medium and large EV sizes. This methodology has been designed to simulate mass-scale deployment of EVs on a country-level, which is not necessarily suited for smaller charging locations. To simplify this, authors in [13] generated EV charging sessions based on standard probability density functions. Typically, arrival and departure times are based on Gaussian distributions, and the daily mileage is based on log-normal distributions. However, this methodology lacks accuracy when dealing with different types of EV drivers because it assumes one probability density function for all drivers. In [14], authors developed an electric vehicle load profile generator. Their method focuses on generating load profiles directly from existing empirical data from EVs. The dataset used comes from three field trials in Germany from 2011 to 2015. Their method allows for a realistic representation of EV demand. However, it cannot be used to model EV charging behavior for different use cases. Authors in [15] modelled charging profiles of EVs based on real-world electric vehicle charging data. They focus on trip level and conditional probabilities to determine whether an EV will be charged after a trip or not.

The method is applied on residential use cases which makes it very user-centric modeling and not necessarily suited for a pool of EVs. In [16], authors propose a synthetic data generator for electric vehicle charging sessions using a large dataset of 1.8 million sessions. The results show great statistical similarities between the generated and real data. While this method works well for large use cases, it is unknown if it can be applied to specific local energy systems.

Since most of these previous works have different input data, different end-objectives and different techniques to model the mobility demand, Table 1 attempts a classification of the previous cited papers.

Table 1. Literature review classification on mobility demand modeling.

Topic	Classification	Papers
Input data	Survey	[4,7,17]
	Limited empirical data	[9,13,16] & [This paper]
	Abundant empirical data	[8,10,11,14,15]
Use cases	Residential	[7–9,15]
	Local energy system (e.g., office building)	[10,14] & [This paper]
	Large-scale use cases (e.g., country level)	[4,11,13,16,17]
Method	Consequential probabilities	[7–10,16]
	Non-consequential probabilities	[4,11,13–15,17] & [This paper]

1.2. Research Gap and Contributions

The list of papers previously cited shows the interest in modeling the charging behavior. There are still some research questions to be answered. First of all, the models presented previously tend to use different types of datasets leading to different methodologies and results. Some of these papers use abundant empirical data (e.g., EV motor power, battery capacity, temperature, etc.) which usually are not available, difficult to access and very user- and vehicle-centric. Other papers use surveys to build their model which is less accurate than empirical data and not easy to deploy. Very few papers have developed a methodology able to be applied on many different use cases such as local energy systems (e.g., office buildings, shops, etc.). Secondly, none of the previous papers have the ability to capture specific charging behaviors of a specific use case—for instance, specific behaviors such as morning vs. afternoon charging shifts or employee vs. visitor charging behavior. Nonetheless, such particular charging behaviors must be included in the model to have a more accurate simulation.

The previous research gaps are the main research questions that are intended to be solved in this paper. The key contributions of the EV charging sessions generator proposed in this paper are:

- The input dataset of the methodology originates from a standard communication protocol widely available in the interoperability of charging infrastructures. The standard communication protocol allows for applying the methodology on many different and specific use cases (office buildings, shops, houses, etc) and can help to investigate/design different use cases;
- The classification of EV driver's profiles with similar charging behaviors in order to improve the modeling and simulation results. The classification is performed using a clustering technique. In addition, the Kernel Density Estimation process is used to better capture details of each cluster as well as particular charging behaviors;
- The modularity of the generator, its ease-of-use and the standardized output data format are key attributes of its scalability and replicability.

1.3. Structure of Paper

The paper is organized as follows: The methodology of the charging sessions generator is explained in Section 2. The use case upon which the generator is tested is presented in

Section 3.1, and the results are analyzed in Section 3.2. In addition, Section 3.3.1 focuses on the validation of the methodology. Finally, to illustrate the benefit of using the generator, the impact of uncoordinated and smart charging is simulated and presented in Section 3.4.2.

2. Materials and Methods

The structure of the methodology proposed in this paper is illustrated in Figure 1. It consists mainly of four different steps. In a first step, a data pre-processing is carried out and detailed in Section 2.1. This step explains the origin of the data, the cleaning of it and the features it contains. Section 2.2 introduces the clustering algorithm used to characterize the EV charging behavior and the need to normalize the dataset. Followed by this, the main generator algorithm with the statistical parameters is explained in Section 2.3. Finally, Section 2.4 details the validation criteria used to assess the performances of the methodology.

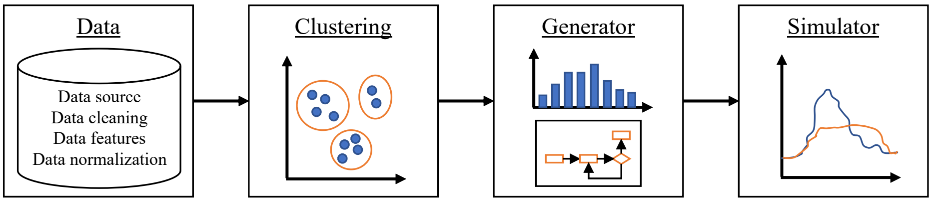


Figure 1. Scheme of the methodology of the generator.

