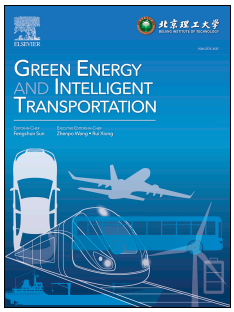
Journal Pre-proof



Clustering Methodologies for Flexibility Characterization of Electric Vehicles Supply Equipment

Marcelo Forte, Cindy P. Guzman, Alexios Lekidis, Hugo Morais

PII: S2773-1537(25)00054-4

DOI: https://doi.org/10.1016/j.geits.2025.100304

Reference: GEITS 100304

To appear in: *Green Energy and Intelligent Transportation*

Received Date: 10 June 2024

Revised Date: 21 August 2024

Accepted Date: 12 September 2024

Please cite this article as: Forte M, Guzman CP, Lekidis A, Morais H, Clustering Methodologies for Flexibility Characterization of Electric Vehicles Supply Equipment, *Green Energy and Intelligent Transportation*, https://doi.org/10.1016/j.geits.2025.100304.

Clustering Methodologies for Flexibility Characterization of Electric Vehicles Supply Equipment

Abstract

The continuous growth of electric vehicles (EVs) poses new challenges to power systems planning and operation due to the need to meet society's decarbonization goals. In this context, clustering has emerged as a powerful tool to help understand and categorize the uncertain behavior of EV users and the electric vehicle supply equipment (EVSE) needs. However, previous studies lack empirical European EV charging data and relevance for practical applications. Therefore, to address such issues, this study evaluates different clustering techniques to identify typical EV charging profiles and, mainly, usage flexibility. The defined methodology comprises three major stages: data preprocessing, clustering application, and validation of results. We conduct benchmarking based on EV energy consumption, arrival, and sojourn times, using K-means, Gaussian mixture model, and Hierarchical clustering. This method allows greater applicability to various datasets from different regions, producing more comprehensive profiles that can provide empirical flexibility data in a visual, intuitive, and relevant approach. A use case considering EV charging data from Caltech University and Greece is utilized to test the proposed methods, demonstrating the versatility of our methodology. Specifically, Caltech features highly flexible prolonged charging sessions, while Greece exhibits quick-stay sessions with less flexibility potential. Both contexts offer opportunities to use the available flexibility for coordination with renewable energy sources and help balance the grid. This information unlocks the potential for future studies, enabling distribution system operators and charge point operators to intelligently and successfully integrate EVs into the energy system.

Keywords: Clustering, Electric Vehicles, Electric Vehicles Supply Equipment, Power Flexibility, Power Systems Planning

1. Introduction

1.1. Motivation

The adoption of electric vehicles (EVs) has experienced rapid growth in the 21st century. This trend is driven by the pressing need to transition global energy demand away from fossil fuels, particularly within the past decade [1]. To achieve this goal, intelligent management methods adapted to transportation should be adopted.

Considering the concerns related to climate change, the European Union (EU) aims to be carbon-neutral by 2050. This objective is the heart of the European Green Deal and in line with the EU's commitment to global climate action under the Paris Agreement [2], since Transport is the only sector where greenhouse gas (GHG) emissions have increased in the past three decades [3]. This sector was responsible for more than a quarter of Europe's energy consumption in 2019, of which roughly 71% came from road transportation, according to a 2022 report [4].

To achieve carbon neutrality, the EU's environment Ministers approved the "Fit for 55 in 2030" package in 2022 [5], mandating that only zero-emission vehicles can be sold in Europe from 2035. The United States of America (USA) and the United Kingdom (UK) are also targeting net-zero emissions by 2050, China and Russia by 2060, and India by 2070. These nations, alongside the EU, represent the biggest contributors to global emissions [6]. In response, governments and car manufacturers have intensified investments in new EV models and tax incentives, contributing to a marked increase in EV adoption over the past five years [7]. Despite the promising growth, the rapid rise of EVs poses significant challenges to power systems, particularly at the distribution level [8, 9]. Uncontrolled EV charging can destabilize the existing power grid, causing voltage fluctuations, system overcurrents, and deterioration in power quality [10]. Additionally, the widespread deployment of electric vehicle supply equipment (EVSE) introduces further complexities [11], such as increased grid strain and limited physical space for infrastructure expansion [12].

While EVs introduce new challenges in distribution network planning [13], they also offer considerable opportunities for distribution system operators (DSOs) and charging point operators (CPOs) [14], notably due to their flexibility potential [15]. Understanding EV charging behavior and the flexibility provided by EVSE usage is, therefore, critical from the perspective of grid management. Furthermore, effective coordination between EVs and renewable energy sources must be considered a pivotal aspect of integrating these technologies into sustainable energy systems [16].

1.2. Background

In the context of EV charging, ***flexibility*** refers to the ability to adjust the charging process in various ways to benefit both the EV owner and the electrical grid. This flexibility can be expressed in terms of time (*temporal flexibility*), meaning the ability to shift charging times to periods when the grid has more capacity, or in terms of power (*power flexibility*), meaning the ability to modify the charging rate (kW) based on grid

needs [17]. In both cases, EV charging data are crucial for understanding the available flexibility [18]. Several authors have studied this topic focused on time-series data, including Genov et al. [19], who presented two distinct methods (tree-based and cluster-based) to forecast flexibility, and Babrowski et al. [20] examined how EVs can help balance the electricity system by shifting their charging load to different times of the day. However, these studies often propose complex methodologies with poor replicability in data with different characteristics.

A noteworthy and recurring finding in the literature is that EV charging data consistently reveal distinct patterns in charging behavior, commonly called ***charging profiles***. A charging profile characterizes the typical times of day when charging sessions are more or less frequent, and captures attributes such as session duration (long or short), energy demand (high or low), and flexibility potential. These profiles have been identified through methods such as simulations [21], temporal data analysis [22], and travel surveys [23]. However, clustering techniques, which have received relatively less attention in the literature, represent one of the most effective approaches for identifying these profiles.

For instance, Helmus et al. [24] employed a clustering approach that first used Gaussian Mixture Model (GMM) clustering to group charging sessions, followed by K-Medoids clustering (similar to K-means) to classify portfolios of charging sessions per user. The study considered features such as session start time, connection duration, time intervals between sessions, and the distance between sessions. Shahriar and Al-Ali [25] conducted one of the most interesting analyses found, utilizing K-means, Hierarchical clustering, and GMM to identify similar groups of charging behavior, on real public EV charging activity during the COVID-19 pandemic. Silhouette coefficient, Calinski-Harabasz, and Davies-Bouldin index were the chosen metrics to evaluate the clustering results.

Few European studies have applied clustering techniques specifically to analyze flexibility in EV charging. Bayram et al. [26] examined the first public AC charging sessions in the UK over four months, focusing on utilization rates, arrival and departure times, energy transfer, and overstay durations. The authors employed the DBSCAN algorithm to cluster charging sessions based on arrival and departure times, similar to the approach used by Sadeghianpourhamami et al. [27], who analyzed EV charging sessions in the Netherlands to identify patterns and quantify flexibility, resulting in the identification of three distinct charging clusters. A GMM and K-means analysis of the charging patterns of EVs using an extensive private dataset from Germany, highlighting the potential for flexibility in the charging processes was employed by Märtz et al. [28]. This research is one of the most complete in the literature due to the detailed methodology description and justifications. However, it does not provide a practical representation of flexibility that allows the results to be employed in future studies.

In fact, there is a lack of useful flexibility data for network planning and management studies. For instance, Jerónimo et al. [29] propose a new flexibility model for CPOs that requires EVSE occupancy rates as inputs to the model. These inputs were obtained through simulations. Carvalho et al. [30] study also relied on simulations to obtain typical flexibility profiles due to the deficit of empirical EV flexibility data.

The list of previously mentioned papers confirms interest in understanding the flexibility potential of charging profiles. However, there remain unanswered questions and opportunities for further research. In particular, it is challenging to get universally applicable results with the given charging data and clustering methods. Many reviewed papers do not validate the results, and the techniques employed to identify typical profiles commonly rely on limited data, producing generic clusters and representation of flexibility needlessly intricate and impractical. This is a crucial point since there is a significant lack of empirical flexibility data for demand response and management studies. It is also important to note that most of these (few) studies utilized datasets from countries outside of Europe. Therefore, there is an opportunity to address these gaps. DSOs and CPOs need to develop strategies and acquire knowledge to make informed decisions for the near future, and clustering can be a useful tool to achieve these goals.

1.3. Main Contributions and Paper Organization

This paper aims to fill the gaps previously highlighted, proposing a novel methodology that produces more comprehensive EV charging profiles that provide empirical flexibility data in a visual, intuitive, and relevant approach, with greater applicability to various datasets from different regions. For this purpose, typical EV charging profiles are obtained through clustering methods and designed to be applied in distribution system planning strategies The methodology is tested on empirical EV charging data from the USA (ACN-Data) and validated on European data (GR-Data). GR-Data is a novel private dataset that includes more than 100 000 sessions since 2021 from Greece, while ACN-Data is open-access. The goal is to increase the knowledge about EV charging flexibility, which can create new income for EV users and more flexibility to be managed by DSOs. In particular, the main contributions can be listed as follows:

- A comprehensive and robust methodology based on clustering techniques that can be readily applied to various charging datasets across different regions, particularly when the objective is to identify typical profiles for characterizing EV flexibility flexibility (temporal and power);
- A benchmark analysis of various clustering methods, specifically K-means, GMM, and Hierarchical Clustering, to verify which yields the best profiles for two datasets from different geographical areas (USA and Europe). To achieve this, we present scores (namely Silhouette coefficient, Davies-Bouldin, and Calinski-Harabasz index) and visual representations of the profiles to allow easy comparison of results for future studies;
- Empirical input data are provided for flexibility-based planning studies such as those proposed by Jerónimo et al. [29]. DSOs and CPOs need these data to manage their assets effectively, optimize infrastructure use, reduce congestion, and minimize additional investments.

