Clustering Applications in Electric Mobility

Marcelo José da Silva Braço Forte
Instituto Superior Técnico — IST, Universidade de Lisboa
Lisboa, Portugal
Email: marcelojsforte@tecnico.ulisboa.pt
November 2023

***Abstract*—The continuous growth of electric vehicles (EVs) has been boosted by the need to achieve society's decarbonization targets. The mass adoption of EVs introduces new challenges in power systems planning and operation, and clustering has emerged as a powerful tool to help better understand and categorize the uncertain behavior of EV users and the electric vehicle supply equipment (EVSE) needs. In this thesis, different clustering techniques were evaluated to identify typical groups of EV charging processes to support characterizing EV charging profiles, EV user behavior profiles, and EVSE accessibility. The defined methodology comprises three major stages: data preprocessing, clustering application, and validation of results. A use case considering both open and private EV charging data (Caltech University and the publicly operated EVSEs in Greece, respectively) has been utilized to test the proposed methods, closing the gap verified in the literature. The experimental results demonstrated that Caltech features highly flexible charging sessions with routine users, while Greece exhibits more frequent EV users and quick-stay sessions. Additionally, there is an excellent opportunity to expand the charging network in Greece at specific locations. 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.**

***Index Terms*—Clustering, Data Analysis, Electric Mobility, Typical profiles, User Behavior.**

I. Introduction

THE world is an ever-changing place, but in between this revolution, one thing seems well-defined: fight climate change. The adoption of electric vehicles (EVs) has experienced rapid growth in the 21st century due to the pressing need to transition global energy demand away from fossil fuels, particularly within the past decade. 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 [1], since Transport is the only sector where greenhouse gas (GHG) emissions have increased in the past three decades in Europe [2]. This sector was responsible for more than a quarter of the EU's total GHG emissions in 2019, of which approximately 71% came from road transportation, increasing 33% between 1990 and 2019, according to a 2022 report by the European Environment Agency [3].

To achieve carbon neutrality, in 2022 the EU's environment ministers approved the "Fit for 55 in 2030" package [4], which orders 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 [5], together with the EU, the biggest polluters in the world. With that in mind, car manufacturers and governments have been investing in new models and tax incentives for the purchase and adoption of EVs, whose popularity has significantly increased over the past five years [6].

Due to all these factors, the number of EVs will certainly increase in the upcoming years. However, the EV rise poses several challenges in the power systems, mainly at the distribution level. Uncontrolled EV charging negatively impacts the existing power grid, including high load peaks, system overcurrents, and degradation of power quality [7].

Utilities, distribution system operators (DSOs), and charge point operators (CPOs) need to quantify the impacts on grid infrastructure and network reinforcement demands to address future challenges associated with EV and electric vehicle supply equipment (EVSE) deployment. The identification of typical profiles is of great relevance for these entities to perform a successful and intelligent integration of EVs in the energy system.

A. Background

EV charging data has been submitted to clustering to identify the most common and recurrent profiles [8]. However, most of these studies lack practical relevance to help DSOs and CPOs with grid management. Additionally, most of these (few) studies used datasets from countries outside of Europe. For instance, Märtz et al. [9] employed Gaussian Mixture Model (GMM) and K-means clustering to analyze the charging behavior of EVs using an extensive private dataset from 2019. Van Kriekinge et al. [10] proposed a methodology to simulate the charging demand for different types of drivers. Typical EV driver profiles with similar charging habits are needed to accomplish this goal. The authors grouped the data by the users into (average value of the plug-in times, parking times, and charged energy). Besides, the focus of the literature goes beyond charging patterns, as Carlton and Sultana [11] performed spatial clustering of public EVSEs to analyze the characteristics of their land use, and how these impact EVSE accessibility in the Chicago Metropolitan Area.

B. EV Charging vs User Behavior vs EVSE Accessibility

The literature often considers EV charging and EV user behavior profiles synonyms. The same does not happen for EVSE accessibility, whose studies utilize EVSE location data and not EV charging data. However, in this work, the two types of profiles are not synonymous, they represent different information: an ***EV charging profile*** aims to characterize the times of day when more or fewer charging sessions occur, whether the sessions are high energy, low energy, with high or low flexibility potential; an ***EV user behavior profile*** intends to give an understanding of whether the user's behavior is routine, or random and without a typical charging frequency. Regarding ***EVSE accessibility***, the aim is to understand the geographic distribution of the corresponding Charging Pools (CPs), whether the current supply is in line with the demand, and whether there are inequalities that prevent the widespread use of EVs. This work assumes **CP** as a site with one or more Charging Stations (CSs), operated by a CPO. Additionally, **CS** is a physical object that includes one or more EVSEs, and **EVSE** is the equipment that provides electricity to an EV.

C. Main Contributions

The main aim of this work is to investigate the possibility of identifying different groups of EV charging processes, through clustering, to provide support in characterizing EV charging profiles, EV user behavior profiles, and EVSE accessibility, based on comprehensive datasets of empirical charging processes. A detailed insight into the complexity of EV charging behavior has enormous significance for the future sizing of distribution grids and charging infrastructures. It can also be helpful in future studies, particularly in the coordination of EVs with solar and wind renewable energies.

D. Paper Organization

The document is organized as follows. Section II presents the proposal of methodologies to achieve the intended objectives, along with a description of the chosen datasets, and the clustering/evaluation methods. Sections III, IV and V perform a detailed explanation of the obtained results, summarizing and commenting on the main findings. Section VI proposes practical applications for the outcomes of this work. Finally, Section VII contains the conclusions and possible future work.

II. Methodology

The overview of the methodological approach for the proposed solution is illustrated in Fig. 1. A description of the datasets' characteristics is done in II-A. The data preprocessing steps are explained in II-B. II-C describes the main characteristics of clustering, in particular those utilized in this study, and II-D presents the selected cluster validation techniques.