2.1. Data Pre-Processing

2.1.1. Dataset and Features

The dataset consists of individual charging sessions where each of them contains four different features: (a) an identification, (b) a plug-in time, (c) a plug-out time; and (d) an energy consumed. These features are usually available since they are part of the communication standard "Open Charge Point Protocol" (OCPP) between charge points and the charge point operator (CPO) [18]. These charging sessions data are required in order to bill a charging session to an EV driver (using Charge Detail Record (CDR) in OCPP).

To obtain a specific behavior per EV driver, a technique is used to replace all charging sessions from an individual EV driver to one specific theoretical charging session. The method consists of computing the mean value of the plug-in times, the parking times (where parking time is the difference between plug-in and plug-out times) and the energy charged, of all the charging sessions of an EV driver. This results in one theoretical charging session per EV driver consisting of only mean values.

Once each EV driver is associated with one theoretical charging session, they still need to be classified based on the frequency of charging. In other words, a new feature shall be associated with a driver to differentiate regular EV drivers from occasional ones. One would use the number of charging sessions of an EV driver to make the distinction. However, this approach would be limited by the fact that the period where these charging sessions took place would not be considered. For instance, an EV driver that charged 20 times in 20 days is assumed to have the same behavior as an EV driver that charged 20 times in 200 days. To include this aspect, the frequency of charging is used instead of just the number of charging sessions. The frequency of charging, denoted $freq$, is defined in (1):

$$freq = \frac{\text{Number of charging sessions}}{\text{Period between first session and last day of the dataset}} \quad (1)$$

The period in the denominator includes the last day of the dataset and not the last day of an EV driver session. Unless this distinction is made, a driver with a single charging session would present a frequency of one which would misleadingly be interpreted as a 100% probability of having one session per day.

2.1.2. Data Cleaning

The raw dataset needs to be cleaned since it contains charging sessions that failed or did not start. From a time perspective, all charging sessions with parking times lower than five minutes are removed. In addition, duplicated charging sessions, based on the plug-in time, are filtered out. Finally, the time is in UTC so it is adjusted to local time.

A second set of filters is applied on the energy measurements. Firstly, all charging sessions with an energy charged lower than 500 Wh are filtered out. In addition, charging sessions with an energy demand higher than 120 kWh are deleted. Finally, an average charging power is computed per charging session by dividing the energy by the parking time. All average power higher than 22 kW is removed since the charge points are AC technology, hence limited to 22 kW.

2.2. Clustering Technique

The objective is to group EV drivers with a similar charging behavior. There exist multiple techniques to group similar objects together. A well-known approach is an unsupervised machine learning technique called clustering. Among the numerous clustering algorithms in the literature, the most common one is the *k-means* clustering. The objective of this clustering algorithm is to minimize the within-cluster sum of squares. It is mathematically expressed in (2):

$$min \sum_{i=1}^{k} \sum_{x \in S_i} \|x - \mu_i\|^2 \tag{2}$$

where k is the number of clusters, S_i a subset of datapoints of cluster i, and μ_i is the mean value of the subset S_i. An in-depth explanation on the working principle of *k-means* clustering can be found in [19]. Examples of clustering of charging sessions using *k-means* algorithm can be found in [20–23].

The *k-means* algorithm requires the number of clusters as input to cluster. Since it is unknown how many clusters there are, different metrics are used from literature to help identify the correct number of clusters. These are the elbow method [21] and the Davies–Bouldin score [22]. The elbow method consists of plotting the clustering results in the function of the number of clusters. The number of clusters to choose should appear in an elbow on the plot. Intuitively, the more clusters are added, the better the objective function is, but the higher the probability is of over-fitting. Hence, the plot should reflect the balance between the maximum number of clusters and the over-fitting issue. The Davies–Bouldin score is another method which computes the ratio of within-cluster and between-cluster distances. The lower the value is, the optimal number of clusters is.

The clustering proposed in this study works in two stages. A first clustering is performed on the charging sessions' plug-in time, parking time and energy charged features. The second clustering is then performed on the frequency of charging of an EV driver. It has been decided to cluster in two stages since clustering all features together results in a bad grouping of EV drivers and low flexibility in choosing the right number of clusters.

Finally, grouping behaviors together has the advantage of being able to modify certain groups in different ways, making the methodology highly modular. For instance, if a cluster has been identified as employee behavior (early arrival time and long parking time), and it is known from the charging site that many new employees will switch to EVs in the coming years; this specific cluster can be modified to mimic this future behavior specific to the site. Section 3.4.1 shows an example of the strength of the methodology in being able to modify certain clusters to draft future scenarios.