The paper is organized as follows. Section 2 presents the proposal of methodologies along with a description of the datasets, and the clustering/evaluation methods. Section 3 performs a detailed explanation of the obtained results, summarizing and commenting on the main findings. Finally, Section 4 contains the conclusions and possible future work.

2. Methodology

The overview of the methodological approach is illustrated in Fig. 1, which represents the research flowchart of this work. A description of the datasets' characteristics is done in Section 2.1. The data preprocessing steps are explained in Section 2.2. Section 2.3 describes the main characteristics of clustering, in particular those utilized in this study, and Section 2.4 presents the selected cluster validation techniques.

2.1. *Data Description and Analysis*

There is no cluster analysis without a dataset. Therefore, it is essential to have an adequate EV charging dataset. Amara-Ouali et al. [31] perform an outstanding study of the best EV open data available, providing the community with a structured and carefully selected list of open datasets ready to be used to foster data-driven research in this field. Furthermore, Calearo et al. [32] present a review of data sources for EVs, categorized into different classes by the type of data and its availability.

Based on these papers, an open dataset was found and will be studied: **ACN-Data** [33], from a parking garage available to the public at Caltech University (USA), containing 54 EVSEs with rated 6.656 kW and one 50 kW DC fast charger. At the time of writing, ACN-Data has 31424 EV charging sessions. The first session was in April 2018 and the last was in September 2021.

In addition to open data, this paper had access to private datasets from several European partners in the context of the **EV4EU** project [34]. For this study, the private dataset of public EVSEs in Greece (**GR-Data**) was selected to find EV charging profiles and characterize the flexibility potential. It was collected from public EVSEs in Greece, mainly located in high-traffic and quick-stay areas such as highways, gas stations, supermarkets, and stores. There are a total of 657 EVSEs with registered sessions in the dataset, of which 70 are DC fast-chargers, ranging from 50 kW (15), 60 kW (42), 120 kW (3), 180 kW (6), and even 300 kW (4). The remaining EVSEs have a maximum power of 22 kW. It has a total of 102685 charging sessions from July 2021 to September 2023. Both datasets are in the format *charging event* (1 row of the dataset, 1 EVSE transaction). Table 1 summarizes the datasets' characteristics. Regarding the information in Table 1, the charging duration feature is not present in either dataset despite being one of the most prevalent fields in typical EV charging datasets [31].

Table 1: Summary of the main characteristics of the chosen datasets.

Datasets	**ACN-Data**	**GR-Data**
File Format	JSON file	CSV file
Time Interval	Apr 2018 - Sep 2021	Jul 2021 - Sep 2023
Total Sessions	31 424	102 685
No. of different EVSEs	55	657
EVSE ID and Location	Only Identification	Both
Plug-in/Plug-out Time	Yes	Yes
Start/End Charging Time	Yes	Only Start Time
Charging Duration	No	No
Sojourn Duration	No	Yes
Energy Delivered	Yes	Yes
EVSEs' Max Power	Yes	Yes
User ID	Yes	Yes

2.2. *Stage 1: Data Preprocessing and Cleaning*

According to earlier research [25, 28, 35], data cleaning and preprocessing are two key processes in obtaining interpretable results from cluster analysis.

2.2.1. *Deal with Outliers and Missing Data*

Some datasets' entries might have **missing information**, including the plug-in/plug-out times or energy consumed, for example. Interpolation using nearby entries can be used to replace these absent values. Another possibility would be to remove the datasets' rows corresponding to missing entries, resulting in a dataset with solely accurate and unaltered data. The optimal alternative should be studied and evaluated for each dataset.

There might also be inaccurate information in some entries, such as an abnormal energy supply in a short period. These points, known as **outliers**, should be handled and eliminated using, for instance, techniques like Interquartile Range (IQR) [36], Elliptic Envelope [37], Isolation Forest [38], or by defining thresholds for data removal.

One of the most crucial steps in clustering corresponds to the **normalization** of the data before clustering, especially when working with several fields/features. Clustering algorithms are sensitive to the scale of the data. Normalizing ensures that each entry contributes equally to the distance calculation between data points, helping to improve the accuracy of the clustering. Consequently, each dataset column should range from 0 to 1.

2.2.2. *Feature Engineering*

Another pertinent step involves generating features that serve to enrich the analysis process, facilitating the extraction of more insightful clustering patterns. According to Table 1, the datasets do not contain the same available fields, preventing similar EV charging profiles from being extracted. Therefore, additional features must be created, and two periods can be obtained: the time (t) the EV was parked and plugged into the EVSE (*Sojourn Time*), and the fraction thereof that is effectively spent on charging (*Charging Time*). With these two indicators, the so-called *Idle Time* can be determined, as a measure of **flexibility** of the charging process. These new features can be defined as

$$\text{Sojourn Time} = t^{\textit{plug-out}} - t^{\textit{plug-in}}, \tag{1}$$

$$\text{Charging Time} = t^{\textit{end charging}} - t^{\textit{start charging}}, \tag{2}$$

$$\text{Idle Time} = \text{Sojourn Time} - \text{Charging Time}. \tag{3}$$

ACN-Data contains all the information required for (1) and (2). However, GR-Data does not provide access to $t^{\textit{end charging}}$,

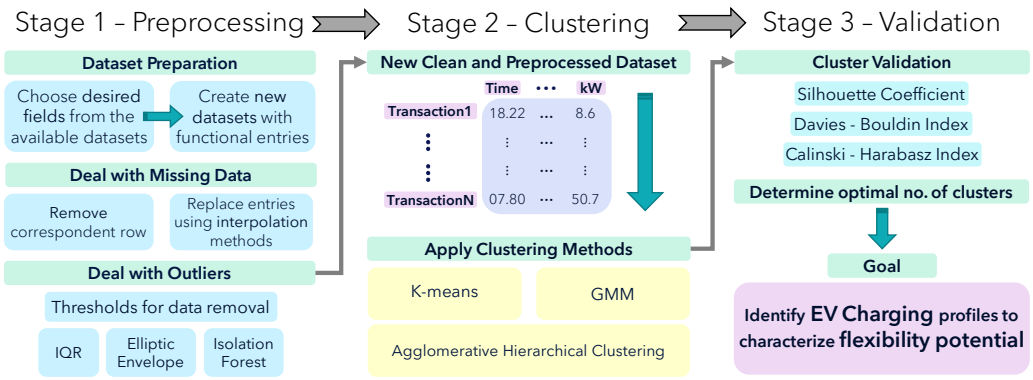


Figure 1: Overview of methodological approach.

and consequently (2) cannot be employed. Instead, it includes the maximum power capacity of the EVSEs. Thus, through (4), it is possible to obtain an average charging time ($AVCT$) value for each session. An adjustment factor (AF) is needed to ensure a more realistic charging time since the process is not performed at a constant power rate; it depends on external factors such as temperature, high grid loads, and the state of charge (SoC) (as the battery becomes fully charged, the charge rate decreases), among others [39]. Therefore, an analysis was performed using ACN-Data to identify the appropriate adjustment factor, as the charging time was already available from expression (1). This analysis involved comparing the energy delivered with the (EVSE max power × charging time). The calculation of this ratio for all ACN-Data sessions resulted in approximately 0.8. This value was then adopted as the adjustment factor in expression (4) to ensure consistency between the two datasets.

$$AVCT_{session\,i} = \frac{Energy\ Delivered_i}{(EVSE\ max\ Power)_i \times AF} \tag{4}$$

2.2.3. Characterization of EV Flexibility

This paper aims to analyze and understand the flexibility in EV charging. The discussion will cover two aspects: temporal and power flexibility.

Temporal flexibility refers to the idle time determined through (3), which indicates the available periods to shift the charging process. On the other hand, **power flexibility** refers to the capability of reducing the charging power to achieve the desired flexibility at a given period. It is obtained based on the typical values for energy delivered, sojourn, charging, and idle times of each cluster. It is determined as follows:

- If the desired flexibility (P^{flex}_{desire}) is less than the idle time ($idleTime$), the power flexibility (P^{flex}) equals the average charging power (μP^{ch}) since the charging process can be shifted without any loss on energy delivery, as represented by (5).

$$P^{flex}_{session\,i} = \mu P^{ch}_i,\ \forall P^{flex}_{desire\,i} \leq idleTime_i \tag{5}$$

- If the desired flexibility exceeds the idle time, the power flexibility is the difference between the average charging power and the minimum power required to guarantee the desired energy (P^{target}), represented by (6). It is important to highlight that P^{target} depends on each EV session.

$$P^{flex}_{session\,i} = \mu P^{ch}_i - P^{target}_i,\ \forall P^{flex}_{desire\,i} > idleTime_i \tag{6}$$

Additionally, the flexibility cannot exceed the corresponding cluster sojourn time. Sections 3.1.7 and 3.2.7 provide a detailed explanation of results to enhance comprehension of this topic.