A. Data Description and Analysis

There is no cluster analysis without a dataset. Therefore, it is essential to have an adequate EV charging dataset. An open dataset will be studied to find EV charging and EV user behavior profiles: **ACN-Data** [12], from a parking garage available to the public at Caltech University (USA). In addition to open data, this paper had access to private datasets from several European partners in the context of the **EV4EU** project [13]. Thus, by performing deep analyses of this data, the current study contributes to filling the lack of European studies identified previously. The private dataset of public EVSEs in Greece (**GR-Data**) was chosen to find EV charging and EV user behavior profiles, and analyze the EVSE accessibility since it features multiple infrastructures across the country, located in quick-stay locations like gas stations, supermarkets, or stores. Both datasets are in the format *charging event* (1 row of the dataset, 1 EVSE transaction). Table I presents a summary of the characteristics of the chosen datasets.

TABLE I
Summary of the main characteristics in the chosen datasets.

Datasets	ACN-Data	GR-Data
File Format	JSON file	CSV file
Time Interval	Apr 2018 - Sep 2021	Jul 2021 - May 2022
Total Sessions	31 424	22 412
No. of different EVSEs	55	312
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
Energy Consumed	Yes	Yes
EVSEs' Max Power	No	Yes
Customer ID	Yes	Yes

B. Stage 1: Data Preprocessing and Cleaning

According to earlier research such as [8], [9], data cleaning and preprocessing are two key processes in obtaining interpretable results from cluster analysis.

*1) **Deal with Outliers and Missing Data**:* Some datasets' entries might have **missing information**. 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. 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) [14], Elliptic Envelope [15], Isolation Forest [16], 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. Therefore, 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) **Feature Engineering**:* Another relevant step comprises creating features not previously included in the dataset that help to analyze and obtain more meaningful clustering. According to Table I, the datasets do not provide all the required fields to obtain EV charging profiles and EV user behavior profiles. Therefore, additional features must be created.

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*).

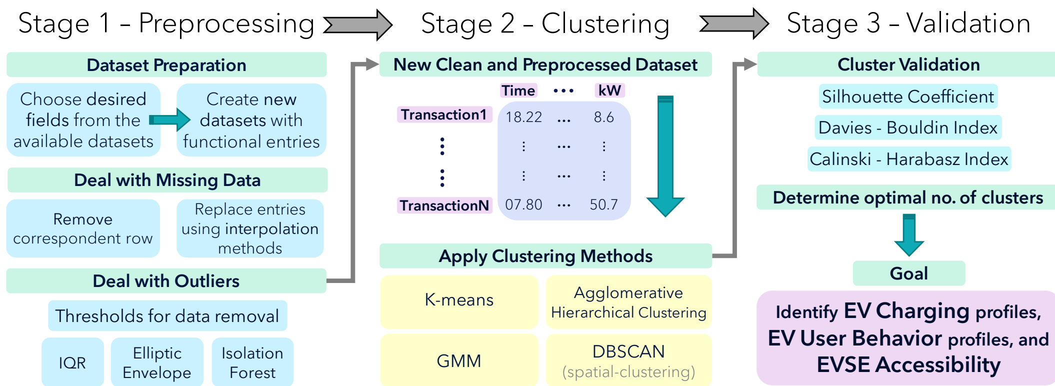

Fig. 1. Overview of methodological approach.

Consequently, 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

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

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

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

ACN-Data contains all the information required for the expressions (1) and (2). However, the GR-Data does not provide access to the end of charging times, 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) (equal to 0.8) guarantees a more realistic charging time since the process is not carried out at a constant power rate [17].

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

Regarding the EV user behavior profiles, the datasets' sessions should be grouped by customer ID to obtain characteristic average values for each EV user. Defining **standard deviations** for temporal fields, including *plug-in time*, *charging time*, or *sojourn time*, can offer valuable details into the variability and dispersion of these fields. Additionally, since the main goal is to get insights into the frequency of charging, a new feature must be associated with the users, defined by

$$\text{Frequency}_{user_i} = \frac{number\ of\ sessions_i}{Period,\ in\ weeks}, \tag{5}$$

where the denominator comprises the number of weeks between the first and last session the EV user attended the Caltech or Greek EVSEs. Thus, the frequency values indicate the average number of days the driver charged its EV, per week.

C. Stage 2: Selected Clustering Methods

Three well-known clustering methods are the choice for identifying groups of similar charging patterns and EV user behavior: K-means, GMM, and Hierarchical clustering. To analyze the distribution and accessibility of public EVSEs, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) will be employed to perform spatial clustering.

These clustering methods are frequently employed in applications related to charging behavior, as mentioned in I-A. Moreover, Shahriar et al. [18] precisely suggest the use of these methods for the analysis of EV charging patterns (this study provides a comprehensive overview of the application of Machine Learning (ML) techniques in EV and EVSE deployment data). Nevertheless, it is important to give a brief introduction to clustering and its methods.

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** [19]. 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 dataset 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). Since there is no clear definition of 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 [19]. The following subsections give a brief introduction to the existing clustering methods. For a complete and detailed explanation, see [19].

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 [20] 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 [21] 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) Hierarchical Clustering: **Hierarchical Clustering** techniques create a sequence of nested partitions, which can be visualized as a tree (*dendrogram*), indicating the merging process and the intermediate clusters. The highest level (root) of the tree consists of all points in one single cluster, whereas the lowest level (leaves) consists of clusters of individual points, each point in its cluster. If the desired number of clusters is known, one can graphically see the level at which k clusters exist. There are two algorithmic approaches [22]:

- **Agglomerative**: Start with the points as individual clusters and, at each step, merge (or agglomerate) the most similar or closest pair of clusters until the desired number of clusters has been found. This requires a definition of cluster similarity or distance. For this, a variety of distance measures can be used, including **single link**, **complete link**, **average link**, or **Ward's method**;
- **Divisive**: Start with one cluster (all points), and at each step, divide a cluster until only clusters of individual points remain. In this case, it is required to decide, at each stage, which cluster to split and how to perform it. It works just the opposite of the Agglomerative approach.