Data Normalization

The features that will be clustered are on different scales, with different ranges, and sometimes on a different order of magnitude. The issue is that clustering techniques rely strongly on the distance between features. Hence, it is important to adjust all feature scales to a common scale by using a normalization technique. Two different common normalization techniques can be found in literature, called *z-score* and *min-max* normalization. The choice between the two depends mainly on the data source and its characteristics. Since the data source is based on human behavior, which is highly stochastic and usually following Gaussian distributions, the technique used is the *z-score* normalization. The formula of the *z-score* normalization is given by (3)

$$x_N = \frac{x_t - \mu}{\rho} \tag{3}$$

where x_t is the non-normalized data point, μ is the mean value, and ρ is the standard deviation.

2.3. Generator Principle

2.3.1. Statistical Distributions

Once the charging sessions have been clustered together, some statistical parameters are extracted to build probability distributions. The main statistical parameters which are extracted per cluster are listed hereunder:

- The probability of having a certain number of charging sessions per day. It has been decided to divide this probability into two probability distributions, mainly one for the working days and one for the weekend days, since the number of sessions are highly different;
- The probability of having an EV plug-in and plug-out at a certain time;
- The probability of having a certain amount of energy to charge.

Each statistical distribution is estimated using a kernel density estimation process from [24] instead of the common Gaussian distribution used in many papers. The advantage relies in a smoother representation of the drivers behavior and hence better results. This is shown in Section 3.3.1.

2.3.2. Pseudo Algorithm

The EV charging sessions generator principle is shown in Algorithm 1. It requires two inputs: (a) the period over which to simulate and (b) the probability distributions per cluster. Using this information, the generator works in two main steps:

- Step (1) For each cluster, and for each day to simulate, a function (called f1) determines the number of charging sessions to generate;
- Step (2) For each charging session to generate, two functions (called f2 and f3) determine the plug-in and plug-out time, and the energy to charge.

Algorithm 1 EV charging session generator

```
Input: Simulation dates, clusters data
for all clusters do
    for all simulation dates do
        f1: Get number of sessions
        for all sessions do
            f2: Get plug-in and plug-out time
            f3: Get energy
Output: Generated charging sessions
```

The first function (f1) chooses a certain number of charging sessions for the day and cluster selected. It will select randomly a number of sessions based on the probability distribution of the frequency feature for a cluster. Similarly, the second function (f2) chooses a plug-in and plug-out time based on the probability distributions of those features. With function (f3), two values are created: (a) the maximum energy that can be charged considering the maximum charging power of the charger during a parking time determined with (f2), and (b) a random energy selection based on the distribution of the energy feature of the cluster. This is carried out to verify that the energy generated for the session is lower than the maximum energy that can be charged.

2.4. Validation Criteria

In order to ensure the validity of the proposed method, the dataset is divided into two subsets. A first subset, called training subset, contains 90% of the dataset, and a second subset, called validation subset, contains the remainder of the dataset. The training subset allows for fitting the model by creating the clusters and to generate EV charging sessions data. Then, the generated EV charging session data can be assessed by comparing it with the validation subset.

The objective is to assess the fit between the generated charging sessions and the real charging sessions. Since the charging sessions are characterized by plug-in time, parking time and energy consumed, histograms of each feature will be shown. Followed by this, a comparison between the histograms from the generated charging sessions and the validation charging sessions will be shown and assessed. The assessment is carried out using the *chi-square histogram distance* evaluation metrics detailed in (4) [25]:

$$\chi^2 = \frac{1}{2}\sum_{i=1}^{n}\frac{(p_i - q_i)^2}{p_i + q_i} \tag{4}$$

where $i = 1, ..., N$ is the i^{th} bin on the total amount of bins N, and p_i and q_i are the occurrences associated with bin i.

3. Results and Discussion

3.1. Use Case

The use case understudy is to simulate the expected charging needs in 2025 of the hospital's parking lot in Brussels. It is an open-access charging location, with paid parking fee. It consists of six charge points containing two Type 2 (IEC 61851) connectors, hence a maximum of 12 connectors, delivering up to 22 kW each. The charging sessions have been logged between May 2018 and January 2022. This period corresponds to 10,477 charging sessions and 424 different EV drivers. To have a better understanding on the use case under study, the number of charging sessions from frequent drivers represents 95.7% of the total number of charging sessions. In addition to a parking lot, solar panels of 590 kWp are installed with an on-site transformer limited to 630 kW. Different buildings are connected to transformer with a low electricity demand compared to the transformer limit.

3.2. Clustering Results

The first step is to cluster the plug-in time, the parking time and the energy charged. Using the elbow method and Davies–Bouldin score, the number of clusters has been set to five. The two methods are shown in Figure 2.

The elbow method does not show a straightforward elbow so it does not help to find the optimal number of clusters. Still, the number of clusters should be between four and seven. On the other hand, the Davies–Bouldin score indicates the optimal number of clusters to be five clusters (lowest score on Figure 2). To have an idea on how EV drivers differ and are grouped, two different clusters are represented, as an example, in Figure 3, by illustrating the plug-in times, parking times and energy needs.