2.3. *Stage 2: Selected Clustering Methods*

Three well-known clustering methods are the choice for identifying groups of similar charging patterns: **K-means**, **GMM**, and **Hierarchical clustering**. These methods are frequently employed in applications related to charging behavior, as previously mentioned in Section 1.2. Moreover, Al-Ogaili et al. [40] and Shahriar et al. [41] precisely suggest the use of these methods for the analysis of EV charging patterns (these studies provide a comprehensive overview of machine learning (ML) techniques applied in EV and EVSE deployment data). Nevertheless, it is important to briefly introduce clustering.

Cluster analysis, often known as **clustering**, is not a specific algorithm, but rather the general problem of partitioning a dataset into natural subgroups called **clusters** [42]. Objects within the same group should be as similar as possible (based on a similarity measure), while objects between different groups should be as dissimilar as possible. Clustering uses almost no information to evaluate the data and does not require a separate training set to determine the model parameters (unsupervised learning approach). It is the main objective of exploratory data analysis, a popular statistical analysis technique applied in various domains, (e.g., image analysis, bioinformatics, and ML).

Due to the absence of a definitive definition for the term "cluster," numerous methods for distinct strategies have been developed. In this work, the notation and nomenclature follow the ones defined by Zaki and Meira [42]. The following subsections introduce the clustering methods used in this article. For a complete explanation of these and further methods, see [42].

2.3.1. Representative-based clustering

Representative-based clustering aims to divide a dataset into k clusters. Each cluster is characterized by a representative point (called **centroid**), commonly chosen as

the mean of within-cluster points. The K-means and Expectation-Maximization (EM) algorithms are examples of representative-based clustering approaches:

- K-means [43] is a greedy technique that minimizes the squared distance between points and their corresponding cluster means. It also conducts hard clustering, meaning that each point is assigned to only one cluster;
- EM [44] generalizes K-means by modeling the data as a mixture of normal distributions and maximizing the likelihood of the data to find the cluster parameters (mean and covariance matrix). It conducts soft clustering since it returns the probability of a point belonging to each cluster. EM is the algorithm utilized by the GMM method.

2.3.2. Hierarchical Clustering

Hierarchical clustering creates a sequence of nested partitions, which can be visualized as a tree, also called *dendrogram*, indicating the merging process and the intermediate clusters. The highest level (root) consists of all points in one single cluster, whereas the lowest level (leaves) consists of clusters of individual points, each point in its cluster.

In this paper, the **Agglomerative** algorithmic approach is the choice [45]. It starts with the points as individual clusters and, at each step, merges (or agglomerates) the most similar or closest pair of clusters until the desired number of clusters has been found. This requires a definition of cluster similarity and, for that, a variety of measures can be employed, including **single link**, **complete link**, **average link**, or **Ward's method** [42].

2.4. *Stage 3: Clustering Validation Techniques*

Since no ground truth is available, internal validation should be used to quantify the performance of the clustering [42]. Three internal validation metrics, Silhouette coefficient [46], Davies-Bouldin index [47], and Calinski-Harabasz index [48] can be employed, based on the studies reviewed in Section 1.2.

2.4.1. Silhouette Coefficient

For each point $\boldsymbol{x}_i$, the silhouette coefficient is

$$s_i = \frac{\mu_{out}^{min}(\boldsymbol{x}_i) - \mu_{in}(\boldsymbol{x}_i)}{\max\{\mu_{out}^{min}(\boldsymbol{x}_i), \mu_{in}(\boldsymbol{x}_i)\}}, \tag{7}$$

where $\mu_{out}^{min}(\boldsymbol{x}_i)$ is the mean of the distances from $\boldsymbol{x}_i$ to points in the closest cluster, and $\mu_{in}(\boldsymbol{x}_i)$ is the mean distance from $\boldsymbol{x}_i$ to points in its own cluster. The total **Silhouette coefficient** [46] is defined as the mean s_i value across all points, given by (8), where a value close to +1 denotes good clustering.

$$SC = \frac{1}{n}\sum_{i=1}^{n} s_i \tag{8}$$

2.4.2. Davies-Bouldin Index

The Davies-Bouldin measure for a pair of clusters C_i and C_j is defined as

$$DB_{ij} = \frac{\sigma_{\mu_i} + \sigma_{\mu_j}}{\delta(\mu_i, \mu_j)}, \tag{9}$$

where μ_i denotes the centroid of cluster C_i, $\sigma_{\mu_i} = \sqrt{var(C_i)}$ represents the dispersion of the points around the respective centroid (square root of the total variance) and $\delta(\mu_i, \mu_j)$ is the distance between the centroids.

The **Davies-Bouldin index** [47] is thus defined as

$$DB = \frac{1}{k} \cdot \sum_{i=1}^{k} \max_{i \neq j}\{DB_{ij}\}, \tag{10}$$

meaning that for each cluster C_i it is chosen the cluster C_j that returns the largest DB_{ij} ratio. Therefore, smaller DB values mean better clustering (clusters are well separated and each one is well represented by its centroid).

2.4.3. Calinski-Harabasz Index

The **Calinski-Harabasz index** [48] is given by

$$CH(k) = \frac{tr(\boldsymbol{S}_B)}{tr(\boldsymbol{S}_W)} \cdot \frac{n-k}{k-1}, \tag{11}$$

where $tr(\boldsymbol{S}_B)$ is the trace of the within-cluster scatter matrix, $tr(\boldsymbol{S}_W)$ is the trace of the between-cluster scatter matrix.

For a good value k (number of clusters), it should result in a high CH value. This way, the Calinski-Harabasz index can be also used to choose the number of clusters that maximize $CH(k)$, an alternative to the elbow method [49].

3. Evaluation of EV Flexibility using Clustering Methods

In this section, the evaluation of EV flexibility using clustering methods is described and discussed. It includes two main subsections, focusing on the data preprocessing steps and the presentation of results from the two datasets analyzed: ACN-Data and GR-Data. In each subsection, the fundamental aspects of the applied methodology are critically and concisely discussed.

The code was written in Python using the Google Colab platform, and the *scikit-learn* library [50] for the preprocessing, clustering, and evaluation methods (most parameters were left at default, while those modified are mentioned throughout the text).

3.1. *ACN-Data: Data preprocessing and Results*

This section is dedicated to discussing the results related to the ACN-Data. First, it is necessary to conduct the data preprocessing steps, which consist of preparing the dataset, dealing with missing data, detecting outliers, and adjusting the data. The main results focus on EV charging profiles and the evaluation of EV flexibility. The following subsections provide more details on the data preprocessing and the results.

3.1.1. Dataset preparation

The first step in obtaining EV charging profiles is data preprocessing, according to the schematic in Fig. 1. Since the dataset is provided in a JSON file, various conversions were necessary to obtain each field in the required format, especially with the help of *Pandas* library to obtain the fields *sojournTime*, *chargingTime*, and *idleTime* based on (1), (2) and (3), respectively. After determining these extra fields, the entries in DateTime format needed to be converted into a suitable numeric structure: for example, 10h17 (10 hours and 17 minutes) becomes 10.28h (10.28 hours) in float format, consequently allowing full use of outlier removal approaches, clustering methods, and graphical representations.

3.1.2. Deal with Missing Data

After analyzing the preprocessed dataset, *endChargingTime* was occasionally missing in some entries, indicating that the charging time was insufficient to obtain a fully charged battery. Thus, this field was assigned with the *disconnectTime* (plug-out) entry in these sessions, leading to an idle time of zero. Regarding the *userID* field, the lack of this information makes it impossible to discover or predict the user corresponding to the session with total certainty.

3.1.3. Outlier Detection

With a fully functional dataset, the next step involved setting thresholds to remove unwanted data. A limit was defined to remove the sessions with a *sojournTime* or *chargingTime* greater than 48 hours and less than 1 minute. Another threshold was set to clear sessions with energy-delivered values greater than 90 kWh, selected considering the characteristics of EVs available on the market during the period of the data (2018-2021). Also eliminated were all null/negative entries and specific cases where there was a higher amount of energy than the maximum value allowed by the EVSE during charging.

Fig. 2 illustrates the distribution of the clean data. There are roughly three main groups: one at the beginning of the day, from 00h00 to 03h00, with scattered sojourn times; another from 06h00 to the end of the day, with longer sojourn times when connecting in the morning; and finally, between 17h00 and 23h59, with higher sojourn times.

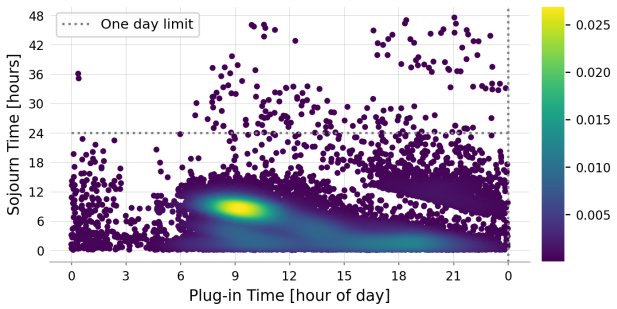


Figure 2: Clean ACN-Data distribution regarding Sojourn and Plug-in Time.