3) Density-based clustering : **Density-based clustering** methods use the density or connectedness properties to find nonconvex clusters. It employs the local density of points to determine the clusters rather than using only the distance between points, such as in K-means or EM. The most popular method is Density-Based Spatial Clustering of Applications with Noise (DBSCAN), which needs the thresholds ϵ (defines the neighborhood of a point) and *minpts* (specifies the minimum number of points to form a cluster) defined a priori.

4) Spectral/Graph clustering: **Graph clustering** can be viewed as an optimization problem over a k-way cut in a graph, with different objectives represented as spectral decompositions of various graph matrices, derived from the original graph data or the kernel matrix, such as the adjacency matrix and Laplacian matrix [23]. The graph can then be split into connected components using a specific graph cut method, and those components are referred to as clusters.

D. Stage 3: Clustering Validation Techniques

Since no ground truth is available, internal validation should be used to quantify the performance of the clustering [19]:

*1) **Silhouette Coefficient**:* For each point $\boldsymbol{x}_i$, the silhouette coefficient is

$$s_i = \frac{\mu_{out}^{min}(\boldsymbol{x}_i) - \mu_{in}(\boldsymbol{x}_i)}{\max\{\mu_{out}^{min}(\boldsymbol{x}_i), \mu_{in}(\boldsymbol{x}_i)\}}, \tag{6}$$

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** [24] is defined as the mean s_i value across all points, given by (7), where a value close to $+1$ denotes good clustering.

$$SC = \frac{1}{n}\sum_{i=1}^{n} s_i \tag{7}$$

*2) **Davies-Bouldin Index**:* The Davies-Bouldin measure for a pair of clusters C_i and C_j is defined as

$$DB_{ij} = \frac{\sigma_{\mu_i} + \sigma_{\mu_j}}{\delta(\mu_i, \mu_j)}, \tag{8}$$

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** [25] is thus defined as

$$DB = \frac{1}{k} \cdot \sum_{i=1}^{k} \max_{i \neq j}\{DB_{ij}\}, \tag{9}$$

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, closer to zero, mean better clustering (clusters are well separated and each one is well represented by its centroid).

*3) **Calinski-Harabasz Index**:* The **Calinski-Harabasz index** [26] is given by

$$CH(k) = \frac{tr(\boldsymbol{S}_B)}{tr(\boldsymbol{S}_W)} \cdot \frac{n-k}{k-1}, \tag{10}$$

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 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 typically used for K-means.

III. EV Charging profiles

This section starts with the datasets' preprocessing description, followed by a description and discussion of the obtained EV charging profiles. The code was written in Python 3.10.11 on a Jupyter Notebook using the Google Colab platform, and the *scikit-learn* library [27] for the preprocessing, clustering, and evaluation methods (most parameters were left at default, while those modified are mentioned throughout the text).

A. ACN-Data: Data Preprocessing

*1) **Dataset preparation and feature engineering** :* The first step in obtaining EV charging and user behavior 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 the DateTime method of *Pandas* library to obtain the fields *sojournTime*, *chargingTime*, and *idleTime* based on (1), (2) and (3), respectively.

After determining the extra fields, the entries in DateTime format needed to be converted into a suitable numeric structure, as mentioned by Shahriar and Al-Ali [8] and by Märtz et al. [9]. Thus, the *connectionTime*, *disconnectTime*, *doneChargingTime*, *sojournTime*, *chargingTime*, and *idleTime* fields were converted into float format: for example, 10h17 (10 hours and 17 minutes) becomes 10.28h (10.28 hours), consequently allowing full use of outlier removal approaches, clustering methods, and graphical representations.

*2) **Deal with Missing Data:*** After analyzing the preprocessed dataset, *doneChargingTime* 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* entry in these sessions, leading to an idle time of zero. Regarding the *userID*, the lack of this information makes it impossible to discover or predict the corresponding session user. Hence, for the user behavior study (Section IV), only sessions with *userID* will be considered.

*3) **Outlier Detection:*** With a fully functional dataset, the next step involved setting thresholds to remove unwanted data. Thus, the limit defined allows removing all sessions with a *sojournTime* or *chargingTime* greater than 48 hours and less than 1 minute. Another threshold was also set to clear sessions with energy-supplied values greater than 100 kWh, selected considering the characteristics of EVs available on the market during the period of the data (2018-2021). All negative entries were also removed. 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.

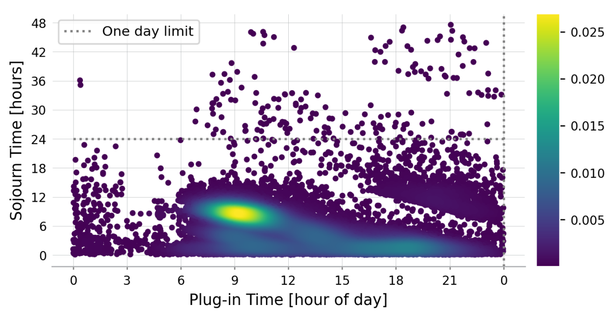

Fig. 2. Clean ACN-Data distribution regarding Sojourn and Plug-in Time.

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 fields' thresholds. Therefore, no outlier removal method will be employed for the ACN-Data dataset.

*4) **Data Adjustment:*** The plug-in time with day and hours became plug-in time at 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 II.