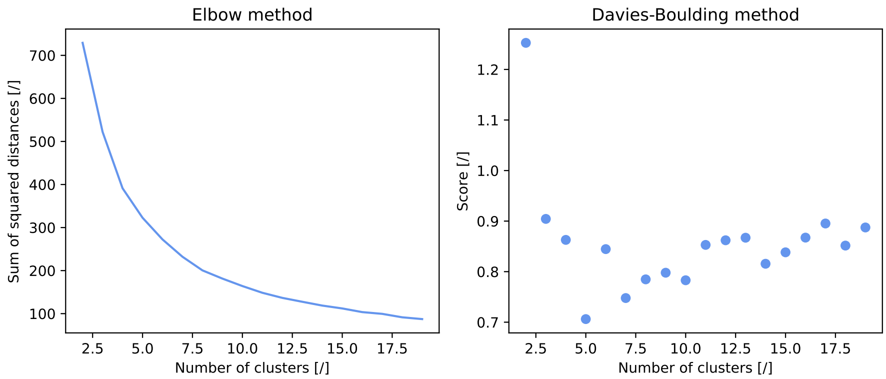
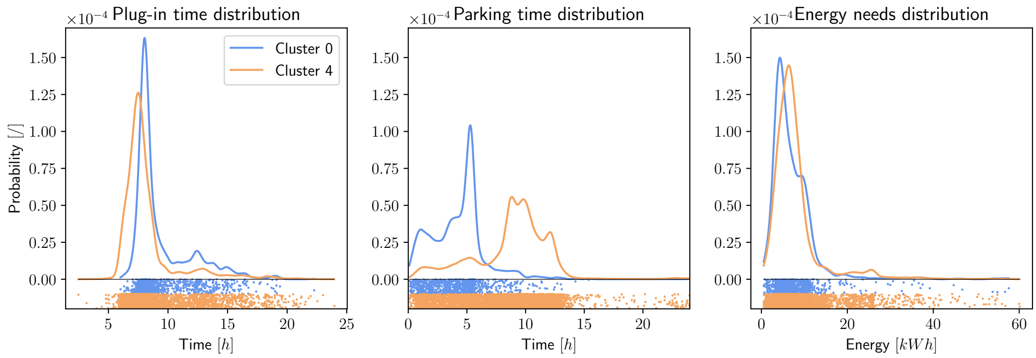


Figure 2. Number of clusters identification for the k-means clustering algorithm.



Note: Scatter points below the zero y-axis are shown to better understand the distribution profiles.

Figure 3. Results of the first clustering.

Both clusters have similar plug-in times and energy demands but a strong difference in parking times. Cluster 0 typically represents an early visitor profile, whereas Cluster 4 typically represents an employee profile, with early arrival and long stay. To have a quantitative understanding, Table 2 summarizes the characteristics of the five clusters.

Table 2. Clusters' quantitative characteristics.

Cluster ID	# of Sessions	# of Drivers	Plug-In Time (Mean Value)	Parking Time (Mean Value)	Energy (Mean [kWh])	Sub-Clusters
Cluster 0	1088	104	Morning (09h26)	Mid (04h15)	Low (7.22)	2
Cluster 1	826	139	Afternoon (15h51)	Mid (05h52)	Low (9.31)	2
Cluster 2	521	39	Morning (09h42)	Long (07h03)	High (40.4)	2
Cluster 3	2	1	Afternoon (16h45)	Very long (38h48)	Low (5.07)	N.A.
Cluster 4	6618	69	Morning (08h15)	Long (08h58)	Mid (7.9)	3

Note: Parking time: "Short" below 3 h, "Mid" for [3–6] h, "Long" for [6–9] h, "Very long" over 9 h. Energy: "Low" below 10 kWh, "High" over 10 kWh.

Table 2 shows a couple of different charging behaviors which are interesting to analyze. First of all, the mean energies are relatively low except for Cluster 2 (which represents 9.2% of the number of total EV drivers). Cluster 2 and Cluster 4 have the same parking behavior (morning arrival and long stay) but not the same charging behavior (higher energy demand). A second interesting result to discuss is that Cluster 3 is represented only by one driver because this driver has a very different behavior compared to the other clusters.

This cluster shows the good performances of the clustering technique since it is able to exclude outliers into an individual cluster. Finally, Cluster 0 and Cluster 1 differ from the other clusters with short parking times, and are themselves differentiated by morning and afternoon arrivals. To conclude, Table 2 shows high differences in behavior between clusters. Hence, it shows the importance to cluster drivers' behavior.

The last column of Table 2 shows the results of the second level clustering. It indicates the number of subclusters based on the frequency of charging. The results show that most of the clusters are divided into two subclusters, mainly in low and high frequency of charging. Nevertheless, for certain clusters, more than two subclusters are required due to the higher frequency divergence of charging.

3.3. Generator Results

3.3.1. Validation

Figure 4 shows three histograms that consist of plug-in times, parking times and energy needs, for both generated charging sessions and the real charging sessions (from the validation subset). The number of charging sessions generated is 1335, which is close to the original number of charging sessions which is 1422, on a period of four months.