Another interesting fact is that some points are more scattered from the three main groups identified previously. These points are the so-called **outliers**. However, when analyzing the data, one can see that these points represent behavior that could have happened and are not errors in the data, even though they are distant from most sessions. The grosser errors, effectively outliers, have already been identified and eliminated when defining the thresholds for the target fields *chargingTime*, *sojournTime*, and energy-delivered. Therefore, no outlier removal method was employed for the ACN-Data dataset.

3.1.4. Data Adjustment

The plug-in time with date and hours became plug-in time with only the hour of the day when the DateTime variables were converted into float values. The drawback of this strategy is that the time frame under consideration was 00h00 to 23h59. Due to their loss of spatial proximity, early and late plug-in times might be clustered separately. As displayed in Fig. 2, there is an instant when charging activity is at its lowest, reaching it around 03h00. To restore the spatial proximity, all charging sessions with plug-in times less than this minimum were relocated to the right side to continue the timeframe after 23h59. The final available fields are represented in Table 2.

Table 2: Summary of the final usable fields in the ACN-Data dataset.

Field name	**Non-Null count**	**Dtype**
connectionTime (Plug-in)	31318	float64
disconnectTime (Plug-out)	31318	float64
endChargingTime	31318	float64
kWhDelivered	31318	float64
EVSE_ID	31318	int64
userID	16355	int64
chargingTime	31318	float64
sojournTime	31318	float64
idleTime	31318	float64

3.1.5. Chosen fields and normalization of the data

Following Shahriar and Al-Ali's article [25] previously mentioned, it became clear that an in-depth study was needed on the choice of fields for clustering. Thus, by analyzing the covariance matrix between the available features, interesting patterns arise. The highly correlated *connectionTime*, *disconnectTime*, and *endChargingTime* fields are redundant, so only one is necessary (*connectionTime* provides intelligible information, and thus it must be chosen). The same reasoning applies to *kWhDelivered* and *chargingTime*. Additionally, *connectionTime* exhibits an inverse relationship with both *sojournTime* and *idleTime*; thus, selecting one is appropriate. Ultimately, the choice became *connectionTime*, *sojournTime*, and *kWhDelivered* fields since this triplet yielded the best results in the first cluster analysis. The remaining fields were eliminated, and the data were normalized to obtain the best possible results, described next.

3.1.6. ***Results: EV Charging profiles***

The number of clusters, k, was chosen based on the scores of Silhouette, Davies-Bouldin, and Calinski-Harabasz. Additionally, it was possible to use the elbow method with the K-means

and the *dendrograms* with Hierarchical clustering to get an initial idea of the most suitable number of clusters.

Besides k, for GMM and Hierarchical clustering, it was also necessary to tune parameters to obtain the best possible scores: for GMM clustering, the types of covariance from *full*, *tied*, *diagonal*, and *spherical*; for Agglomerative Hierarchical clustering, the distance measure (recall Section 2.3.2). The remaining parameters were left default. The optimum number of clusters for each method was chosen according to the k that simultaneously leads to the best scores and the most interpretable and meaningful results. Table 3 summarizes the optimal scores obtained for each clustering method.

Table 3: Summary of the selected metrics for each ACN-Data clustering method.

	K-means	**GMM**	**Hierarchical**
Best k	8	8	6
Parameters	-	Tied Covariance	Ward's Method
Elbow Method	k={5, 6, 7, 8}	-	-
Silhouette	0.329	0.313	0.325
Davies-Bouldin	1.006	1.007	1.097
Calinski-Harabasz	17561.08	15226.62	15496.63

Therefore, by analyzing Table 3 and the obtained profiles, the K-means method produced the best results, which will be examined in greater detail. Fig. 3 presents the distribution of the adjusted EV charging profiles regarding the Plug-in Time, Sojourn Time, and kWh (energy delivered) fields.

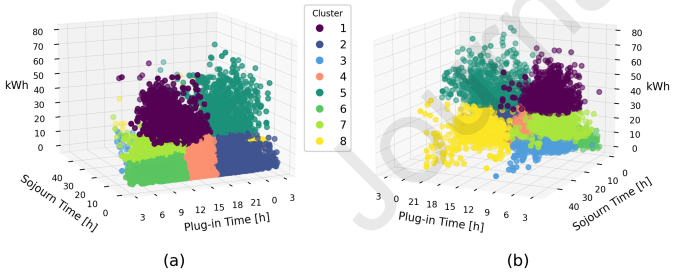


Figure 3: 3D distribution of the adjusted K-means EV Charging profiles for the ACN-Data dataset. (a) Azimuth = -115°. (b) Azimuth = 115°.

From Fig. 3, one can see that the profiles are relatively well-defined and have little overlap. An intriguing result that is immediately apparent is the separation of the high consumption profiles (clusters 1 and 5), which are virtually divided by the plane defined by kWh $\approx$ 30, from the low and medium consumption profiles (clusters 2, 3, 4, 6, 7, and 8).

Additionally, it is noticeable that there are more short-term sessions, which impacts the number of profiles, with the longer sojourn times comprised in clusters 3 and 8. Table 4 lists the quantitative characteristics of the eight profiles. It is important to note that Table 4 presents mean values for the clusters' characteristics, where the profiles are defined as *Low energy*: below 10 kWh; *Medium energy*: between 10 kWh and 30 kWh; *High energy*: over 30 kWh; *Short-term*: sojourn time below 2h; *Medium-term*: between 2h and 4h; *Long-term*: over 4h.

Examining Table 4, one can see that cluster 5 behaves slightly differently from the others, with a plug-out time close to 05h00 and around 600 sessions, indicating that this profile is the least common. A deeper analysis revealed that roughly half of the cluster's sessions start and end on the same day (late afternoon). The remaining sessions only end the next day, with a higher incidence in the morning, suggesting that EVs stay connected to the EVSE during the night. Since cluster 5 comprises two distinct behaviors, the average plug-out time does not fully reflect all sessions.

3.1.7. *Results: Evaluation of EV Flexibility*

The main results of the evaluation of EV flexibility are discussed based on temporal and power flexibility. It builds on the EV charging profile results, starting with the **temporal flexibility** potential of each cluster. Specifically, cluster 3 exhibits a mean idle time that surpasses the mean charging time, suggesting that the EVs spend more time parked without charging than actually charging (recall Table 4). This indicates a high flexibility potential. In general, sessions with shorter sojourn times also present less potential for flexibility. However, most profiles offer great opportunities, with high idle times at different moments of the day. This behavior is in line with the location of the EVSEs (Caltech University). The most representative clusters (clusters 3, 4, and 7) precisely demonstrate typical workplace behavior: EVs connected in the morning/early afternoon and unplugged at the end of the working day, with high flexibility (refer to Table 4). Fig. 4 illustrates the temporal flexibility characterization of the typical profiles found, in a lattice format according to Table 4.

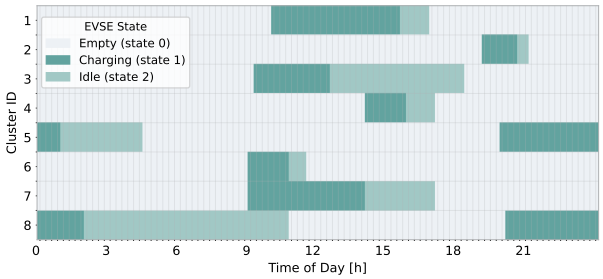


Figure 4: Temporal flexibility characterization of the EV Charging profiles for the ACN-Data dataset.

As previously mentioned, flexibility can also represent the capability of reducing the charging power to achieve the desired flexibility at a given time for reasons such as coordinating the charging process with renewable energy sources, avoiding power surges, or balancing the grid, for example. This **power flexibility** is especially important for peak shaving and smart charging methods, which are increasingly important as EVs grow in popularity [51]. Fig. 5 displays the normalized power reduction capacity per cluster as a function of the desired flexibility.

For instance, in cluster 1, the idle time is 1 hour and 10 minutes (recall Table 4), indicating that the charging process can be

Table 4: Mean quantitative characteristics of the K-means EV Charging profiles for the ACN-Data dataset.

Cluster ID	No. of Sessions	Plug-in Time	Plug-out Time	Energy [kWh]	Sojourn Time	Charging Time	Idle Time	Profile
1	1174	10h12	16h51	34.735	6h 38min	5h 28min	1h 10min	Morning to afternoon, high energy, long-term stay
2	5420	19h14	21h05	7.215	1h 51min	1h 28min	22min	Evening short-term stay, low energy
3	6305	09h30	18h22	4.765	8h 52min	3h 12min	5h 41min	Morning to afternoon, low energy, long-term stay
4	6588	14h05	17h11	6.165	3h 06min	1h 52min	1h 15min	Afternoon medium-term stay, low energy
5	609	19h51	04h33	39.390	8h 42min	5h 16min	3h 26min	Evening to next morning, high energy, long-term stay
6	4671	09h15	11h43	4.987	2h 28min	1h 43min	45min	Morning medium-term stay, low energy
7	5399	09h12	17h11	14.322	7h 59min	4h 55min	3h 05min	Morning to afternoon long-term stay, medium energy
8	1152	19h51	10h49	12.777	14h 58min	6h 01min	8h 57min	Evening to next morning, medium energy, long-term

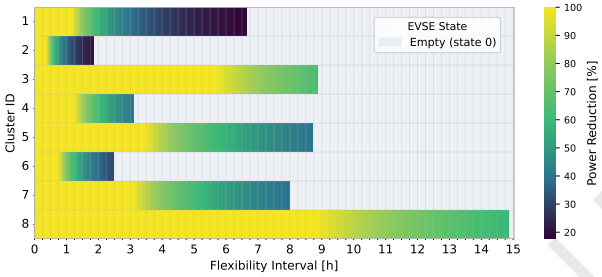


Figure 5: Normalized power flexibility characterization of the EV Charging profiles for the ACN-Data dataset.

shifted by that period without requiring any adjustment to the charging power. This provides 100% flexibility, meaning that the charging power might have a reduction of 100% compared to the mean charging time of the cluster. However, if more flexibility is required, the charging power cannot be reduced by 100% to ensure that the requested energy is delivered to the EV at the end of the sojourn time. For instance, if the desired flexibility is 2 hours, the charging time decreases from 5 hours and 28 minutes to 4 hours and 38 minutes. In this reduced charging period, the EV can only recharge about 29 kWh, leaving a shortfall of around 6 kWh. Dividing the 6 kWh by the 2 hours gives an approximate power of 3 kW, corresponding to around 58% of the mean charging power verified for cluster 1. Fig. 6 provides a visual representation of the previous example.