TABLE II
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
doneChargingTime (End Charging)	31318	float64
kWhDelivered	31318	float64
stationID	31318	float64
userID	16355	float64
chargingTime	31318	float64
sojournTime	31318	float64
idleTime	31318	float64

B. ACN-Data: EV Charging profiles

*1) **Chosen fields and normalization of the data**:* Following Shahriar and Al-Ali's article [8] 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 *doneChargingTime* fields are redundant, so only one is necessary. The same 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 a first cluster analysis. The remaining fields were eliminated, and the data were normalized to obtain the best possible results, described next.

*2) **Results**:* 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, 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* covariance; for Agglomerative Hierarchical clustering, the distance measure. 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 III summarizes the selected metrics for each clustering method, revealing that K-means produced the best scores.

TABLE III
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

However, when analyzing the resulting profiles comprehensively, one verifies that GMM yielded superior results, which will be detailed next. Fig. 3 presents the distribution of the adjusted EV charging profiles regarding the Plug-in Time, Sojourn Time, and kWh (energy delivered) fields.

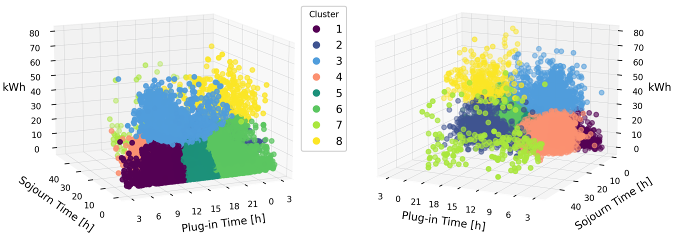

Fig. 3. 3D distribution of the adjusted GMM EV Charging profiles for the ACN-Data dataset.

From Fig. 3, one can see that the profiles are well-defined and have little overlap. An intriguing result that is immediately apparent is the separation of the high consumption profiles (clusters 3 and 8), which are virtually divided by the plane defined by kWh $\approx$ 20, from the low and medium consumption profiles (clusters 1, 2, 4, 5, and 6). It is noticeable that there are more short/medium-term sessions, which impacts the number of profiles. The longer sojourn time sessions are comprised in clusters 2, 4, and 7.

Additionally, the GMM method was able to group the most different sessions into a single cluster (cluster 7), allowing the remaining profiles to reveal typical average values more adjusted to the sessions they contain.

Another noteworthy point concerns idle times (or the flexibility potential). After analyzing the session characteristics of each cluster, it emerged that cluster 4 exhibits an average idle time that surpasses the average charging time, meaning that the EVs spend more time parked without charging than actually charging. 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 aligns with the EVSE's location (Caltech University) and typical workplace behavior: EVs connected in the morning/early afternoon and unplugged at the end of the working day, with high flexibility.

C. GR-Data: Data Preprocessing

*1) **Dataset preparation and feature engineering**:* Since this dataset had the same format as the ACN-Data, the steps followed for the conversion of the entries were identical, only adapting to 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 entries in DateTime format were also converted to float values.

*2) **Deal with Missing Data**:* After analyzing the preprocessed dataset, there were no missing entries in the *Plug-in Time*, *Plug-Out Time*, *stationID*, or *userID* fields. However, the *kWhDelivered* and *maxPowerEVSE* entries were occasionally missing, making it impossible to determine the average charging time and idle 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) **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 2021-2022 EV sales in Europe). There are only 32 sessions with more than 24 hours of parking stay. Consequently, by removing these sessions, the clustering results will be improved, yielding more meaningful clusters. All negative entries were also removed.

*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. 4. Table IV contains all the usable fields from the final preprocessed dataset.

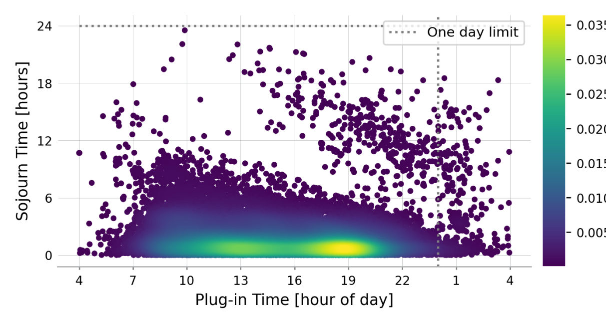

Fig. 4. Final adjusted GR-Data distribution depending on Sojourn Time and Plug-in Time.

TABLE IV
SUMMARY OF THE FINAL USABLE FIELDS IN THE GR-DATA DATASET.

Field name	Non-Null count	Dtype
Start_datetime (Plug-in)	21801	float64
End_datetime (Plug-out)	21801	float64
kWhDelivered	21801	float64
stationID	21801	object
maxPowerEVSE	21081	float64
userID	21801	object
sojournTime	21801	float64
averageChargingTime	21801	float64
idleTime	21801	float64

D. GR-Data: EV Charging profiles

*1) **Chosen fields and normalization of the data**:* A similar analysis described in subsection III-B1 was also 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 results, detailed next.

*2) **Results**:* Following the method described in III-B2 for the ACN-Data dataset, it was possible to obtain Table V, which summarizes the optimal scores obtained and the specific characteristics selected for each clustering method.

TABLE V
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	0.326	0.309	0.322
Davies-Bouldin	0.983	1.061	1.036
Calinski-Harabasz	10715.45	9259.41	11777.57

Therefore, by analyzing Table V and the resulting profiles, the K-means method produced the best clusters, 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.

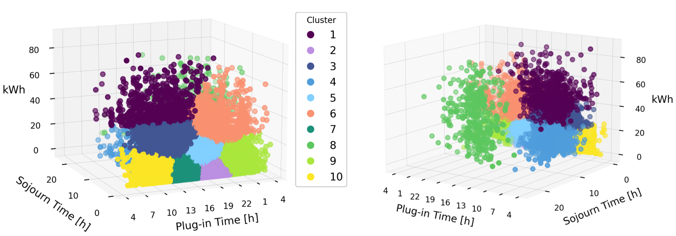

Fig. 5. 3D distribution of the adjusted K-means EV Charging profiles for the GR-Data dataset.