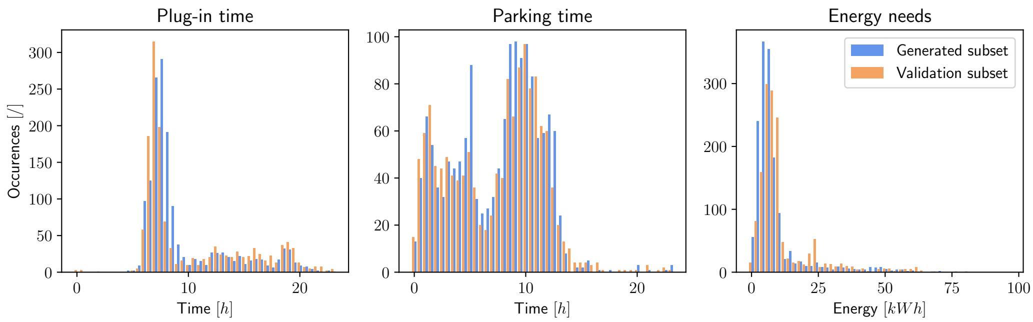


Figure 4. Comparison of plug-in time, parking time and energy needs between generated and real charging sessions.

The results show a good match between the generated and the real charging sessions. The methodology proposed in this paper is able to capture different subtleties. For instance, regarding the plug-in times, the method correctly generates many charging sessions in the morning but tends to underestimate the afternoon charging sessions. Similarly, the generated parking times follow correctly the validation subset but with an underestimation of short parking times. Finally, the generated energy needs fit the validation subset, with an underestimation of higher energy demand. This is due to a low number of charging sessions with high energy demand (as shown in Table 2). In general, Figure 4 shows the great representativeness of the methodology proposed in this paper with a close match between the generated subset and the validation subset.

3.3.2. The Impact of Clustering and Kernel Density Distribution

Two important novelties are included in the methodology of this paper which are the introduction of clustering and the use of kernel density distributions. It is therefore interesting to assess their impact on the validation subset. Four different scenarios are tested and assessed by computing the *chi-square histogram distance* using (4). The results are presented in Table 3.

Table 3. Chi-square distance evaluation for different scenarios.

Scenario ID	Description	Plug-In Time	Parking Time	Energy Needs
1	Gaussian distribution without clustering	491.8	135.7	261.3
2	Gaussian distribution with clustering	379.1	167.7	136.7
3	Kernel distribution without clustering	123.9	48.9	221.4
4	Kernel distribution with clustering	103.7	37.9	181.5

From Table 3, it can be observed that the introduction of kernel distribution has a positive impact, by reducing the *chi-square distance*, on the plug-in time and parking time but not on the energy needs. The main reason is that the energy needs distribution in Figure 4 strongly follows a Gaussian distribution. It can also be observed that the introduction of clustering has a positive impact on the three features, where scenario 4 has the lowest chi-square distance. It is important to note that the clustering main objective is not to improve the model, but rather to make it modular by allowing changes in groups of EV drivers.

3.3.3. The Evolution in Charging Behavior

It has been observed that the charging session generator tends to underestimate the energy demand. From the example of the previous section, the total energy demand from the generated charging sessions is 14,480 kWh, which is 19.2% lower than the real validation total energy demand of 17,933 kWh. To understand such underestimation, Figure 5 shows the daily energy demand of the full dataset.

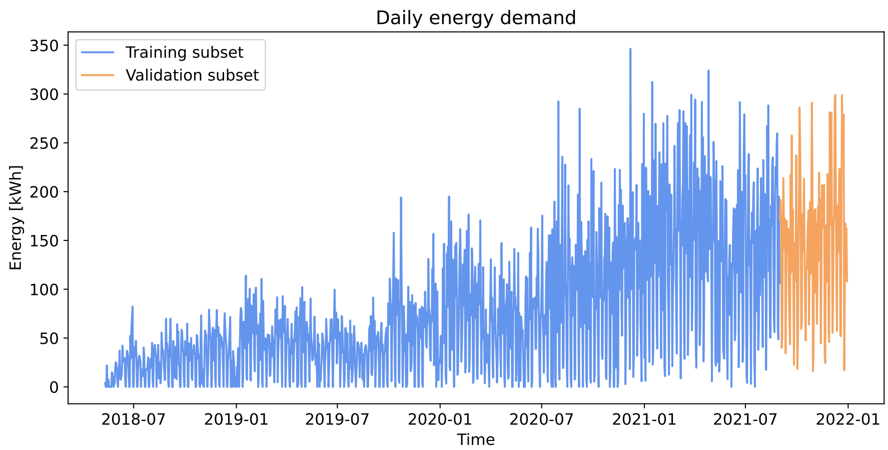


Figure 5. Daily charging energy demand evolution.