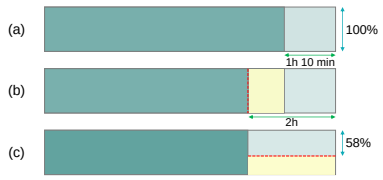


Figure 6: Visualisation of a two-hour flexibility example for cluster 1 of ACN-Data. (a) Typical behavior of the cluster. (b) With 2 hours of flexibility, amount of energy missing in yellow. (c) Final behavior with 2 hours of flexibility, with the corresponding reduction in power.

When the required flexibility matches the sojourn time of 6 hours and 38 minutes, the charging power can only be reduced by approximately 18% (the flexibility cannot exceed the corresponding cluster sojourn time, as expected). To better interpret the results, Fig. 7 reveals the power flexibility in absolute values per each cluster session, and Fig. 8 illustrates the power flexibility considering the total number of sessions.

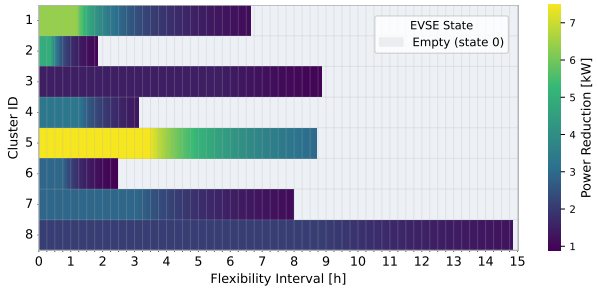


Figure 7: Power flexibility characterization of the EV Charging profiles for the ACN-Data dataset per cluster session.

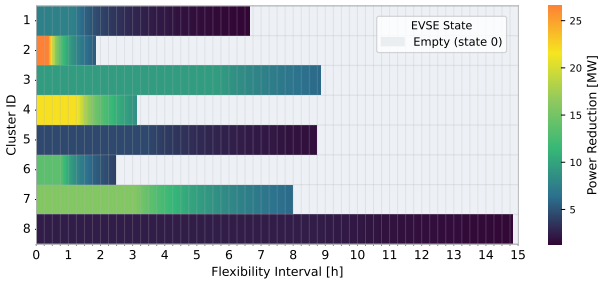


Figure 8: Power flexibility characterization of the EV Charging profiles for the ACN-Data dataset, considering the total number of sessions per cluster.

3.2. *GR-Data: Data preprocessing and results*

This section is devoted to the discussion of the results related to the GR-data. Similar to subsection 3.1, it is first essential

to address the data preprocessing steps, which involve dataset preparation, dealing with missing data, outlier detection, and data adjustments, and the main results focus on EV charging profiles and the assessment of EV flexibility. Further details on data preprocessing and the results are provided in the following subsections for GR-Data.

3.2.1. Dataset preparation

Since the GR-Data dataset had the same format as the ACN-Data, the steps followed for the conversion of the entries were identical, only changing the fields' names according to the GR-Data's specific characteristics. With the *averageChargingTime* created through (4) and the *sojournTime* present in the dataset, it was then possible to obtain the *idleTime* through (3). Finally, the DateTime fields were also converted to float values.

3.2.2. Deal with Missing Data

After analyzing the preprocessed dataset, there were no missing entries in the *Start_datetime*, *End_datetime*, *EVSE_ID*, or *userID* fields. However, some sessions were missing the EVSE Max Power entry, making it impossible to determine the average charging time. Thus, these sessions were discarded.

Additionally, some sessions presented an average charging time higher than the sojourn time, indicating that the vehicle was effectively charging during the entire parking period and that the adjustment factor was too harsh for these particular sessions. Accordingly, the *averageChargingTime* was assigned with the value of the *sojournTime* entry in these sessions, leading to a corresponding idle time of zero.

3.2.3. Outlier Detection

The defined thresholds match those specified for the previous dataset, with slight differences: 24-hour charging time and sojourn time limit, only sessions with more than 1 minute of sojourn time, and maximum energy delivered of 100 kWh (considering the 2021-2023 EV sales in Europe). All null or negative entries were also removed. There are only 137 sessions with more than 24 hours of parking stay in the dataset. Consequently, the clustering results were improved by removing these sessions, yielding more meaningful clusters. All null/negative entries and specific cases involving excessive energy during charging beyond the EVSE's maximum capacity were also removed.

3.2.4. Data Adjustment

To restore the spatial proximity, all charging sessions with plug-in times less than 04h00 (instant of minimum charging activity) were relocated to the right side to continue the timeframe after 23h59. The final clean and preprocessed dataset is illustrated in Fig. 9 regarding the instant of plug-in (Plug-in Time) and the duration of parking stay (Sojourn Time). Table 5 contains all the usable fields from the final preprocessed GR-Data dataset.

Table 5: Summary of the final usable fields in the GR-Data dataset.

Field name	Non-Null count	Dtype
Start_datetime (Plug-in)	95759	float64
End_datetime (Plug-out)	95759	float64
kWhDelivered	95759	float64
EVSE_ID	95759	object
userID	95759	object
maxPowerEVSE	95759	float64
sojournTime	95759	float64
averageChargingTime	95759	float64
idleTime	95759	float64

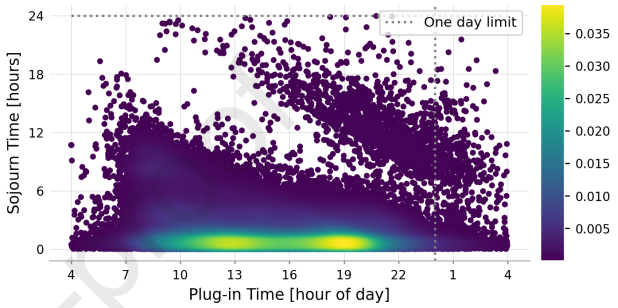


Figure 9: Final adjusted GR-Data distribution depending on Sojourn Time and Plug-in Time.

3.2.5. Chosen fields and normalization of the data

A similar breakdown described in Section 3.1.5 was performed with the private dataset GR-Data, yielding highly similar results. Consequently, the chosen fields were *Start_datetime*, *sojournTime*, and *kWhDelivered*, allowing a comparable analysis between the profiles found in both datasets. The remaining features were removed, and the data were normalized to obtain the best possible outcomes, detailed next.

3.2.6. *Results: EV Charging profiles*

Following the method described in Section 3.1.6 for the ACN-Data dataset, it was possible to obtain Table 6, which summarizes the optimal scores obtained and the specific characteristics chosen for each clustering method.

Table 6: Summary of the selected metrics for each GR-Data clustering method.

	K-means	GMM	Hierarchical
Best k	10	8	7
Parameters	-	Tied Covariance	Ward's Method
Elbow Method	k={7, 8, 9, 10}	-	-
Silhouette	0.314	0.305	0.288
Davies-Bouldin	0.989	1.002	1.029
Calinski-Harabasz	48149.50	45802.46	42924.90

Therefore, by analyzing Table 6 and the resulting profiles, the K-means method produced the best results, which will be examined in greater detail. Fig. 10 presents the distribution of the adjusted EV charging profiles regarding the Plug-in Time, Sojourn Time, and kWh (energy delivered) fields.

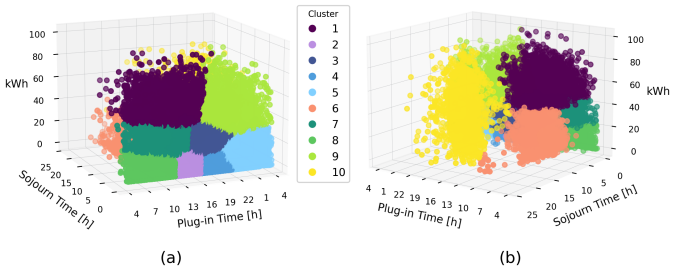


Figure 10: 3D distribution of the adjusted K-means EV Charging profiles for the GR-Data dataset. (a) Azimuth = -115°. (b) Azimuth = 125°.

By observing Fig. 10, one can see that the results are relatively similar to those obtained for the ACN-Data dataset, apart from the higher number of sessions, clearly visible in the schematic. There is, however, a greater separation between sessions for morning/early afternoon plug-in times, with five clusters with morning plug-in times (clusters 1, 6, 7, and 8) and only three in the late afternoon/early evening (clusters 5, 9, and 10). The reduced number of profiles in the evening demonstrates that the sessions during this period have a more similar behavior than the sessions during the day.