By observing Fig. 5, one can see that the results are relatively similar to those obtained for the ACN-Data dataset. There is, however, greater separation between the sessions as five clusters were found with plug-in times in the morning (clusters 1, 3, 4, 7, and 10) and only three with plug-in times in the evening (clusters 6, 8 and 9). There are also two clusters during the middle/late afternoon (clusters 2 and 5). Therefore, one can conclude that, in this dataset, the sessions during the day differ significantly from each other, translating into a higher number of daily profiles when compared to ACN-Data. Additionally, the reduced number of clusters in the evening suggests that the sessions during this period exhibit more similar behavior than those during the day.

After analyzing the session characteristics of each cluster, one verifies that clusters 2 and 7 comprise the highest number of sessions. This means that the short, low-energy, and low-flexibility potential sessions are the most frequent. The later the drivers plug in, the more energy they consume. Morning and afternoon profiles are generally lower energy.

Another interesting point is that the most different sessions (with higher sojourn times and, consequently, higher flexibility potential) fall into distinct clusters: cluster 8 contains the sessions that only end the next day, regardless of the plug-in time, while cluster 4 comprises the sessions that start in the morning and only end in the afternoon of the same day. The remaining clusters represent typical charging sessions at quick-stay locations.

IV. EV USER BEHAVIOR PROFILES

Due to the similarity in data preprocessing, IV-A presents the steps performed for the two datasets under analysis. Then, the obtained profiles are detailed for each dataset separately.

A. ACN-Data & GR-Data: Chosen fields and normalization of the data

The first step consisted of creating new datasets, focusing solely on sessions with a *userID* (remember Tables II and IV). The sessions were grouped by user, replacing all individual driver sessions with a single theoretical charging session composed by the *mean* of the plug-in times, *mean* of the sojourn times, *standard deviation* (Std) of plug-in times, and *standard deviation* of sojourn times. Then, a threshold was defined to eliminate users with less than three recorded sessions, which were random and unsuitable for finding behavior patterns. As a result, the ACN-Data dataset dropped from 571 to 338 users, and the GR-Data dataset from 3184 to 1228 users. In addition to these fields, a new feature must be associated with the users to differentiate regular EV drivers from occasional ones: the frequency field, obtained through (5). Table VI outlines the usable fields from the user behavior datasets.

TABLE VI
SUMMARY OF THE USABLE FIELDS IN THE ACN-DATA AND GR-DATA USER BEHAVIOR DATASETS.

Field name	ACN-Data Non-Null count	GR-Data Non-Null count	Dtype
mean Plug-in Time	338	1228	float64
mean Std of Plug-in Time	338	1228	float64
mean sojournTime	338	1228	float64
mean Std of sojournTime	338	1228	float64
frequency	338	1228	float64

The following stage involved selecting fields for clustering, a comparable but more time-consuming process than the selection for EV charging profiles, mainly due to the subjectivity of *user behavior* in the literature. Nevertheless, in this study, the fields *std of plug-in time*, *std of sojourn time*, and *frequency* were chosen, as this triplet yielded the most interpretable profiles among the available fields. The selected fields were normalized to obtain the best possible results, detailed next.

B. ACN-Data: EV User Behavior profiles

*1) **Results**:* Once again, following the method described in III-B2, Table VII summarizes the optimal scores obtained and the specific characteristics chosen for each clustering method.

TABLE VII
SUMMARY OF THE SELECTED METRICS FOR EACH ACN-DATA USER BEHAVIOR CLUSTERING METHOD.

	K-means	GMM	Hierarchical
Best k	4	4	4
Parameters	-	Tied Covariance	Ward's Method
Elbow Method	k={4, 5, 6}	-	-
Silhouette	0.362	0.421	0.333
Davies-Bouldin	0.922	0.800	0.927
Calinski-Harabasz	175.08	131.72	151.62

Table VII reveals that GMM produced the best Silhouette and Davies-Boudin scores, while K-means yielded a higher Calinski-Harabasz. When further analyzing the resulting profiles, one verifies that GMM grouped most users into a single cluster while the remaining clusters comprised extreme users. These profiles can be valuable for specific applications; however, in the context of EV user behavior profiles, K-means was the most consensual method. Fig. 6 presents the distribution of the EV user behavior profiles regarding the Std of the Plug-in Time, Std of the Sojourn Time, and Frequency fields.

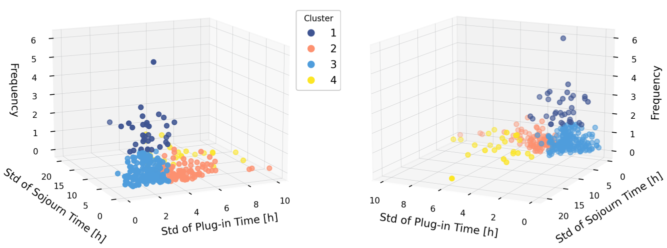

Fig. 6. 3D distribution of the K-means EV User Behavior profiles for the ACN-Data dataset.

From Fig. 6, one can see that the profiles are well-defined and have minimal overlap. An interesting result is the separation of high-frequency users (cluster 1), which are virtually divided by the plane defined by Frequency $\approx$ 1.5 from the low and medium frequency profiles. Additionally, it is noticeable that most users exhibit relatively constant behavior, with low standard deviations of plug-in and sojourn times. However, there are also users with high standard deviations, mainly present in cluster 4, indicating that their charging behavior is indeed random and lacks routine.

After analyzing the user characteristics of each cluster, one verifies that cluster 3 comprises most users, characterized by routine behavior without high deviations. Therefore, most users of the Caltech EVSEs have a predictable behavior since they start charging in the morning and only end in the late afternoon, with reduced deviations. The remaining clusters ultimately differentiate the most extreme users in each field.

C. GR-Data: EV User Behavior profiles

*1) **Results**:* Table VIII summarizes the metrics and parameters selected for each clustering method. GMM revealed the best Silhouette and Davies-Boudin scores. However, it achieved that by placing most users in one cluster, while the remaining clusters comprised extreme users. On the other hand, K-means produced a higher Calinski-Harabasz index and superior overall profiles.

TABLE VIII
SUMMARY OF THE SELECTED METRICS FOR EACH GR-DATA USER BEHAVIOR CLUSTERING METHOD.

	K-means	**GMM**	**Hierarchical**
Best k	5	4	5
Parameters	-	Tied Covariance	Ward's Method
Elbow Method	k={4, 5, 6, 7}	-	-
Silhouette	0.324	0.471	0.285
Davies-Bouldin	0.964	0.870	1.042
Calinski-Harabasz	558.37	391.75	483.32

Fig. 7 presents the distribution of the EV user behavior profiles regarding the Std of the Plug-in Time, Std of the Sojourn Time, and Frequency fields.

Comparing the results to those obtained for the ACN-Data (refer to IV-B), there is an increase in users that led to a growth in the number of clusters with distinct standard deviations of plug-in time (clusters 1-3). One profile with the most frequent users (cluster 4) remains, while cluster 5 consists of the users with the highest standard deviations of sojourn time (which do not follow a charging routine).

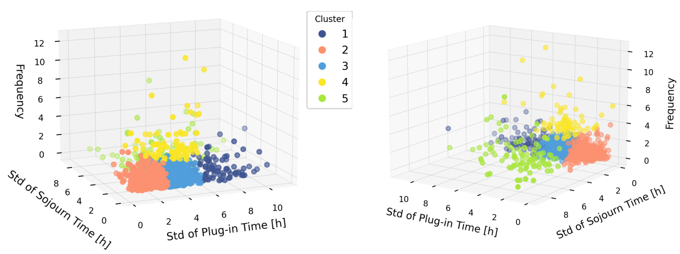

Fig. 7. 3D distribution of the K-means EV User Behavior profiles for the GR-Data dataset.

The analysis of the user characteristics of each cluster revealed that K-means did not find a correlation between the plug-in time's standard deviation and the plug-in time's mean value. This means that the Greek users do not demonstrate a specific routine of only charging in the morning or the evening, for example. Nonetheless, cluster 2 contains the most routine users who recharge at lunchtime, with minor deviations regarding the plug-in time and sojourn time fields.

One noteworthy aspect is that the charging frequency in GR-Data user profiles is higher than in the profiles obtained for ACN-Data. Further examination reveals that GR-Data profiles demonstrate an overall higher frequency but a lower sojourn time, while ACN-Data profiles demonstrate the opposite. This is justified by the location of the EVSEs and the consequent user behavior since the GR-Data's EVSEs are either situated along highways, gas stations, or in quick-stay areas like supermarkets. On the other hand, the ACN-Data's EVSEs are located in a garage, thus allowing for longer sojourn times.

V. EVSE ACCESSIBILITY

One of the barriers noted for massifying EVs is the scarcity of Charging Pools (CPs), especially publicly available infrastructures for those who cannot recharge at home. This study precisely aims to analyze the accessibility of EVSEs and understand whether the current supply is in line with the demand. The GR-Data dataset was chosen to conduct an in-depth study on this topic.

A. GR-Data dataset

The GR-Data dataset provides details on the location of the EVSEs, namely the address, zip code, and city. However, the geographic coordinates are needed to analyze the spatial distribution of the EVSEs, i.e., in the format (*latitude*, *longitude*). Therefore, it was necessary to modify the preprocessed dataset (recall Table IV) by creating a new column with the location of the EVSEs in the format "*address, zip code, city, country*" for each session (row of the dataset). By removing the remaining fields and the duplicated entries, a new dataset was created with 124 unique CPs. Then, the **OpenCageGeocode API** [28] was employed to obtain the geographic coordinates of each address. It is a service that provides the corresponding geographic coordinates accurately by inputting an address.

DBSCAN is particularly attractive for studying the accessibility of EVSEs as it allows finding irregularly shaped clusters. Specifically, it can find data points that do not fit into any group, labeling them for *cluster -1*. A small number of clusters indicates a uniform distribution over the territory of Greece, ideally equal to one for a perfectly uniform distribution.

*1) **Results** :* As previously mentioned in II-C3, the threshold ϵ and the $minpts$ value must be defined a priori. In the literature, the **Haversine distance metric** - a metric option in *scikit-learn*'s DBSCAN method [29] - is often employed to calculate the distance between coordinate pairs (*latitude, longitude*). Considering the number of existing EVSEs, $minpts = 1$ was defined, meaning that at least two different CPs are needed to form a cluster. The threshold ϵ was set to 5 km, selected considering the size of the Greek cities, meaning that a CP belongs to the neighborhood of another if it is less than 5 km away. Table IX reveals the most relevant information found for each cluster, including the utilization rate of the most attended EVSE (number of sessions divided by the number of weeks with recorded sessions).

TABLE IX
KEY CHARACTERISTICS OF THE GR-DATA DBSCAN CLUSTERS.

Cluster ID	Region	No. of CPs	Total no. of sessions	Maximum* Utilization Rate
-1	-	49	3118	6
0	Athens	26	9356	21
1	Athens	7	1475	11
2	Athens	2	1062	7
3	Thessaloniki	12	2649	31
4	Athens	10	2990	11
5	Chania	2	74	7
6	Ioannina	2	116	3
7	Patras	3	335	5
8	Kalamata	2	90	2
9	Larissa	2	172	3
10	Heraklion	7	364	3

*Utilization rate for the most utilized EVSE in each cluster.

Four clusters are in the Athens region, the most populous Greek urban area, with cluster 0 being the largest with 26 CP locations. Cluster 3 with 12 CP locations is found in Greece's second-largest region, Thessaloniki. The largest cluster outside these areas is cluster 10, which contains 7 CPs in Heraklion, on the Crete island. Cluster -1 corresponds to 49 non-clustered sites, on highways or in areas with a reduced EVSE network.

This concentration of CPs in city centers is an excellent representation of the disparity of EVSE accessibility. Smaller towns and even suburban areas lack sufficient charging infrastructure for the imminent rise in EV sales. Additionally, Table IX reveals that the utilization rate of the most popular EVSE in each cluster is often low, regardless of the size of the areas they serve. Most sessions occur in cluster 0; however, the most requested EVSE is found in cluster 3, with roughly 31 weekly charging sessions. Its popularity may derive from its convenient location (large multinational retailer) and its maximum power capacity of 50 kW, surpassing the 22 kW of the remaining EVSEs in the cluster. Expansion of the EVSE network within this location represents a potentially advantageous strategy regarding demand and economics.

VI. Practical Applications

The present work provides various practical applications. The EV charging profiles, EV user behavior profiles, and EVSE accessibility yield valuable information for CPOs and DSOs that assists in grid management and the correct insertion of EVs into the energy system. In the literature, these outcomes have been exploited for various practical applications. For instance, Xiong et al. [30] discovered EV user behavior profiles and used them as input for a model that can apply to different **scheduling** EV charging algorithms. Nespoli et al. [31] also relied on clustering to identify typical profiles, a fundamental step in obtaining the **forecast** results.

Additionally, there are further applications not extensively explored in the literature. EV charging profiles can be employed to provide empirical **flexibility** data for various future investigations. For instance, Jerónimo et al. study [32] could be adapted to benefit from this empirical data rather than resorting to simulation algorithms (the authors present a new flexibility model for CPOs, starting with their characterization based on charging and occupancy rates). Fig. 8 illustrates the temporal characterization of the typical profiles found.

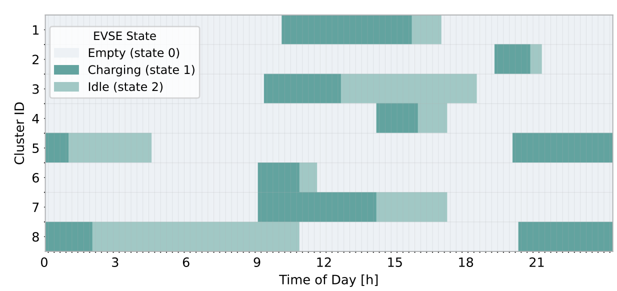

Fig. 8. Flexibility characterization of the EV Charging profiles for the ACN-Data dataset.

Regarding EV user behavior profiles, they can be applied to create **customized charging tariffs**. Since these tariffs can affect charging behavior and result in advantages for the user (e.g., cost efficiency), the environment (e.g., collaboration with renewable energy sources), and grid management [33], they may be a powerful tool for the growth of EVs. Finally, the EVSE accessibility study can assist CPOs in the appropriate **EVSE placement**, assuming geographic and economic factors.

VII. Conclusions

This work performed a robust assessment of the possible applications of clustering in EV-related data. In particular, EV charging profiles, EV user behavior profiles, and EVSE accessibility were found by applying clustering methods to datasets of empirical charging processes: ACN-Data (open data) and GR-Data (private data from one of the Greek EV4EU project partners).

The experimental results demonstrated the feasibility of utilizing clustering techniques to extract comprehensive insights into the EV charging process, the behavior of EV drivers, and the accessibility of EVSEs, confirming all the objectives. The EV charging and EV user behavior profiles were obtained using K-means, GMM, and Hierarchical clustering, with subsequent comparison of methods and approaches. DBSCAN was employed to obtain information about the accessibility and distribution of Greek EVSEs.

ACN-Data is characterized by highly flexible EV charging profiles since EVs spend more time parked than charging, and GR-Data predominantly contains quick-stay sessions, in line with the EVSEs' locations (publicly available infrastructures). The analysis of this information can be helpful in future projects, particularly in the coordination of EVs with solar and wind renewable energies. Concerning the EV user behavior profiles, most ACN-Data users choose infrequent, long charging routines, typically every two weeks. In contrast, GR-Data includes more frequent users without a specific charging routine, as they prefer short sessions at no particular time of the day, on different EVSEs. This information can be applied to create personalized charging tariffs that can benefit the user, the environment, and the grid. Furthermore, studying the accessibility of EVSEs revealed the geographic distribution of the corresponding publicly-operated Greek CPs, confirming that the most populated cities had the most extensive charging networks during the 2021-2022 data period and indicating the possibility for additional EVSEs in strategic locations.

The results of this work seek to help DSOs and CPOs to perform a successful and intelligent integration of EVs into the energy system, providing them with valuable information about the charging behavior of EVs and EVSEs.

VIII. System Limitations and Future Work

The conducted study presents some limitations due to the available data. The EV charging profiles from ACN-Data and GR-Data reflect the data's nature since they only depict charging patterns in those specific regions/countries. Hence, future work may include further clustering studies with newly available datasets from different regions/countries to increase knowledge about EVs and EVSEs. Besides, in an era characterized by large volumes of data, clustering has emerged as a powerful ally for processing and extracting valuable information for forthcoming studies. By seamlessly integrating simulated studies with real-world data enhanced through clustering, a giant leap can be made toward understanding and guiding a sustainable future.

References

[1] E. Union, " EU Action: 2050 long-term strategy," n.d. [Online]. Available: https://bit.ly/3tjNqTE

[2] E. E. Agency, "Is Europe reducing its greenhouse gas emissions?" 2022. [Online]. Available: https://bit.ly/45fhugw

[3] ——, "Transport and environment report 2022," 2023. [Online]. Available: https://bit.ly/3LPb94u

[4] European Council, "Fit for 55," Jun. 2023, publisher: General Secretariat of the European Council. [Online]. Available: https://europa.eu/!yfBkpH

[5] Z. Liu, Z. Deng, S. J. Davis, C. Giron, and P. Ciais, "Monitoring global carbon emissions in 2021," *Nature Reviews Earth & Environment*, vol. 3, no. 4, pp. 217–219, Mar. 2022.

[6] M. Kane, "Global Plug-In Electric Car Sales Increased 61% In July 2022," Sep. 2022. [Online]. Available: https://bit.ly/408Tlas

[7] C. B. Jones, M. Lave, W. Vining, and B. M. Garcia, "Uncontrolled Electric Vehicle Charging Impacts on Distribution Electric Power Systems with Primarily Residential, Commercial or Industrial Loads," *Energies*, vol. 14, no. 6, p. 1688, Jan. 2021.

[8] S. Shahriar and A. R. Al-Ali, "Impacts of COVID-19 on Electric Vehicle Charging Behavior: Data Analytics, Visualization, and Clustering," *Applied System Innovation*, vol. 5, no. 1, p. 12, Jan. 2022.

[9] A. Märtz, U. Langenmayr, S. Ried, K. Seddig, and P. Jochem, "Charging Behavior of Electric Vehicles: Temporal Clustering Based on Real-World Data," *Energies*, vol. 15, no. 18, p. 6575, Sep. 2022.

[10] G. Van Kriekinge, C. De Cauwer, N. Sapountzoglou, T. Coosemans, and M. Messagie, "Electric Vehicle Charging Sessions Generator Based on Clustered Driver Behaviors," *World Electric Vehicle Journal*, vol. 14, no. 2, p. 37, Feb. 2023.

[11] G. J. Carlton and S. Sultana, "Electric vehicle charging station accessibility and land use clustering: A case study of the Chicago region," *Journal of Urban Mobility*, vol. 2, p. 100019, Dec. 2022.

[12] Z. J. Lee, T. Li, and S. H. Low, "ACN-Data – A Public EV Charging Dataset," 2021. [Online]. Available: https://ev.caltech.edu/dataset

[13] P. Ferreira, "EV4EU," Jun. 2022. [Online]. Available: https://www.inesc-id.pt/ev4eu-launches-today/

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

[15] P. J. Rousseeuw and K. V. Driessen, "A Fast Algorithm for the Minimum Covariance Determinant Estimator," *Technometrics*, vol. 41, no. 3, pp. 212–223, Aug. 1999.

[16] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation Forest," in *2008 Eighth IEEE International Conference on Data Mining*, Dec. 2008, pp. 413–422, iSSN: 2374-8486.

[17] C. Daake, M. Cammerer, and M. Hackmann, "P3 Charging Index Report 07/22," P3 GROUP GMBH, Tech. Rep. 07/22, 2022. [Online]. Available: http://bit.ly/3KENHGJ

[18] S. Shahriar, A. R. Al-Ali, A. H. Osman, S. Dhou, and M. Nijim, "Machine Learning Approaches for EV Charging Behavior: A Review," *IEEE Access*, vol. 8, pp. 168 980–168 993, 2020.

[19] M. J. Zaki and W. Meira, Jr, *Data Mining and Machine Learning: Fundamental Concepts and Algorithms*. Cambridge Uni. Press, 2020.

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

[21] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society. Series B*, vol. 39, no. 1, pp. 1–38, 1977.

[22] M. Steinbach, G. Karypis, and V. Kumar, "A Comparison of Document Clustering Techniques," University of Minnesota, Report, May 2000.

[23] H. Jia, S. Ding, X. Xu, and R. Nie, "The latest research progress on spectral clustering," *Neural Computing and Applications*, vol. 24, no. 7-8, pp. 1477–1486, Jun. 2014.

[24] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, Nov. 1987.

[25] D. L. Davies and D. W. Bouldin, "A Cluster Separation Measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224–227, Apr. 1979.

[26] T. Calinski and J. Harabasz, "A dendrite method for cluster analysis," *Communications in Statistics*, vol. 3, no. 1, pp. 1–27, 1974.

[27] "scikit-learn 1.3.0 documentation." [Online]. Available: https://scikit-learn.org/stable/index.html#

[28] "OpenCage." [Online]. Available: https://opencagedata.com/

[29] M. Amiruzzaman, R. Rahman, M. R. Islam, and R. M. Nor, "Evaluation of DBSCAN algorithm on different programming languages: An exploratory study," in *2021 5th International Conference on Electrical Engineering*. Dhaka, Bangladesh: IEEE, Nov. 2021, pp. 1–6.

[30] Y. Xiong, B. Wang, C.-C. Chu, and R. Gadh, "Electric Vehicle Driver Clustering using Statistical Model and Machine Learning," in *2018 IEEE PESGM*. Portland, OR: IEEE, Aug. 2018, pp. 1–5.

[31] A. Nespoli, E. Ogliari, and S. Leva, "User Behavior Clustering Based Method for EV Charging Forecast," *IEEE Access*, vol. 11, 2023.

[32] A. Jerónimo, P. Carvalho, C. Jesus, L. Dias, L. M. Ferreira, and H. Morais, "Modeling demand response of Charge Point Operators to consider flexibility in grid planning," in *SEST 2023*. Mugla, Turkey: SEST, Sep. 2023. [Online]. Available: https://bit.ly/3RukHW5

[33] F. Daneshzand, P. J. Coker, B. Potter, and S. T. Smith, "EV smart charging: How tariff selection influences grid stress and carbon reduction," *Applied Energy*, vol. 348, p. 121482, Oct. 2023.