From Figure 5, it can be observed that the daily energy demand has a strong evolution over the four years of the dataset. Two main reasons can explain such evolution which are an increase of the number of users and the evolution from PHEVs to BEVs. The drawback of the methodology proposed in this paper is that it includes low energy demand charging sessions (from 2018 until 2020) in the statistical distributions. Consequently, it will reproduce such energy demands in the validation subset, hence lowering the overall energy demand. An interesting future research would be to find a methodology that can model the evolution.

3.4. Simulation Results

3.4.1. Scenario Construction

The final objective of the methodology proposed in this paper is to generate EV charging session data based on clustered EV drivers' behavior. Generating EV charging session data directly from the clusters will reproduce the same scenario as the existing one (from the existing data) which is not the goal. To generate future scenarios, some changes in the clusters are required to build more realistic (future) scenarios. How these scenarios are built is the main subject of this section.

It is known from literature that the EV penetration follows an exponential behavior, and that it will likely still follow an exponential behavior in the future [2]. Knowing this, the clusters are expected to follow an exponential behavior. Consequently, an extrapolation technique is used to evaluate the trend of each cluster and the expected future behavior. In this paper, a *nonlinear least squares* method is used to extrapolate the clusters behaviors [26]. Results of this method are shown in Figure 6.

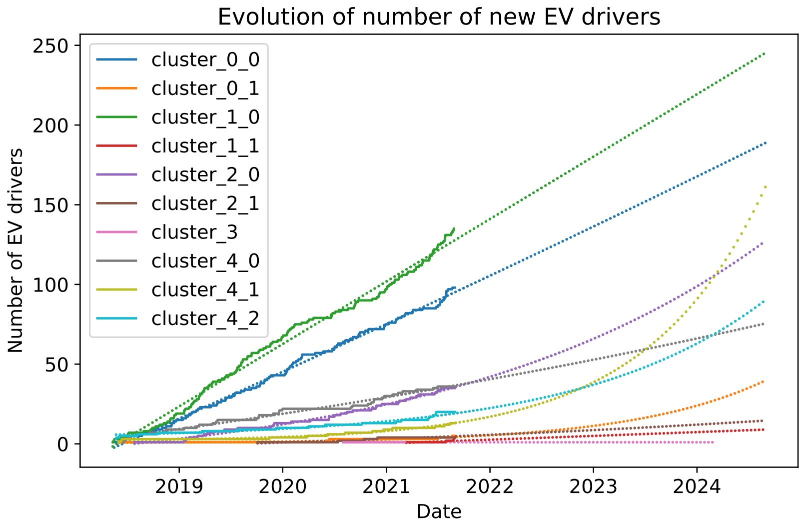


Figure 6. Extrapolation of number of charging sessions per cluster.

From the extrapolation technique, a scenario can be drafted for 2025. The clusters are then accordingly modified, and the scenario will be simulated in Section 3.4.2.

3.4.2. Uncoordinated vs. Smart Charging

The use case understudy is a hospital parking lot which desires to install new charge points to meet the future EV charging demand in 2025. The first objective is to analyze the impact of installing new charge points using uncoordinated charging on the local energy system. The second objective is to analyze the positive impacts of coordinating the charging processes using a smart charging algorithm.

In order to achieve this, the generated charging sessions produced by the generator are simulated in a simulator available in [6]. The smart charging algorithm is a model predictive control algorithm that minimizes the cost of charging and the peak powers. In this paper, perfect forecast is used to simplify the simulations. The results are presented in Figure 7 showing the charging power profile dynamics by building, for each quarter hour, the mean values, 1st and 3rd quartiles values and the maximum values of full simulation.

Figure 7 shows the strong impact of uncoordinated charging on the parking lot of the hospital. The maximum values go up to 728 kW, exceeding the maximum capacity of the on-site 630 kW transformer. In addition, important injection can be seen during the day, mainly because the EVs are fully charged. When controlling the charge points in a smart way, the maximum peaks, the 3rd quartile and the mean values are strongly reduced to take advantage of overproduction during the day. This is the peak shaving and valley filling principle. The self-consumption changes from 27.5% for uncoordinated charging to 59% for smart charging.