There are also middle/late afternoon profiles (clusters 2, 3, and 4), which indicates that, in this dataset, the sessions throughout the day are considerably more different from each other than the ones observed in ACN-Data, translating into a higher number of daily profiles.

Table 7 lists the mean quantitative characteristics of the ten profiles, confirming that clusters 2 and 4 are the most typical profiles, as they comprise the highest number of sessions. This means that the short, low-energy, and low-flexibility potential sessions are the most frequent, occurring at lunchtime and after work. Additionally, one can see that the later the drivers plug in, the more energy they consume. Morning (cluster 8) and afternoon (cluster 4) profiles are generally characterized by lower energy delivered. In fact, compared to ACN-Data, most Greek profiles exhibit shorter charging times due to the higher power EVSEs and, consequently, shorter sojourn times.

3.2.7. *Results: Evaluation of EV Flexibility*

The main results of the evaluation of EV flexibility are discussed based on temporal and power flexibility. It begins with an analysis of the **temporal flexibility** potential for each cluster based on the results of the EV charging profiles. Specifically, analyzing Table 7, one can verify that the most different sessions (with higher sojourn times and, consequently, higher flexibility potential) fall into distinct clusters, namely clusters 6 and 10. Cluster 10 contains the sessions that only end the next day, regardless of the plug-in time, while cluster 6 comprises the sessions that start in the morning and only end in the afternoon of the same day. These two profiles provide the highest flexibility potential since the remaining clusters represent typical charging at quick-stay locations such as supermarkets, highways, or gas stations, precisely where most of Greece's public EVSEs are located.

The typical profiles can thus be employed to provide empirical flexibility data for various future investigations. For instance, Jerónimo et al. study [29] could be adapted to benefit from this empirical data rather than resorting to simulation algorithms. Fig. 11 illustrates the temporal characterization of the typical profiles found, in a lattice format, similar to the representation in [29].

Despite shorter sojourn times compared to ACN-Data, Fig. 11 reveals that DSOs and CPOs still have ample opportunity to utilize existing flexibility, particularly during morning, afternoon, and evening periods, according to the behavior of clusters 8, 6, and 10, respectively. Furthermore, clusters 5 and 9 provide additional flexibility in the early evening hours, in contrast to ACN-Data results (remember Fig. 4).

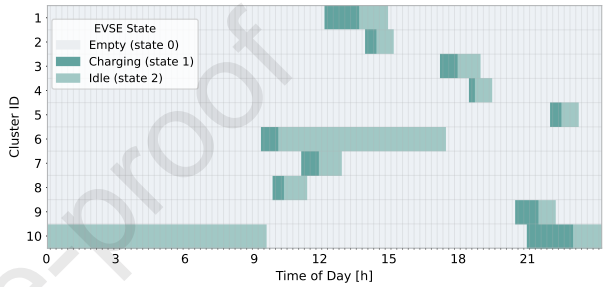


Figure 11: Temporal flexibility characterization of the EV Charging profiles for the GR-Data dataset.

Regarding **power flexibility**, Fig. 12 represents the normalized power reduction capacity per cluster based on the desired flexibility. As the charging times are short, it is natural to have high flexibility during the brief sojourn time available in each cluster, particularly in cluster 6. Due to the fast charging of EVSEs, the flexibility of cluster 6 remains high even during extended periods, and, remarkably, charging power can be reduced by up to 91% without compromising energy delivery within the sojourn time. Cluster 10 exhibits a similar performance. On the other hand, cluster 9 has the lowest capacity for reducing charging power — it can only lower charging power by 37% for flexibility equal to the sojourn time of 1 hour and 45 minutes. This is explained in Table 7, where one can verify that cluster 9 (and cluster 1) includes more prolonged charging times than idle times.

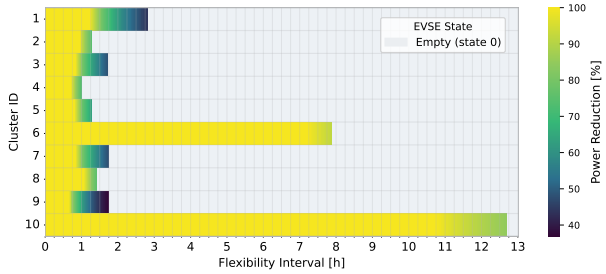


Figure 12: Normalized power flexibility characterization of the EV Charging profiles for the GR-Data dataset.

Fig. 13 reveals the power flexibility in absolute values per each cluster session, and Fig. 14 illustrates the power flexibility considering the total number of sessions. A notable observation is that clusters 2 and 4 exhibit the highest absolute power

Table 7: Mean quantitative characteristics of the K-means EV Charging profiles for the GR-Data dataset.

Cluster ID	No. of Sessions	Plug-in Time	Plug-out Time	Energy [kWh]	Sojourn Time	Charging Time	Idle Time	Profile
1	4636	12h10	14h59	52.968	2h 49min	1h 35min	1h 14min	Morning to afternoon high energy, medium-term stay
2	20071	13h57	15h13	6.761	1h 16min	20min	56min	Early afternoon low energy, short-term stay
3	7883	17h07	18h49	26.768	1h 43min	53min	49min	Afternoon to evening medium energy, short-term stay
4	19993	18h22	19h21	5.754	59min	16min	43min	Early evening low energy, short-term stay
5	8635	21h57	23h13	12.128	1h 16min	28min	48min	Night low energy, short-term stay
6	5520	9h27	17h20	12.229	7h 54min	42min	7h 12min	Morning to afternoon medium energy, long-term
7	7196	11h05	12h49	28.083	1h 44min	54min	49min	Morning to early afternoon medium energy, short-term
8	16090	9h57	11h21	7.346	1h 25min	21min	1h 03min	Morning low energy, short-term stay
9	4068	20h18	22h03	46.850	1h 45min	1h 06min	38min	Evening to night high energy, short-term stay
10	1667	20h56	9h36	34.598	12h 41min	1h 58min	10h 43min	Evening to next morning medium energy, long-term

reduction capacity despite delivering less energy (as noted in Table 7). Cluster 8, typical of morning charging, also displays a high absolute power reduction capacity, which could contribute to helping balance the electrical grid during this period. Therefore, there is significant potential for reducing power consumption in these clusters, even for a short duration.

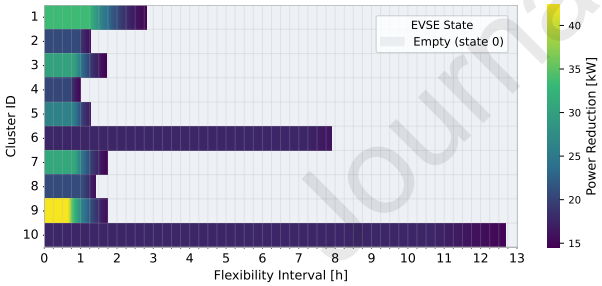


Figure 13: Power flexibility characterization of the EV Charging profiles for the GR-Data dataset per cluster session.

Another intriguing finding relates to profiles 6 and 10. Cluster 10 is found to be highly time-flexible, but it has limited absolute power reduction capacity. This happens because it includes a smaller number of sessions, which makes it less representative. In contrast, cluster 6 may deliver less energy, but it compensates for it by having more sessions, which results in a greater absolute power reduction capacity.

3.3. Knowledge and critical analysis of the obtained results

Identifying the optimum number of clusters (along with the most appropriate type of covariance and distance metric) was complex and time-consuming. The aim was to find typical and meaningful profiles, which required a more in-depth analysis than just selecting the parameters that produced the best Silhouette, Davies-Bouldin, or Calinski-Harabasz scores. It was necessary to consider the meaningfulness of the results as well.

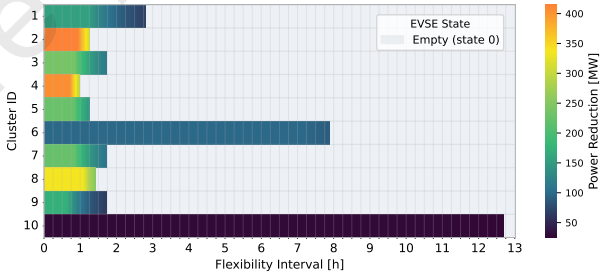


Figure 14: Power flexibility characterization of the EV Charging profiles for the GR-Data dataset, considering the total number of sessions per cluster.

The EV charging profiles previously mentioned provide valuable information about flexibility. The analysis of this information can be helpful in future projects, particularly in the collaboration of EVs with solar and wind renewable energies. For instance, according to the ACN-Data dataset (recall Fig. 4), there is no flexibility during evening peak hours (between 19h00 and 21h00). However, there is significant flexibility during the day due to high sojourn times and extended charging times, in line with the location of the EVSEs. This flexibility can be utilized to reduce charging power during high demand times to coordinate with solar renewable energy sources, and thanks to clusters 5 and 10, there is an opportunity to assist with wind curtailment which mainly occurs at night [52]. However, the low maximum power of the chargers may limit the level of flexibility that can be achieved.

Regarding the GR-Data dataset, the temporal flexibility of the EV charging profiles is limited (remember Fig. 11). As publicly operated EVSEs, the most typical usage involves fast, short-term sessions with low energy delivered. Yet, due to the large number of profiles, there is always some flexibility during strategic hours of the day, which allows EVs to help balance and coordinate the electrical grid with solar energy sources. A

long-term night-time profile is also available (cluster 10), opening the possibility to help with wind curtailment (although not very representative, as seen in Table 7). Due to the high power capabilities of Greek public EVSEs, short idle times can still result in a potentially significant reduction of the aggregated charging power, which is a positive achievement for the DSOs and CPOs management.

In both ACN-Data and GR-Data, there is no flexibility available during peak times. Yet, this lack of flexibility mainly affects late-evening and early-night periods. The GR-Data dataset could be filtered to only include sessions from specific cities or strategic areas to explore potential local flexibility, as it combines data from EVSEs across Greece.

An important consideration when analyzing charging profiles on AC chargers is the characteristics of EVs' onboard chargers. The onboard charger determines the maximum AC charging rate that an EV can handle, which affects both the charging duration and behavior. If two EVs are connected to the same EVSE but have onboard chargers with different power ratings, they can exhibit different charging profiles. Therefore, it is valuable to include these vehicle-specific factors when studying charging patterns. However, the lack of detailed information about EVs' onboard chargers prevented the inclusion of this data in the study presented in this paper. Future research could greatly benefit from incorporating this information to better differentiate between charging behaviors driven by EV characteristics and those resulting from user-specific patterns.

Nonetheless, the study presented in this paper demonstrates that clustering can be a powerful tool to extract valuable and practical information when applied to EV charging data.

4. Conclusions and Future Work

This paper proposes a three-stage method to evaluate various clustering techniques for identifying electric vehicle (EV) charging profiles, with a particular emphasis on usage flexibility. In particular, typical profiles were found by applying clustering methods to datasets of empirical charging processes, including ACN-Data (an open dataset) and GR-Data (a private dataset from the European project EV4EU: Electric Vehicle Management for Carbon Neutrality in Europe).

The experimental results demonstrated the effectiveness of clustering techniques in extracting comprehensive insights into the EV charging process, confirming the methodology's adaptability across different datasets. It includes a benchmark analysis of various clustering techniques to identify the most effective approach for profiling, including K-means, Gaussian Mixture Model (GMM), and Hierarchical clustering. The analysis was validated using the Silhouette coefficient, Davies-Bouldin index, and Calinski-Harabasz index. Among the parameters tested, tied covariance for GMM and Ward's method for Hierarchical clustering consistently yielded the most effective results. However, K-means ultimately produced the best outcomes across both datasets, achieving a good balance between meaningful profiles and high relevance scores. The resulting profiles highlight the potential of EVs as a crucial tool due to their charging flexibility. Specifically, ACN-Data is characterized by highly flexible profiles, as most EVs spend more time parked than actively charging, typically during a standard workday. In contrast, GR-Data predominantly features quick-stay sessions, reflecting the nature of the electric vehicle supply equipment (EVSE) locations, which are publicly accessible infrastructures. Still, due to the high number of clusters and short charging durations, flexibility is available at critical times throughout the day. The findings of this study aim to assist distribution system operators (DSOs) and charging point operators (CPOs) in the successful and intelligent integration of EVs into the energy system by providing valuable empirical information on EV charging behavior and associated usage flexibility.

Future work may include further clustering studies using newly available datasets from various regions/countries, covering different fields selected according to a defined goal to increase knowledge about EVs and EVSEs. The choice of fields and the ultimate objective will determine the nature of the results obtained, with multiple possibilities yet to be explored. Another avenue for future work could involve studies using regional EV data to achieve a more specific and targeted understanding of flexibility, in contrast to the broader scope of this research. Finally, integrating the methodology presented in this article with previous studies that relied on simulated data creates a giant opportunity for understanding and guiding a sustainable future that we aspire to share with everyone.

CRediT authorship contribution statement

Acknowledgements

References

[1] I. E. A. (IEA), Global EV Outlook 2023, Tech. rep., IEA, Paris (2023).
URL https://www.iea.org/reports/global-ev-outlook-2023

[2] E. Union, EU Action: 2050 long-term strategy (n.d.).
URL https://climate.ec.europa.eu/eu-action/climate-strategies-targets/2050-long-term-strategy_en

[3] E. E. Agency, Is Europe reducing its greenhouse gas emissions? — European Environment Agency (2022).
URL https://www.eea.europa.eu/themes/climate/eu-greenhouse-gas-inventory

[4] E. E. Agency, Transport and environment report 2022 — European Environment Agency (2023).
URL https://www.eea.europa.eu/publications/transport-and-environment-report-2022

[5] European Council, Fit for 55: towards more sustainable transport, publisher: General Secretariat of the European Council (Jun. 2023).
URL https://europa.eu/!yfBkpH

[6] Z. Liu, Z. Deng, S. J. Davis, C. Giron, P. Ciais, Monitoring global carbon emissions in 2021, Nature Reviews Earth & Environment 3 (4) (2022) 217–219. doi:10.1038/s43017-022-00285-w.
URL https://www.nature.com/articles/s43017-022-00285-w.pdf

[7] M. Kane, Global Plug-In Electric Car Sales Increased 61% In July 2022 (Sep. 2022).
URL https://bit.ly/408Tlas

[8] A. Pamidimukkala, S. Kermanshachi, J. M. Rosenberger, G. Hladik, Barriers and motivators to the adoption of electric vehicles: A global review, Green Energy and Intelligent Transportation 3 (2) (2024) 100153. doi:https://doi.org/10.1016/j.geits.2024.100153.
URL https://www.sciencedirect.com/science/article/pii/S2773153724000057

[9] E. Veldman, R. A. Verzijlbergh, Distribution grid impacts of smart electric vehicle charging from different perspectives, IEEE Transactions on Smart Grid 6 (1) (2015) 333–342. doi:10.1109/TSG.2014.2355494.

[10] C. B. Jones, M. Lave, W. Vining, B. M. Garcia, Uncontrolled Electric Vehicle Charging Impacts on Distribution Electric Power Systems with Primarily Residential, Commercial or Industrial Loads, Energies 14 (6) (2021) 1688. doi:10.3390/en14061688.
URL https://www.mdpi.com/1996-1073/14/6/1688

[11] E. Hopkins, D. Potoglou, S. Orford, L. Cipcigan, Can the equitable roll out of electric vehicle charging infrastructure be achieved?, Renewable and Sustainable Energy Reviews 182 (2023) 113398. doi:10.1016/j.rser.2023.113398.
URL https://linkinghub.elsevier.com/retrieve/pii/S1364032123002551

[12] M. Nicholas, D. Hall, N. Lutsey, Quantifying the evse gap across U.S. markets (Jan. 2019).
URL https://theicct.org/publication/quantifying-the-electric-vehicle-charging-infrastructure

[13] M. Li, Y. Wang, P. Peng, Z. Chen, Toward efficient smart management: A review of modeling and optimization approaches in electric vehicle-transportation network-grid integration, Green Energy and Intelligent Transportation (2024) 100181doi:https://doi.org/10.1016/j.geits.2024.100181.
URL https://www.sciencedirect.com/science/article/pii/S2773153724000331

[14] F. Gonzalez Venegas, M. Petit, Y. Perez, Active integration of electric vehicles into distribution grids: Barriers and frameworks for flexibility services, Renewable and Sustainable Energy Reviews 145 (2021) 111060. doi:https://doi.org/10.1016/j.rser.2021.111060.

[15] C. Develder, N. Sadeghianpourhamami, M. Strobbe, N. Refa, Quantifying flexibility in EV charging as DR potential: Analysis of two real-world data sets, in: 2016 IEEE International Conference on Smart Grid Communications (SmartGridComm), IEEE, Sydney, Australia, 2016, pp. 600–605. doi:10.1109/SmartGridComm.2016.7778827.
URL http://ieeexplore.ieee.org/document/7778827/

[16] E. E. Michaelides, V. N. Nguyen, D. N. Michaelides, The effect of electric vehicle energy storage on the transition to renewable energy, Green Energy and Intelligent Transportation 2 (1) (2023) 100042. doi:https://doi.org/10.1016/j.geits.2022.100042.
URL https://www.sciencedirect.com/science/article/pii/S2773153722000421

[17] V. Barthel, J. Schlund, P. Landes, V. Brandmeier, M. Pruckner, Analyzing the Charging Flexibility Potential of Different Electric Vehicle Fleets Using Real-World Charging Data, Energies 14 (16) (2021) 4961. doi:10.3390/en14164961.
URL https://www.mdpi.com/1996-1073/14/16/4961

[18] L. Sica, F. Deflorio, Estimation of charging demand for electric vehicles by discrete choice models and numerical simulations: Application to a case study in turin, Green Energy and Intelligent Transportation 2 (2) (2023) 100069. doi:https://doi.org/10.1016/j.geits.2023.100069.
URL https://www.sciencedirect.com/science/article/pii/S2773153723000051

[19] E. Genov, C. D. Cauwer, G. V. Kriekinge, T. Coosemans, M. Messagie, Forecasting flexibility of charging of electric vehicles: Tree and cluster-based methods, Applied Energy 353 (2024) 121969. doi:10.1016/j.apenergy.2023.121969.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261923013338