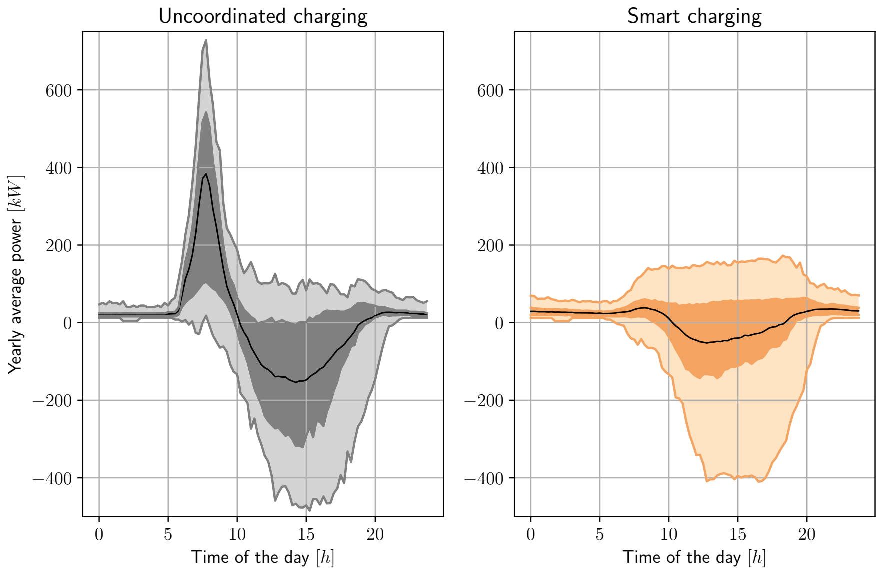


Figure 7. Six-month simulation of uncoordinated and smart charging.

4. Conclusions

In this paper, an EV charging sessions generator is presented. It enables the creation of charging sessions data based on historical data of a specific charging location. The historical data have been analyzed to group different types of EV drivers together. For each group, specific sets of statistical parameters are extracted, which are then used by the generator. The full methodology is applied on a use case of a hospital which plans to expand its EV fleet. The results are presented in two parts.

A first part is dedicated on the clustering of EV charging sessions from historical data. The different types of EV drivers are presented with some important characteristics. The results show the big differences in behavior between the EV drivers and the importance of grouping such drivers. A second part focuses on the actual generation of EV charging sessions data. The generator is validated by comparing the historical data set with a newly generated session according to populated clusters. A scenario is then defined and analyzed by simulating uncoordinated charging and smart charging from the generated charging sessions. The results indicate the strong impact on power and energy demand when adding new EV drivers to the population. The analyzed scenario highlights the need for grid reinforcement or smart charging technologies to avoid overloading and peak demands due to an increase of charging sessions of specific EV driver types. The good results of the validation process demonstrate the potential of this generator to simulate new scenarios while the scenario analysis demonstrates its usefulness to analyse future EV transition scenarios.

Future research could include a method to follow the evolution of the energy demand of EVs since drivers are switching from plug-in hybrid EVs to battery EVs requiring more charging needs.

Author Contributions: Conceptualization, C.D.C. and G.V.K.; methodology, G.V.K.; software, G.V.K.; validation, G.V.K., C.D.C. and N.S.; formal analysis, G.V.K.; data curation, C.D.C.; writing—original draft preparation, G.V.K.; writing—review and editing, C.D.C. and N.S.; visualization, G.V.K.; supervision, C.D.C. and N.S.; project administration, T.C. and M.M.; funding acquisition, T.C. and M.M. All authors have read and agreed to the published version of the manuscript.

Funding: The authors would like to thank the Agency for Innovation and Entrepreneurship (VLAIO) for providing funding of this research under Grant No. HBC.2018.0519 and Flux50 for support to our team.

Data Availability Statement: The datasets used during the current study are not available since they are protected by privacy reasons, mainly because of the presence of EV users' radio frequency identification (RFID).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

APS	Announced Pledges Scenario
BEV	Battery Electric Vehicle
CPO	Charge Point Operator
CDR	Charge Detail Record
DSO	Distribution System Operator
EV	Electric Vehicle
IEA	International Energy Agency
LES	Local Energy System
OCPP	Open Charge Point Protocol
PHEV	Plug-in Hybrid Electric Vehicle
RAMP	Remote-Areas Multi-energy systems load Profiles
RFID	Radio Frequency Identification
TSO	Transmission System Operator

References

1. European Parliament. 2021/0197(COD)-CO2 Emission Standards for Cars and Vans. 2022. Available online: https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?reference=2021/0197(COD)&l=en (accessed on 1 February 2023).
2. International Energy Agency (IEA). *Global EV Outlook 2022*; IEA: Paris, France, 2022; pp. 98–100. Available online: https://www.iea.org/reports/global-ev-outlook-2022 (accessed on 1 February 2023).
3. Firouzjah, K.G. Profit-based electric vehicle charging scheduling: Comparison with different strategies and impact assessment on distribution networks. *Int. J. Electr. Power Energy Syst.* **2022**, *138*, 107977. [CrossRef]
4. Mangipinto, A.; Lombardi, F.; Sanvito, F.D.; Pavičević, M.; Quoilin, S.; Colombo, E. Impact of mass-scale deployment of electric vehicles and benefits of smart charging across all European countries. *Appl. Energy* **2022**, *312*, 118676. [CrossRef]
5. Dubey, A.; Santoso, S. Electric Vehicle Charging on Residential Distribution Systems: Impacts and Mitigations. *IEEE Access* **2015**, *3*, 1871–1893. [CrossRef]
6. Van Kriekinge, G.; De Cauwer, C.; Sapountzoglou, N.; Coosemans, T.; Messagie, M. Peak shaving and cost minimization using model predictive control for uni- and bi-directional charging of electric vehicles. *Energy Rep.* **2021**, *7*, 8760–8771. [CrossRef]
7. Fischer, D.; Harbrecht, A.; Surmann, A.; McKenna, R. Electric vehicles' impacts on residential electric local profiles – A stochastic modelling approach considering socio-economic, behavioural and spatial factors. *Appl. Energy* **2019**, *233–234*, 644–658. [CrossRef]
8. Muratori, M. Impact of uncoordinated plug-in electric vehicle charging on residential power demand. *Nat. Energy* **2018**, *3*, 193–201. [CrossRef]
9. Iversen, E.B.; Morales, J.M.; Madsen, H. Optimal charging of an electric vehicle using a Markov decision process. *Appl. Energy* **2014**, *123*, 1–12. [CrossRef]
10. Gruosso, G.; Gaiani, G.S. A Model of Electric Vehicle Recharge Stations based on Cyclic Markov Chains. In Proceedings of the IECON 2019-45th Annual Conference of the IEEE Industrial Electronics Society, Lisbon, Portugal, 14–17 October 2019; Volume 1, pp. 2586–2591. [CrossRef]
11. Gaete-Morales, C.; Kramer, H.; Schill, W.P.; Zerrahn, A. An open tool for creating battery-electric vehicle time series from empirical data, emobpy. *Sci. Data* **2021**, *8*, 1–18. [CrossRef] [PubMed]
12. Lombardi, F.; Balderrama, S.; Quoilin, S.; Colombo, E. Generating high-resolution multi-energy load profiles for remote areas with an open-source stochastic model. *Energy* **2019**, *177*, 433–444. [CrossRef]
13. Jiang, R.; Zhang, Z.; Li, J.; Zhang, Y.; Huang, Q. A coordinated charging strategy for electric vehicles based on multi-objective optimization. In Proceedings of the 2017 2nd International Conference on Power and Renewable Energy, ICPRE, Chengdu, China, 20–23 September 2017, pp. 823–827. [CrossRef]
14. Schäuble, J.; Kaschub, T.; Ensslen, A.; Jochem, P.; Fichtner, W. Generating electric vehicle load profiles from empirical data of three EV fleets in Southwest Germany. *J. Clean. Prod.* **2017**, *150*, 253–266. [CrossRef]
15. Brady, J.; O'Mahony, M. Modelling charging profiles of electric vehicles based on real-world electric vehicle charging data. *Sustain. Cities Soc.* **2016**, *26*, 203–216. [CrossRef]
16. Lahariya, M.; Benoit, D.F.; Develder, C. Synthetic Data Generator for Electric Vehicle Charging Sessions: Modeling and Evaluation Using Real-World Data. *Energies* **2020**, *13*, 4211. [CrossRef]

17. Harris, C.B.; Webber, M.E. An empirically-validated methodology to simulate electricity demand for electric vehicle charging. *Appl. Energy* **2014**, *126*, 172–181. [CrossRef]
18. Open Charge Allaince-OCPP2.0. 2022. Available online: https://www.openchargealliance.org/protocols/ocpp-20/ (accessed on 1 February 2023).
19. Aggarwal, C.C. *Data Mining*; Springer International Publishing: Cham, Switzerland, 2015; Volume 1, pp. 162–163. [CrossRef]
20. Wang, Y.; Cai, S.; Shen, Y.; Shao, D.; Gong, X.; Zhou, C.; Chu, D. The analysis of electrical vehicles charging behavior based on charging big data. In Proceedings of the 2019 IEEE 4th International Conference on Cloud Computing and Big Data Analytics, ICCCBDA, Chengdu, China, 12–15 April 2019; Volume 1, pp. 63–67. [CrossRef]
21. Crozier, C.; Apostolopoulou, D.; McCulloch, M. Clustering of Usage Profiles for Electric Vehicle Behaviour Analysis. In Proceedings of the Proceedings-2018 IEEE PES Innovative Smart Grid Technologies Conference Europe, ISGT-Europe, Sarajevo, Bosnia and Herzegovina, 21–25 October 2018; Volume 1, pp. 1–6. [CrossRef]
22. Xydas, E.; Marmaras, C.; Cipcigan, L.M.; Jenkins, N.; Carroll, S.; Barker, M. A data-driven approach for characterising the charging demand of electric vehicles: A UK case study. *Appl. Energy* **2016**, *162*, 763–771. [CrossRef]
23. Xiong, Y.; Wang, B.; Chu, C.C.; Gadh, R. Electric Vehicle Driver Clustering using Statistical Model and Machine Learning. In Proceedings of the IEEE Power and Energy Society General Meeting, Portland, OR, USA, 5–10 August 2018; Volume 2018. [CrossRef]
24. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
25. Pele, O.; Werman, M. The Quadratic-Chi Histogram Distance Family. In *Lecture Notes in Computer Science, Proceedings of the Computer Vision—ECCV 2010, Crete, Greece, 5–11 September 2010*; Daniilidis, K., Maragos, P., Paragios, N., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; Volume 6312, pp. 749–762. [CrossRef]
26. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [CrossRef]