[20] S. Babrowski, H. Heinrichs, P. Jochem, W. Fichtner, Load shift potential of electric vehicles in Europe, Journal of Power Sources 255 (2014) 283–293. doi:10.1016/j.jpowsour.2014.01.019.
URL https://linkinghub.elsevier.com/retrieve/pii/S0378775314000342

[21] J. Zhang, J. Yan, Y. Liu, H. Zhang, G. Lv, Daily electric vehicle charging load profiles considering demographics of vehicle users, Applied Energy 274 (2020) 115063. doi:10.1016/j.apenergy.2020.115063.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261920305754

[22] F. M. Andersen, H. K. Jacobsen, P. A. Gunkel, Hourly charging profiles for electric vehicles and their effect on the aggregated consumption profile in Denmark, International Journal of Electrical Power & Energy Systems 130 (2021) 106900. doi:10.1016/j.ijepes.2021.106900.
URL https://linkinghub.elsevier.com/retrieve/pii/S014206152100140X

[23] G. Pareschi, L. Küng, G. Georges, K. Boulouchos, Are travel surveys a good basis for EV models? Validation of simulated charging profiles against empirical data, Applied Energy 275 (2020) 115318. doi:10.1016/j.apenergy.2020.115318.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261920308308

[24] J. R. Helmus, M. H. Lees, R. van den Hoed, A data driven typology of electric vehicle user types and charging sessions, Transportation Research Part C: Emerging Technologies 115 (2020) 102637. doi:10.1016/j.trc.2020.102637.
URL https://linkinghub.elsevier.com/retrieve/pii/S0968090X19315414

[25] S. Shahriar, A. R. Al-Ali, Impacts of COVID-19 on Electric Vehicle Charging Behavior: Data Analytics, Visualization, and Clustering, Applied System Innovation 5 (1) (2022) 12. doi:10.3390/asi5010012.
URL https://www.mdpi.com/2571-5577/5/1/12

[26] I. S. Bayram, A. Saad, R. Sims, A. Babu, C. Edmunds, S. Galloway, Statistical Characterization of Public AC EV Chargers in the U.K., IEEE Access 11 (2023) 70274–70287. doi:10.1109/ACCESS.2023.3293091.
URL https://ieeexplore.ieee.org/document/10175513/

[27] N. Sadeghianpourhamami, N. Refa, M. Strobbe, C. Develder, Quantitive analysis of electric vehicle flexibility: A data-driven approach, International Journal of Electrical Power & Energy Systems 95 (2018) 451–462. doi:10.1016/j.ijepes.2017.09.007.
URL https://linkinghub.elsevier.com/retrieve/pii/S0142061516323687

[28] A. Märtz, U. Langenmayr, S. Ried, K. Seddig, P. Jochem, Charging Behavior of Electric Vehicles: Temporal Clustering Based on Real-World Data, Energies 15 (18) (2022) 6575. doi:10.3390/en15186575.
URL https://www.mdpi.com/1996-1073/15/18/6575

[29] A. Jerónimo, P. Carvalho, C. Jesus, L. Dias, L. M. Ferreira, H. Morais, Modeling demand response of Charge Point Operators to consider flexibility in grid planning, in: International Conference on Smart Energy Systems and Technologies (SEST), SEST, Mugla, Turkey, 2023, pp. –.
URL https://ev4eu.eu/wp-content/uploads/2023/09/SEST_Jeronimo_Revised-1.pdf

[30] P. M. Carvalho, J. D. Peres, L. A. Ferreira, M. D. Ilic, M. Lauer, R. Jaddivada, Incentive-based load shifting dynamics and aggregators response predictability, Electric Power Systems Research 189 (2020) 106744. doi:10.1016/j.epsr.2020.106744.
URL https://linkinghub.elsevier.com/retrieve/pii/S0378779620305472

[31] Y. Amara-Ouali, Y. Goude, P. Massart, J.-M. Poggi, H. Yan, A Review of Electric Vehicle Load Open Data and Models, Energies 14 (8) (2021) 2233. doi:10.3390/en14082233.
URL https://www.mdpi.com/1996-1073/14/8/2233

[32] L. Calearo, M. Marinelli, C. Ziras, A review of data sources for electric vehicle integration studies, Renewable and Sustainable Energy Reviews 151 (2021) 111518. doi:10.1016/j.rser.2021.111518.
URL https://linkinghub.elsevier.com/retrieve/pii/S1364032121007966

[33] Z. J. Lee, T. Li, S. H. Low, ACN-Data – A Public EV Charging Dataset (2021).
URL https://ev.caltech.edu/dataset

[34] P. Ferreira, EV4EU launches today! (Jun. 2022).
URL https://www.inesc-id.pt/ev4eu-launches-today/

[35] A. Satre-Meloy, M. Diakonova, P. Grünewald, Cluster analysis and prediction of residential peak demand profiles using occupant activity data, Applied Energy 260 (2020) 114246. doi:10.1016/j.apenergy.2019.114246.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261919319336

[36] J. W. Tukey, Exploratory data analysis, Addison-Wesley series in behavioral science, Addison-Wesley Pub. Co, Reading, Mass, 1977.

[37] P. J. Rousseeuw, K. V. Driessen, A Fast Algorithm for the Minimum Covariance Determinant Estimator, Technometrics 41 (3) (1999) 212–223. doi:10.1080/00401706.1999.10485670.

URL https://www.tandfonline.com/doi/abs/10.1080/00401706.1999.10485670

[38] F. T. Liu, K. M. Ting, Z.-H. Zhou, Isolation Forest, in: 2008 Eighth IEEE International Conference on Data Mining, 2008, pp. 413–422, iSSN: 2374-8486. doi:10.1109/ICDM.2008.17.

[39] C. Daake, M. Cammerer, M. Hackmann, P3 Charging Index Report 07/22 – Comparison of the fast charging capability of various electric vehicles, Tech. Rep. 07/22, P3 GROUP GMBH (2022).
URL http://bit.ly/3KENHGJ

[40] Al-Ogaili, T. J. Tengku Hashim, N. A. Rahmat, A. K. Ramasamy, M. B. Marsadek, M. Faisal, M. A. Hannan, Review on Scheduling, Clustering, and Forecasting Strategies for Controlling Electric Vehicle Charging: Challenges and Recommendations, IEEE Access 7 (2019) 128353–128371. doi:10.1109/ACCESS.2019.2939595.
URL https://ieeexplore.ieee.org/document/8825773/

[41] S. Shahriar, A. R. Al-Ali, A. H. Osman, S. Dhou, M. Nijim, Machine Learning Approaches for EV Charging Behavior: A Review, IEEE Access 8 (2020) 168980–168993. doi:10.1109/ACCESS.2020.3023388.
URL https://ieeexplore.ieee.org/document/9194702/

[42] M. J. Zaki, W. Meira, Jr, Data Mining and Machine Learning: Fundamental Concepts and Algorithms, 2nd Edition, Cambridge University Press, 2020. doi:10.1017/9781108564175.

[43] L. M. L. Cam, J. Neyman, Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability: Weather modification, University of California, 1967, google-Books-ID: IC4Ku_7dBFUC.

[44] A. P. Dempster, N. M. Laird, D. B. Rubin, Maximum Likelihood from Incomplete Data via the EM Algorithm, Journal of the Royal Statistical Society. Series B (Methodological) 39 (1) (1977) 1–38.
URL http://www.jstor.org/stable/2984875

[45] M. Steinbach, G. Karypis, V. Kumar, A Comparison of Document Clustering Techniques, Report, University of Minnesota Digital Conservancy (May 2000).
URL http://conservancy.umn.edu/handle/11299/215421

[46] P. J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, Journal of Computational and Applied Mathematics 20 (1987) 53–65. doi:10.1016/0377-0427(87)90125-7.
URL https://linkinghub.elsevier.com/retrieve/pii/0377042787901257

[47] D. L. Davies, D. W. Bouldin, A Cluster Separation Measure, IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-1 (2) (1979) 224–227. doi:10.1109/TPAMI.1979.4766909.
URL http://ieeexplore.ieee.org/document/4766909/

[48] T. Calinski, J. Harabasz, A dendrite method for cluster analysis, Communications in Statistics - Theory and Methods 3 (1) (1974) 1–27. doi:10.1080/03610927408827101.
URL http://www.tandfonline.com/doi/abs/10.1080/03610927408827101

[49] T. Kodinariya, P. Makwana, Review on determining of cluster in k-means clustering, International Journal of Advance Research in Computer Science and Management Studies 1 (2013) 90–95.

[50] scikit-learn 1.3.0 documentation (2023).
URL https://scikit-learn.org/stable/index.html#

[51] V. Heinisch, L. Göransson, R. Erlandsson, H. Hodel, F. Johnsson, M. Odenberger, Smart electric vehicle charging strategies for sectoral coupling in a city energy system, Applied Energy 288 (2021) 116640. doi:10.1016/j.apenergy.2021.116640.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261921001756

[52] J. Dixon, W. Bukhsh, C. Edmunds, K. Bell, Scheduling electric vehicle charging to minimise carbon emissions and wind curtailment, Renewable Energy 161 (2020) 1072–1091. doi:10.1016/j.renene.2020.07.017.
URL https://linkinghub.elsevier.com/retrieve/pii/S0960148120310934

Highlights (options)

- Quantification of EV charging process's flexibility.
- Benchmarking of clustering methods for comprehensive EV charging profiles
- Evaluation of power flexibility with real datasets.
- Definition of flexibility data to be use din distribution grids planning

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: