# A Comprehensive Review of Clustering Methods Applications in Electric Mobility

Marcelo Forte*,[a], Cindy P. Guzman[a], Lucas Pereira[b], Hugo Morais[a]

[a]*Department of Electrical and Computer Engineering, INESC-ID—Instituto de Engenharia de Sistemas e Computadores-Investigação e Desenvolvimento, Instituto Superior Técnico (IST), Universidade de Lisboa, Rua Alves Redol, 9, Lisboa, 1000-029, Lisbon, Portugal*
[b] *Interactive Technologies Institute (ITI), Laboratory for Robotics and Engineering Systems (LARSyS), Av. Rovisco Pais, 1, Lisboa, 1049-001, Lisbon, Portugal*

**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 the power systems planning and operation, mainly due to the uncertainty related to EV users' behavior and charging needs. Some of the difficulties motivated by the uncoordinated behavior of EVs are the occurrence of voltage instabilities, system overcurrents, and harmonic distortion. In this context, clustering can help better understand and categorize the behavior of EVs and electric vehicle supply equipment (EVSE) usage, with multiple research studies devoted to the study of clustering methods to offer solutions for these problems. This manuscript comprehensively presents a review of clustering methods applications for electric mobility that focus on the possibility of identifying different groups of EV charging processes, through clustering, to provide support in characterizing EV charging profiles, EV user behavior, and EVSE accessibility and location. For that, we present a roadmap that starts with cluster analysis, in which the most utilized mathematical clustering and validation techniques are detailed. Then, several EV charging datasets are described, followed by a review of research works focusing on clustering applications in EV data, considering three main categories, namely EV charging profiles, EV user behavior, and EVSE accessibility and location.

*Keywords:* Clustering, Electric Vehicles, User Behavior, Charging Profiles, Charging Stations Placement

*Corresponding author
*Email addresses:* marcelojsforte@tecnico.ulisboa.pt (Marcelo Forte*), cindy.lascano@tecnico.ulisboa.pt (Cindy P. Guzman), lucas.pereira@tecnico.ulisboa.pt (Lucas Pereira), hugo.morais@tecnico.ulisboa.pt (Hugo Morais)

## 1. Introduction

The adoption of electric vehicles (EVs) has experienced rapid growth in the 21st century, driven by the pressing need to transition global energy demand away from fossil fuels, particularly within the past decade [30]. People are facing a dramatic transformation in their lifestyle to become carbon neutral, with the United Nations (UN) placing the fight against climate change under one of the goals of Sustainable Development [1]. At the 2015 United Nations Climate Change Conference (COP 21), 196 countries reached the first-ever universal and legally binding climate change agreement. This agreement sets out a worldwide action plan to "limit global warming to well below 2 °C, preferably to 1.5 °C, compared to pre-industrial levels" [2]. This ambitious plan requires a significant reduction in greenhouse gas (GHG) emissions.

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 [3], since Transport is the only sector where greenhouse gas (GHG) emissions have increased in the past three decades [4]. This sector was responsible for more than a quarter of Europe's energy consumption 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 [5]. In addition to GHG, burning fossil fuels, whether in power plants or in internal combustion engine vehicles (ICEVs), releases harmful pollutants that can significantly degrade air quality.

To achieve carbon neutrality, in 2022 the EU's environment ministers approved the "Fit for 55 in 2030" package [6], 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 [7], 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 adoption of EVs [8], whose popularity has significantly increased over the past five years [9].

Even though the production and disposal of EVs are currently less eco-friendly than those of an ICEV (mainly due to the production of its batteries [10]), an analysis of the entire life cycle of an EV shows that it is still cleaner than an ICEV, as revealed by Zhang et al. [11]. Their study demonstrates that EVs could potentially provide a 45% reduction in GHG emissions compared to ICEVs, considering the energy cost of production, assembly, transportation, and usage (the authors assumed 300 000 km as the average lifetime of a passenger vehicle). As the share of electricity from renewable energy sources (RES) is set to increase in the future, as well as making batteries more sustainable, EVs should become even less harmful to the environment [12]. According to the World Energy Transitions Outlook 2023 [13], the share of RES in electricity generation should increase from 28% in 2020 to 91% in 2050.

Due to all these factors, the number of EVs will certainly increase in the upcoming years. Therefore, reliable control and understanding of the charging process of an EV will be essential for its successful penetration into the power system [14, 15]. There is a broad consensus that the crucial factor is not the increased energy demand but rather the potential load peaks resulting from many simultaneous EV charging processes. Uncoordinated EV charging has negative impacts on the existing power grid, including high load peaks, higher energy use, and degradation of power quality [16, 18, 19].

Energy system modelers (ESMs), distribution system operators (DSOs), utilities, and urban planners need to quantify the impacts on grid infrastructure and network reinforcement demands to address future challenges and opportunities associated with EV mass adoption [20]. The identification of charging profiles is of great importance if these entities are to realize the intelligent and successful integration of EVs into the energy system. Clustering methods represent one of the most effective approaches for identifying these profiles, but there is limited coverage of these techniques in existing literature.

Most existing review papers do not focus on charging behavior from a data-driven approach, which hinders prompt identification and comparison of different methodologies. For example, Shafiei and Ghasemi-Marzbali [21] give a comprehensive review of fast charging stations for EVs and the challenges related to them, such as locating and determining their optimal capacity, the problem of energy storage, and the overall management system. Al-Ogaili et al. [22] review scheduling, clustering, and forecasting strategies for EV charging. They also present a review of data-driven and other approaches, such as optimization, and how they have been used for EV charging strategies in the literature. Shahriar et al. [23], on the other hand, solely focused on reviewing the existing machine learning (ML) approaches, including supervised, unsupervised (namely clustering studies), and deep learning, used in the analysis and prediction of EV charging behavior currently in the literature. Andrenacci and Valentini [24] review the factors influencing EV charging behavior, emphasizing the importance of understanding these behaviors for infrastructure planning and energy demand forecasting. Despite mentioning key influences already identified in the literature, including mobility patterns, socio-economic factors, and infrastructure availability, the paper does not focus on comparing methodologies to obtain this charging behavior information. Furthermore, while clustering methods are mentioned as helpful in determining charging patterns, it lacks a detailed analysis of these techniques for EV charging data. Perhaps the most related work to the one proposed in this article is presented in [25], where Nazari et al. suggest that clustering can be used to address the problem of uncertain introduction of EV load in the power system. The authors review recent literature on the application of clustering methods based on the user's behavior, driving cycle, batteries, and charging stations.

The main objective of this study differs from the previously mentioned ones as it aims to analyze the current literature on the application of clustering methods for EVs that focus on the possibility of identifying different groups of EV charging processes, through clustering, to provide support in characterizing EV charging profiles, EV user behavior, and EVSE accessibility and optimal location. The studies referenced and presented in the following subsections result from a complex research work where priority was given to the most recent, relevant, and accurate studies on each topic. The Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) guidelines [26] were adopted to conduct the present systematic review. The majority of papers reviewed were discovered using advanced online article search tools, namely Google Scholar, IEEE Xplore, and Science Direct, by combining the keywords *"EV charging profiles"*, *"EV user behavior profiles"*, *"Electric mobility profiles"*, *"EVSE optimal location"*, and *"EVSE accessibility"*, with *"Clustering"*. Some studies have also been found through citations in the aforementioned analyses [21, 22, 23, 25]. Considering the number of papers discovered using the mentioned keywords, no selection process was required. The present study will help answer the research questions: What are the most appropriate clustering methodologies to characterize EVs and EVSEs? Do EV-related public datasets exist that allow benchmarking clustering methodologies? What are the most interesting applications of clustering in the EV domain?

The paper is organized as follows. Section 2 provides an overview of the history and present state of EVs and the EV charging process. Section 3 reviews the most commonly applied clustering techniques in the literature in the context of EV charging, and Section 4 describes the characteristics of the most well-known EV open datasets. Section 5 reviews the applications of clustering techniques on EV data. Finally, Section 6 discusses the shortcomings of existing studies, and future research recommendations are presented.

## 2. Background

### *2.1. History & Current State of EVs*

It is hard to pinpoint the invention of the electric car to one inventor or country. Instead, a series of 19th-century breakthroughs in batteries and motors led to the first EV on the road. Around 1881, French engineer Gustave Trouve reportedly created the first battery-powered EV, a 160 kg tricycle [27]. The taxi "Electroboat" was the first EV in the USA, introduced by William Morris in 1889. By 1900, 38% of vehicles sold in the USA were electric [28], but their popularity declined with the mass production of the accessible gasoline-powered Model T by Henry Ford in 1908. Renewed interest in EVs emerged after the 1973 Arab oil crisis encouraged the search for alternative energy sources [29]. Since then, automakers have been developing prototypes in response to new transportation emissions restrictions. With the Paris Agreement in 2015, reducing GHG emissions has been a priority, and EVs became part of the solution. Today, almost all automakers offer at least one EV model.

Recent EV sales reports – including battery electric vehicles (BEVs) and plug-in hybrid electric vehicles (PHEVs) – reveal that the EV stock has increased exponentially in recent years [30], especially BEVs. In 2024, global EV sales achieved 17.5 million units (up 25% relative to 2023), representing around 20% of the market share, and the EV car stock represented 4% of the global fleet [30]. EV sales in Europe stagnated in 2024, with 3.3 million EVs sold, as political subsidies have decreased in the main car markets, such as Germany and France. The European market for EVs represents 24% (14.1 million) of the global EV stock [30]. In Stated Policies Scenario (STEPS) [31], it is foreseen a high growth in global EV sales, possibly reaching a 58% sales share and 15% stock market share in 2030 [30].

Although there are different technologies available to power electric motors, battery packs are the primary power source for these new EVs [28]. Nowadays, BEVs can travel up to 700 km on a single charge, unlike early models that would often last less than 100 km due to battery constraints. The capacity of the batteries, the overall efficiency of the EV, and the management strategies directly impact the range [32]. There is an urgent demand to build charging stations to meet the needs of drivers. According to [30], the number of public EVSEs reached more than 5.45 million in 2024, of which around one-third were fast chargers.

The “Fit for 55” package [6] introduced a regulation to set mandatory national targets for EU member states to deploy publicly accessible alternative fuels infrastructure, the Alternative Fuels Infrastructure Regulation (AFIR) [33]. Article 3 mandates 1.3 kW of publicly available charger per BEV and 0.8 kW per PHEV starting in 2024, a target already met by the EU average, since the member states combined had 2.6 kW/EV at the end of 2024. However, there is still a scarcity of charging ports, as demand has been increasing every year [30]. These advancements in smart charging technologies have enabled the exploration of previously inaccessible topics. EVSEs generate a wide range of charging data, allowing for detailed user behavior analysis and the development of charging management programs for various applications. Specifically, clustering techniques applied to empirical charging data have identified distinct usage patterns, contributing to more accurate demand modeling and more efficient infrastructure planning [22].

In addition to light-duty personal vehicles, electrification is reaching other categories, including two or three-wheelers, commercial, and heavy-duty vehicles [35]. In fact, 27 nations (including the USA and EU) have committed to achieving 100% sales of zero-emission buses and trucks by 2040 [36].

### *2.2. Negative Impacts of Uncoordinated EV Charging*

As mentioned, the accelerated adoption of EVs represents a significant shift towards sustainable transportation. However, without effective management, uncoordinated EV charging (where the charging process occurs without considering the grid capacity or renewable energy availability) can severely impact the electrical grid, economic stability, and environmental goals [16, 17]. Identifying these aspects is essential for understanding the need for clustering methods and their applications in real-world operational scenarios.

#### *2.2.1. Grid Stress and Increased Peak Demand*

Uncoordinated EV charging often aligns with peak residential electricity demand, typically when drivers return home in the evening. This circumstance can lead to a dramatic increase in peak load, deepening grid stress, and potentially leading to infrastructure overloads, voltage instability, and an increased probability of blackouts. As EV adoption rises, these issues could require costly grid upgrades to prevent disruptions.

A study by Jones et al. [16] examined the impacts of uncontrolled EV charging on various types of distribution feeders, including residential, commercial, and industrial, using real-world data and simulations. The study found that in a home-dominant charging scenario, where most EVs recharge during evening hours, peak loads on residential feeders increased significantly, leading to thermal overloading of power lines by as much as 15%. The study ends by emphasizing the need for smart-charging strategies to distribute the load more evenly and reduce the risk of grid overload.

#### *2.2.2. Increased Greenhouse Gas Emissions*

Uncoordinated charging can result in increased GHG emissions when the process occurs during peak hours, which often require the grid to rely on fossil fuel-based power generation. The higher carbon intensity of the grid during these times can offset some of the environmental benefits of EVs.

A study by Kang et al. [37] highlights the delicate balance between reducing CO2 emissions and managing peak power demand in EV charging coordination. The study found that while coordinated charging can reduce annual CO2 emissions by up to 18%, it may also create new peaks in power demand. This underscores the potential for increased greenhouse gas emissions if charging is not strategically managed, particularly given the prevalence of higher-emission sources in the energy mix during peak periods.

These findings reveal the importance of smart-charging strategies that consider both the grid’s carbon intensity and demand patterns to maximize the environmental benefits of EV adoption. These demand patterns can be obtained through clustering methods, specifically EV charging and EV behavior profiles.

### 2.2.3. *Economic Costs and Market Disruption*

The economic implications of uncontrolled EV charging are significant, as they can lead to higher consumer electricity prices. This is mainly due to the need for utilities to activate peaker plants during periods of high demand, which are expensive to operate and typically produce higher carbon emissions [38].

The study by von Bonin et al. [39] analyzed the effects of dynamic tariffs and photovoltaic (PV) incentives on EV charging behavior. Optimized charging strategies, including dynamic tariffs and PV-based incentives, demonstrated cost reductions of up to 33.7% for households. Similarly, the EV4EU D4.5 report [40] evaluated various demand response (DR) programs across Greece, Portugal, and Slovenia. Programs incorporating real-time pricing and time-of-use tariffs revealed substantial cost savings for EV users.

Both studies indicated that uncoordinated user responses to price signals could still result in local peaks and potential market disruptions, as synchronized charging behaviors could exacerbate peak loads, driving up electricity prices and contributing to market volatility. In addition, both studies required charging and behavior profiles that served as the basis for defining dynamic tariffs, proving the necessity and applicability of these types of profiles.

### 2.2.4. *Challenges in Renewable Energy Integration*

One of the potential benefits of EVs is their ability to support the integration of RES by acting as a flexible load that can absorb excess generation [18]. However, uncontrolled EV charging can exacerbate issues of renewable energy curtailment, where excess generation cannot be effectively used or stored due to mismatched demand.

Amiruddin et al. [41] found that optimized EV charging patterns could further enhance renewable integration, reducing the need for battery storage by 84%, cutting emissions by 23.7%, and increasing renewable energy penetration by 10%. These findings underscore the critical importance of developing smart charging strategies and V2G capabilities to align EV demand with renewable generation, mitigating integration challenges, and maximizing environmental and economic benefits. Once again, typical charging and user behavior profiles play a pivotal role in these strategies, with clustering allowing for straightforward and comprehensive obtaining of these profiles.

## 3. Cluster Analysis

In the literature, cluster analysis has received a lot of attention and has been researched extensively. There are papers such as [42], published in 1969, that helped to investigate and develop various mathematical clustering and classification techniques. Nevertheless, it is important to first give a brief introduction.

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** [43]. 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 that is applied in a variety of domains, including pattern recognition, image analysis, bioinformatics, data dimensionality reduction, and machine learning [44]. **Figure 1** provides an illustration of clustering.

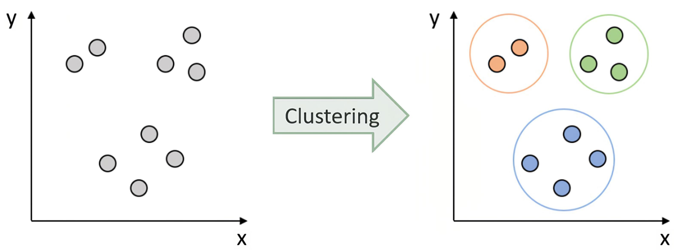


Figure 1: Simple illustration of clustering in two dimensions (Adapted from [45]).

The term “cluster” does not have a universally accepted definition, since various interpretations are used for different analytical purposes and reflect the diverse structures of data, contexts of problems, and application areas. Consequently, numerous clustering methods have been developed, including representative-based, hierarchical, density-based, and spectral (or graph) clustering. In this paper, the notation and nomenclature follow the ones defined by Zaki and Meira [43], presented in **Figure 2**. In the following subsections, the most commonly employed methods for finding EV charging profiles, EV user behavior profiles, and EVSE accessibility and location are comprehensively described, providing a basis for understanding the methodologies defined by the studies discussed in Section 5.

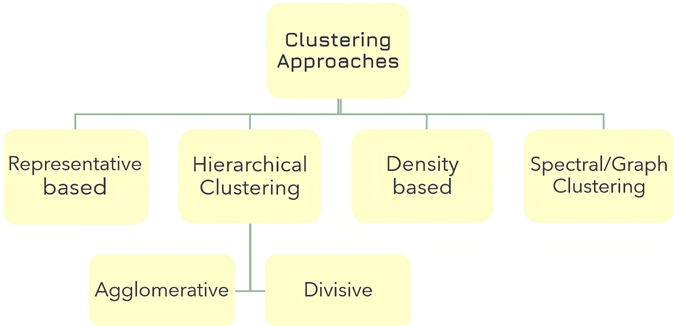


Figure 2: A summary of clustering methods (Based on [43]).

### *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 [46] 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 [47] generalizes K-means by modeling the data as a mixture of normal distributions (Gaussian Mixture Model (GMM)) and maximizing the likelihood of the data to discover the cluster parameters (the mean and covariance matrix). It is a soft clustering approach since it returns the probability of a point belonging to each cluster.

#### *3.1.1. K-means Clustering*

The goal of K-means (**Algorithm 1**) is to find a clustering that minimizes the Sum of Squared Errors (SSE) score, which measures the accuracy or goodness of the clustering, defined as

$$SSE(C) = \sum_{i=1}^{k} \sum_{x_j \in C_i} \|x_j - \mu_i\|^2, \tag{1}$$

where $\boldsymbol{x}_j \in \mathbb{R}^d$ is a point from a given dataset $\mathbf{D}^{n \times d}$ and $\boldsymbol{\mu}_i \in \mathbb{R}^d$ is the centroid of the cluster $C_i$.

As stated in the pseudo-code **Algorithm 1**, the points are initially assigned to the clusters at random, with the integer $k$ being the number of clusters. The **elbow method** is typically used to determine the optimal $k$ [48]. The points are then iteratively assigned to new centroids based on how close they are (line 4). In each iteration, the centroids are updated based on the mean of the assigned points (line 7). The process repeats until the centroids stop changing (defined by a threshold), and the algorithm converges.

K-means is typically run multiple times, with the run with the lowest SSE value being selected to report the final clustering. This happens because the method begins with a random guess for the initial centroids. In terms of computational complexity, from **Algorithm 1** and assuming $t$ iterations, the total time for K-means is given as $O(tnkd)$.

**Algorithm 1: K-means**

**Input:** $(\mathbf{D}, k, \epsilon)$

Initialize the cluster centroids $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_k \in \mathbb{R}^d$ randomly

**repeat**

  **foreach** *data point* $\boldsymbol{x}_j$ **do**

    calculate distance and assign each $\boldsymbol{x}_j$ to the closest $\boldsymbol{\mu}_i$:

$$C_i := arg\min_i \|\boldsymbol{x}_j - \boldsymbol{\mu}_i\|^2$$

  **end**

  **foreach** *cluster* $C_i$ **do**

    compute and update centroids for each cluster:

$$\boldsymbol{\mu}_i := \frac{1}{|C_i|} \sum_{\boldsymbol{x}_j \in C_i} \boldsymbol{x}_j$$

  **end**

**until** $\sum_{i=1}^{k} \|\boldsymbol{\mu}_i^t - \boldsymbol{\mu}_i^{t-1}\|^2 \leq \epsilon$ ;

*3.1.2. Expectation-Maximization Clustering*

Given $n$ points $\boldsymbol{x}_j$ in a $d-$dimensional space, let $\mathbf{X}$, $\mathbf{X} = (X1, X2, \ldots, Xd)$, be the vector random variable across the $d-$attributes. EM (**Algorithm 2**) assumes that each cluster $C_i$ is characterized by a multivariate normal distribution

$$f(\boldsymbol{x} \,|\, \boldsymbol{\mu_i}, \boldsymbol{\Sigma_i}) = \frac{1}{(2\pi)^{(d/2)}|\boldsymbol{\Sigma_i}|^{1/2}} \cdot \exp\left\{-\frac{1}{2}\,(\boldsymbol{x} - \boldsymbol{\mu_i})^T\, \boldsymbol{\Sigma_i}^{-1}\,(\boldsymbol{x} - \boldsymbol{\mu_i})\right\}, \tag{2}$$

where the cluster $C_i$ centroid $\boldsymbol{\mu}_i \in \mathbb{R}^d$ and the covariance $\boldsymbol{\Sigma_i} \in \mathbb{R}^{d \times d}$ are both unknown parameters and $f(\boldsymbol{x} \,|\, \boldsymbol{\mu_i}, \boldsymbol{\Sigma_i})$ is the probability density at $\boldsymbol{x}$ attributable to cluster $C_i$.

A Gaussian Mixture Model over all the $k$ clusters defines the probability density function of $\mathbf{X}$, given as

$$f(\boldsymbol{x}) = \sum_{i=1}^{k} f(\boldsymbol{x} \,|\, \boldsymbol{\mu_i}, \boldsymbol{\Sigma_i}) P(C_i), \tag{3}$$

where the prior probabilities $P(C_i)$ satisfy $\sum_{i=1}^{k} P(C_i) = 1$.

Thus, the Gaussian Mixture Model is characterized by the mean $\boldsymbol{\mu}_i$, the covariance $\boldsymbol{\Sigma}_i$, and the *mixture parameters* for each of the $k$ clusters, written compactly as

$$\boldsymbol{\theta} = \{\boldsymbol{\mu}_1, \boldsymbol{\Sigma_1}, P(C_i), \ldots, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, P(C_k)\}. \tag{4}$$

After all the key points described, moving forward is thus doable. The goal of EM is to find the maximum likelihood estimates for the parameters $\boldsymbol{\theta}$. To achieve that, EM executes a two-step iterative algorithm that starts from an initial guess for the parameters $\boldsymbol{\theta}$.

In the **Expectation Step**, given the current estimates for $\boldsymbol{\theta}$, EM computes the cluster posterior probabilities through the Bayes theorem

$$\begin{aligned} w_{ij} = P(C_i|\boldsymbol{x}_j) &= \frac{P(\boldsymbol{x}_j|C_i)P(C_i)}{\sum_{a=1}^{k} P(\boldsymbol{x}_j|C_a)P(C_a)} \\ &= \frac{f_i(\boldsymbol{x}_j)P(C_i)}{\sum_{a=1}^{k} f_a(\boldsymbol{x}_j)P(C_a)}, \end{aligned} \tag{5}$$

since each cluster is modeled as a multivariate normal distribution [43]. Therefore, $P(C_i|\boldsymbol{x}_j)$ can be considered the weight contribution of $\boldsymbol{x}_j$ to cluster $C_i$.

Next, in the **Maximization Step**, EM recalculates $\boldsymbol{\theta}$ using the weights $w_{ij}$, as can be seen in **Algorithm 2**. The algorithm ends when $\sum_{i=1}^{k} \|\boldsymbol{\mu}_i^t - \boldsymbol{\mu}_i^{t-1}\|^2 \leq \epsilon$, where $\epsilon$ is the convergence threshold, and $t$ denotes the iteration.

**Algorithm 2: Expectation-Maximization**

**Input:** $(\mathbf{D}, k, \epsilon)$

Initialise centroids $\mu_1, \mu_2, \ldots, \mu_k \in \mathbb{R}^d$ randomly

$\mathbf{\Sigma}_i \leftarrow \mathbf{I},\ P(C_i) \leftarrow \frac{1}{k},\ \forall i = 1, \ldots, k$

**repeat**

  **for** *i = 1, . . . , k and j = 1, . . . , n* **do**

    Expectation Step (calculate posterior probability):

$$w_{ij} := \frac{f_i(\boldsymbol{x}_j)P(C_i)}{\sum_{a=1}^{k} f_a(\boldsymbol{x}_j)P(C_a)}$$

  **end**

  **for** *i = 1, . . . , k* **do**

    Maximization Step (recalculate $\boldsymbol{\theta}$):

$$\boldsymbol{\mu}_i := \frac{\sum_{j=1}^{n} w_{ij} \cdot \boldsymbol{x}_j}{\sum_{j=1}^{k} w_{ij}}$$

$$\mathbf{\Sigma}_i := \frac{\sum_{j=1}^{n} w_{ij}(\boldsymbol{x}_j - \mu_i)(\boldsymbol{x}_j - \mu_i)^T}{\sum_{j=1}^{k} w_{ij}}$$

$$P(C_i) := \frac{\sum_{j=1}^{n} w_{ij}}{n}$$

  **end**

**until** $\sum_{i=1}^{k} ||\boldsymbol{\mu}_i^t - \boldsymbol{\mu}_i^{t-1}||^2 \leq \epsilon$ ;

For the Expectation Step, inverting $\mathbf{\Sigma}_i$ and computing its determinant takes $O(kd^3)$, and evaluating the density $f_i(\boldsymbol{x})$ takes $O(nkd^2)$. For the Maximization Step, the time is dominated by the $\mathbf{\Sigma}_i$ update. Assuming $t$ iterations, the computational complexity of the EM method is $O(t(kd^3 + nkd^2))$.

### *3.2. Hierarchical Clustering*

**Hierarchical Clustering** techniques create 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) 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 own 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 to get Hierarchical clusters [49]:

- **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. Chameleon is a well-known example [50];
- **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.

The Agglomerative approach is by far the most widely used in the literature. Thus, it will be examined in greater depth next.

#### *3.2.1. Agglomerative Hierarchical Clustering*

Agglomerative Hierarchical Clustering starts with each of the $n$ points in a separate cluster. Then, the two closest clusters are repeatedly merged until all points are members of the same cluster, as shown in the pseudo-code given in **Algorithm 3**. Given a set of clusters $C = \{C_1, C_2, \ldots, C_m\}$, first, the closest pair of clusters $C_i$ and $C_j$ are found and merged into a new cluster, $C_{ij}$. Next, the set of clusters is updated, removing $C_i$ and $C_j$ and adding $C_{ij}$. This process is repeated until $C$ contains exactly $k$ clusters.

**Algorithm 3: Agglomerative H. Clustering**

**Input:** $(\mathbf{D}, k)$

Initialize each cluster with a single point $C \leftarrow C_i = \{\boldsymbol{x}_i\}, \forall i = 1, \ldots, n$

Compute the distance matrix $\boldsymbol{\Delta} \leftarrow \|\boldsymbol{x}_i - \boldsymbol{x}_j\|, \forall i = 1, \ldots, n\,;\ \forall j = 1, \ldots, n$

**repeat**

  Find the closest pair of clusters: $C_i, C_j \in C$

  Merge clusters $C_{ij} \leftarrow C_i \cup C_j$

  Update $C \leftarrow (C \setminus \{C_i, C_j\} \cup \{C_{ij}\}$ and $\boldsymbol{\Delta}$ to reflect new clustering

**until** $|C| = k$;

Finding the closest pair of clusters is the algorithm's key step. For this, a variety of distance measures can be employed [51] (see **Figure 3**), including:

- **Single link**: The distance between two clusters is defined as the minimum distance between a point in $C_i$ and a point in $C_j$. First developed by Florek et al. [52] and then independently by McQuitty (1957) and Sneath (1957) [53];

- **Complete link**: The distance between two clusters is defined as the maximum distance between a point in $C_i$ and a point in $C_j$. Developed by Sørenson in 1948 [54];

- **Average link**: The distance between two clusters is defined as the average pairwise distance between points in $C_i$ and $C_j$. Developed by Sokal and Michener (1958) [55] to avoid the extremes introduced by either single or complete link;

- **Mean distance**: The distance between two clusters is defined as the distance between the centroids of the two clusters. The earliest use known of this strategy is that of Sokal and Michener (1958) [55].

But possibly the most employed measure is **Ward's Method**, introduced by Joe H. Ward, Jr. in 1963 [56]. The distance between two clusters is defined as the increase in the sum of squared errors when the two clusters are merged. The objective is to minimize the total within-cluster variance. It can be seen as a weighted version of the mean distance measure, as it weights the distance between centroids by half of the harmonic mean of the cluster size.

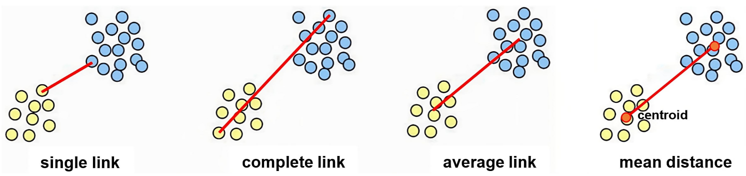


Figure 3: Different distance measures (Adapted from [57]).

When two clusters $C_i$ and $C_j$ combine to form $C_{ij}$, the distances between $C_{ij}$ and each of the remaining clusters $C_r (r \neq i, r \neq j)$ must be updated in the matrix $\boldsymbol{\Delta}$. For all of the cluster proximity measures, the **Lance-Williams** [58] formula offers a general equation to recompute the distances:

$$\delta(C_{ij}, C_r) = \alpha_i \cdot \delta(C_i, C_r) + \alpha_j \cdot \delta(C_j, C_r) + \beta \cdot \delta(C_i, C_j) + \gamma \cdot |\delta(C_i, C_r) - \delta(C_j, C_r)|, \tag{6}$$

where the parameters $\alpha_i, \alpha_j, \beta$ and $\gamma$ differ from one measure to another [58].

In terms of computational complexity, Agglomerative clustering initially takes $O(n^2)$ time to create the distance matrix $\boldsymbol{\Delta}$, and updating/deleting distances from it takes $O(\log(n))$ time for each operation, leading to a total of $O(n^2 \log(n))$.

### 3.3. Density-based clustering

**Density-based clustering** methods use the density or connectedness properties to find nonconvex clusters. This type of clustering 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) [59], analyzed in greater depth next.

#### 3.3.1. DBSCAN

Unlike representative-based clustering that can only discover ellipsoid-shaped or convex clusters, DBSCAN [59] is a density-based clustering method, therefore can find nonconvex clusters. It uses the local density of points rather than using only the distance between points to determine the clusters. The neighborhood of $\boldsymbol{x} \in \mathbb{R}^d$ is defined as

$$N_\epsilon(\boldsymbol{x}) = \delta(\boldsymbol{x}, \boldsymbol{y}) \leq \epsilon\,, \tag{7}$$

where $\delta(\boldsymbol{x}, \boldsymbol{y})$ represents the distance between points $\boldsymbol{x}$ and $\boldsymbol{y}$ (usually Euclidean distance, but might be other metrics). The threshold $\epsilon$ needs to be specified.

In order to fully understand the algorithm, it is first necessary to define some important concepts. $\boldsymbol{x}$ is a *core point* if there are at least *minpts* points in its $\epsilon$-neighbourhood ($N_\epsilon(\boldsymbol{x}) \geq minpts$, with *minpts* a user-defined threshold). A *border point* does not meet the *minpts* threshold, but it belongs to the $\epsilon$-neighbourhood of another point $z$, $\boldsymbol{x} \in N_\epsilon(z)$. Finally, a *noise point* is neither a core nor a border point (outlier). $\boldsymbol{x}$ is *density reachable* from $\boldsymbol{y}$ if there is a set of core points leading from $\boldsymbol{y}$ to $\boldsymbol{x}$. Two points $\boldsymbol{x}$ and $\boldsymbol{y}$ are *density connected* if there exists a core point $z$ such that both $\boldsymbol{x}$ and $\boldsymbol{y}$ are density reachable from $z$. The pseudo-code for the DBSCAN method is shown in **Algorithm 4**.

**Algorithm 4: DBSCAN**

1 **DBSCAN** ($\boldsymbol{D}, \epsilon, minpts$)**:**
2     $Core \leftarrow \emptyset, k \leftarrow 0$ `// core points and cluster id`
3     **foreach** $\boldsymbol{x}_i \in \boldsymbol{D}$ **do** `// find core points`
4         Compute $N_\epsilon(\boldsymbol{x}_i)$
5         $id(\boldsymbol{x}_i) \leftarrow \emptyset$
6         **if** $N_\epsilon(\boldsymbol{x}_i) \geq minpts$ **then** Core $\leftarrow$ Core $\cup\{\boldsymbol{x}_i\}$
7     **end**
8     **foreach** $\boldsymbol{x}_i \in$ *Core, with* $id(\boldsymbol{x}_i) = \emptyset$ **do**
9         $k \leftarrow k + 1$
10         $id(\boldsymbol{x}_i) \leftarrow k$ `// assign` $\boldsymbol{x}_i$ `to cluster id` $k$
11         **DENSITYCONNECTED**($\boldsymbol{x}_i, k$)
12     **end**
13     $C \leftarrow \{C_i\}_{i=1}^k, C_i \leftarrow \{\boldsymbol{x} \in \boldsymbol{D} \mid id(\boldsymbol{x}) = i\}$ `// define clusters`
14     $Noise \leftarrow \{\boldsymbol{x} \in \boldsymbol{D} \mid id(\boldsymbol{x}) = \emptyset\}$
15     $Border \leftarrow \boldsymbol{D} \setminus \{Core \cup Noise\}$

16 **DENSITYCONNECTED** ($\boldsymbol{x}, k$)**:**
17     **foreach** $\boldsymbol{y} \in N_\epsilon(\boldsymbol{x})$ **do**
18         $id(\boldsymbol{y}) \leftarrow k$ `// assign` $\boldsymbol{y}$ `to cluster` $k$
19         **if** $\boldsymbol{y} \in Core$ **then**
20         DENSITYCONNECTED ($\boldsymbol{y}, k$)
21     **end**

First, DBSCAN computes the $\epsilon$-neighborhood for each point $\boldsymbol{x}_i$, checks if it is a core point (lines 3–7), and sets the cluster *id* null for all points. Next, starting from each unassigned core point, the method finds all its density connected points recursively, which are assigned to the same cluster (line 11). Some border points may be accessible from core points in more than one cluster. As DBSCAN is a sequential algorithm, they will be arbitrarily assigned to the first created cluster that can incorporate that specific border point.

Regarding the computational complexity, it takes $O(n^2)$ to compute the neighborhood for each point when the dimensionality is high. With $N_\epsilon(\boldsymbol{x})$ computed, the algorithm needs only a single pass over all points to find the density connected clusters, leading to the overall complexity $O(n^2)$.

### *3.4. Spectral/Graph clustering*

The goal of **Graph clustering** is to cluster the nodes by using the edges and their weights, which represent the similarity between the incident nodes. 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 from the kernel matrix, such as the adjacency matrix and Laplacian matrix.

The concept of spectral clustering is based on spectral graph theory. It approaches the data clustering problem as a graph partitioning problem, building an undirected weighted graph with each point in the dataset as a vertex and the similarity value between any two points as the weight of the edge linking the two vertices [60]. The graph can then be split into connected components using a specific graph cut method, and those components are referred to as clusters.

In the literature, at the time of writing, this method is not employed for the identification of EV user behavior and charging profiles, although some studies for EVSE accessibility and location utilized this method.

**Table 1** summarizes the advantages and disadvantages of the previously described clustering methods.

Table 1: Advantages and disadvantages of the aforementioned clustering methods.

| **Clustering Method** | **Advantages** | **Disadvantages** |
|---|---|---|
| K-means [46] | Most widely used method since it is simple to implement, highly flexible, easy to adapt to different types of data, scales to large datasets, and generalizes to different shapes and sizes of clusters | Requires specifying the number of clusters a priori; it can be sensitive to outliers and as the number of dimensions increases, its scalability decreases; favors spherical or round clusters; sensitive to the choice of the initial cluster centers |
| EM [47] | It has a wide range of applications, but it is best recognized in ML for its usage in unsupervised learning tasks such as density estimation and clustering; the Expectation and Maximization Steps are often easily implemented | It can be sensitive to initialization values; it converges to the local optimum only, with slow convergence; as K-means, it can be challenging to determine the number of clusters |
| Agglomerative H. Clustering [49] | *Dendrogram* allows seeing the progressive grouping of the data, presenting a visual illustration of the clustered data; it is robust since it does not require, a priori, the number of clusters that can be chosen later | Higher time complexity can be a problem in larger datasets, therefore works better with small datasets; it is sensitive to noise and outliers; it has a wide range of potential distance measures, making its application less straightforward and simple |
| DBSCAN [59] | Suitable for handling big datasets with noise; able to locate clusters of different densities and shapes; does not require specifying the number of clusters a priori | Need to select additional parameters, such as *minpts* and $\epsilon$, that affect the outcomes; poor scalability, leading to inferior clustering if the data density is not uniform; computationally expensive |
| Spectral/Graph Clustering [60] | It can handle large datasets and high-dimensional data with many features; does not make strong assumptions about the shape of the clusters | It is computationally expensive and sensitive to the choice of the similarity metric, just like Agglomerative H. Clustering; requires the number of clusters a priori |

### *3.5. Time Series clustering*

The previously mentioned algorithms are the best known and typically used in the literature as they allow performing analysis on static data, i.e., data that is not a function of time (which corresponds to most available datasets). However, time series clustering has significant potential in the electric mobility domain, particularly for analyzing and understanding patterns in EV usage, charging behavior, and mobility trends over time. By grouping similar temporal data sequences, such as daily charging loads, driving patterns, or fleet utilization rates, time series clustering enables more informed decision-making for infrastructure planning, demand forecasting, and energy management [61]. The **dynamic time warping** (DTW) [62] metric is the most popular distance measure for clustering time series data since it measures the similarity between two temporal sequences that do not align perfectly in time, speed, or length.

Time series data is increasingly attractive in data analytics due to the expanding deployment of smart meters [63], but is not commonly explored in the literature in the context of EV charging data. Nevertheless, for more information, see [64], where other techniques besides DTW are described.

### 3.6. Clustering Validation Techniques

In the context of identifying charging behavior patterns of EV users and the optimal location of EVSEs, there is no access to ground-truth partitioning, since the work consists precisely in finding these patterns from the data. Therefore, internal validation should be used to quantify the performance of the clustering [43]. Some of the most extensively applied methods in the literature, such as the Silhouette coefficient [65], the Davies-Bouldin index [66], and the Calinski-Harabasz index [67], may be utilized to study EVSE location, EV user behavior, and charging profiles.

#### 3.6.1. Silhouette Coefficient

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

$$s_i = \frac{\mu_{out}^{min}(\boldsymbol{x}_i) - \mu_{in}(\boldsymbol{x}_i)}{\max\{\mu_{out}^{min}(\boldsymbol{x}_i), \mu_{in}(\boldsymbol{x}_i)\}}, \tag{8}$$

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 point in its own cluster.

The total **Silhouette coefficient** [65] is defined as the mean $s_i$ value across all points, given by (9), where a value close to +1 denotes good clustering.

$$SC = \frac{1}{n}\sum_{i=1}^{n} s_i \tag{9}$$

#### 3.6.2. Davies-Bouldin Index

The Davies-Bouldin index provides the average similarity between clusters, where similarity is a metric that compares the distance between clusters with cluster size. The Davies-Bouldin measure for a pair of clusters $C_i$ and $C_j$ is defined as

$$DB_{ij} = \frac{\sigma_{\mu_i} + \sigma_{\mu_j}}{\delta(\mu_i, \mu_j)}, \tag{10}$$

where $\mu_i$ 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** [66] is thus defined as

$$DB = \frac{1}{k} \cdot \sum_{i=1}^{k} \max_{i \neq j}\{DB_{ij}\}, \tag{11}$$

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).

#### 3.6.3. Calinski-Harabasz Index

The **Calinski-Harabasz index** [67] is defined as the ratio between the within-cluster dispersion and the between-cluster dispersion. It is given by

$$CH(k) = \frac{tr(\boldsymbol{S}_B)}{tr(\boldsymbol{S}_W)} \cdot \frac{n-k}{k-1}, \tag{12}$$

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. Thus, the Calinski-Harabasz index can also help in selecting the $k$ that maximizes $CH(k)$, an alternative to the elbow method typically used for K-means [48].

## 4. EV Charging Datasets

There is no cluster analysis without a dataset. Therefore, it is essential to have an adequate EV charging dataset. Amara-Ouali et al. [68] perform an outstanding study of the best EV open data available, providing the community with a structured list of open datasets ready to foster data-driven research in this field. Furthermore, Calearo et al. [69] present a review of data sources for EVs, categorized into different classes by the type of data and its availability.

Based on these papers, the **ACN-Data** dataset [70] was found to be one of the most widely used in the literature in the context of EV data analysis. Zachary J. Lee, Tongxin Li, and Steven H. Low are responsible for the public release of this dataset. In [71], the authors describe the characteristics of the dataset, how they managed to get the data, and proved that this dataset has several possible applications, including clustering analysis of EV charging data using GMM. At the time of writing, ACN-Data has 31424 EV charging sessions found. The first session was in Apr 2018, and the last was in Sep 2021. Data are broken out by charging station name/identification, transaction date, transaction start/end time, energy use, and customer ID.

Another dataset found in [68] is from the city of Boulder, Colorado. **Boulder** dataset [72] has data from Jan 2018 until Sep 2023 and has 148136 EV charging sessions (1 row of the dataset, 1 EVSE transaction). Thus, it has one of the most recent information about EV charging found online, frequently updated on the website. Boulder does not have the customer ID field. Instead, it features extra fields, like the length of charging time, GHG emission reductions, and gasoline savings from all city-owned EV charging stations.

**Palo Alto** dataset [73] is one of the largest datasets online. It covers six years and features 259416 EV charging sessions, from Jul 2011 until Dec 2020 in the city of Palo Alto, California. It has the same fields as the Boulder dataset plus the customer ID, like the ACN-Data.

**ElaadNL** dataset [74] is one of the best-known European datasets. An overview of 10000 random charging events, including 15-minute meter values per transaction for 2019, is currently available on the official website. It is also possible to access private data by requesting a code. In the literature, it can be seen that previous studies referencing this dataset indicate different fields and time periods when compared to the dataset currently made available to the general public. At the time of writing, the ElaadNL dataset features a Max Power entry, an absent field from the previously mentioned open datasets, which allows the investigation of EV charging flexibility.

It is necessary to choose only a subset of the total available fields from the datasets to perform clustering and produce meaningful results. According to the selected entries, the final outputs may be either EVSE accessibility, EV charging profiles, or EV user behavior. A complete list of the fields found in each of the aforementioned datasets is presented in **Table 2**.

In addition to these datasets, known as *Charging Event Datasets* or *Transaction Datasets*, it is also possible to obtain data in time series format, called *Meter Values Datasets*. However, this differs from the literature's norm for studying EV charging profiles or user behavior. For instance, the file obtained through the ElaadNL dataset website [74] also provides a meter values version for the same period (2019). Another dataset found online [75] corresponds to meter values of household consumption of more than 700 EV owners who joined an 18-month smart charging trial, known as the Electric Nation project [76], in the UK.

## 5. Applications of Clustering in EV Data

The literature often considers EV charging profiles and EV user behavior synonyms. Authors name their work using one term or another depending on the dataset and the chosen fields. The same does not happen for EVSE accessibility, whose studies utilize EVSE location data and not EV charging data. For example, Shen et al. [77] grouped the charging sessions from the ACN-Data dataset by each user and then performed clustering, naming their work EV user charging behavior identification. Shahriar and Al-Ali [78] also utilized the ACN-Data but chose the features of each session without clustering the data by user, naming it charging behavior clusters. Ultimately, the two studies found groups with similar charging behavior characteristics. Thus, this section presents the work done in each of these areas, divided into the subsections EV User Behavior & Charging Profiles, EVSE Accessibility and Location, and Other Applications.

Table 2: Summary of the available fields in the open datasets reviewed.

| Datasets | ACN-Data | Boulder | Palo Alto | ElaadNL |
|---|---|---|---|---|
| Format | JSON file | CSV file | CSV file | CSV file |
| Time Interval | 25 Apr 2018 - 14 Sep 2021 | 20 Jan 2018 - 09 Sep 2023 | 29 Jul 2011 - 31 Dec 2020 | 01 Jan 2019 - 31 Dec 2019 |
| Total Sessions | 31424 | 148136 | 259416 | 10000 |
| EVSE Identification and Location | Only Identification | Both | Both | Only Identification |
| Start and End Times (plug-in and plug-out) | Yes | Yes | Yes | Yes |
| Charging Duration (effective time of charging) | No | Yes | Yes | Yes |
| Energy Consumed [kWh] | Yes | Yes | Yes | Yes |
| Port Type | No | Yes | Yes | No |
| Customer Identifier | Yes | No | Yes | Yes |
| GHG and Fuel Savings | No | Yes | Yes | No |
| Customer Postal Code | No | No | Yes | No |
| Max Power (per charging session) | No | No | No | Yes |

### 5.1. *EV User Behavior & Charging Profiles*

EV charging data has been submitted to clustering methods to identify the most common and recurrent charging profiles. As previously mentioned, these studies aim to identify valuable information to assist ESMs, DSOs, utilities, and urban planners in correctly implementing EVs in the power system. EV batteries also represent a flexibility potential that may become increasingly valuable to the energy system as RES increases in prevalence [79]. Previous studies have proposed various methods to identify clusters of similar charging patterns.

For example, Shen et al. [77] employed the K-means algorithm to cluster EV charging behavior, aiming to improve the efficiency of scheduling EVs within a 5G-enabled smart-grid environment. To enhance the accuracy of the clustering process, especially when dealing with sparse or irregular data, the authors introduced a hybrid intelligence concept known as Human-in-the-Loop (HITL). This approach involves manual supervision and adjustment of the clustering results to correct potential errors that might arise from the algorithm. To obtain typical user behavior, the authors grouped data based on individual users, employing features such as the average charging time, the standard deviation of charging time, and the standard deviation of connection time. This method yielded three distinct clusters: two representing users with stable and predictable charging behaviors and a third representing EV drivers with unpredictable behaviors. Subsequently, the K-Nearest Neighbors (K-NN) algorithm was applied to classify new data, effectively identifying whether new users exhibit stable or unstable charging patterns based on the clusters found.

Similarly, Xiong et al. [80] investigated EV user behavior by organizing the data at the user level, representing each user with a tuple consisting of average arrival time, average departure time, standard deviation of arrival time, and standard deviation of departure time. Additionally, the authors included the Pearson correlation coefficient between stay duration and energy consumption to capture the relationship between these variables. Using this enriched data, K-means clustering was employed to identify four distinct user profiles. One cluster exhibited highly predictable behavior, with arrival and departure times falling within a well-defined range and minimal variance, while another represented users with random travel schedules and energy consumption patterns. After identifying these clusters, the authors employed a multilayer perceptron (deep learning approach) to study EV user charging records and generate classifications based on the clustering labels, improving the ability to predict and manage future charging demands.

Van Kriekinge et al. [81] 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 performed clustering with data from a private dataset, replacing all charging sessions with one specific theoretical charging session per EV driver (average value of the plug-in times, parking times, and charged energy) with the goal of obtaining profiles of EV user behavior. The result is a mean behavior for each EV user. The clustering proposed in this study works in two stages: the first step consists of clustering the mean features per EV user to find users with typical behavior, and the second step is conducted only on how frequently each EV driver charges their vehicle, always utilizing the K-means algorithm. The results indicated five clusters, with big differences in behavior between the EV drivers. In addition, the Kernel Density Estimation (KDE) process allows capturing the details of each cluster, and the specific charging behaviors, helping in the final simulation stage, which demonstrated a strong impact on power and energy demand when adding new EV users to the population.

Gerossier et al. [82] employed Hierarchical Clustering to classify EV charging behavior into four distinct groups. The study processed time series data to extract individual charging sessions, categorizing them based on the start-up time (initial plug-in time) and the duration of the charging process. This methodology is thoroughly documented and presented in their work. Most customers were found to belong to the first group, characterized by charging activities predominantly occurring during the evening and morning hours.

To find the typical behavior of users in fast charging EVSEs, Capeletti et al. [83] performed agglomerative clustering with kernel density estimation on private data from 10 charging stations on highways, using features such as stage of charge, energy, power, time, and location to understand user dynamics during charging events. Despite defining their work as user behavior, the characterizations of the 5 clusters found are much more in line with the notion of EV charging profiles found in the majority of studies in the literature, since they characterize the charging sessions (energy delivered, recharge time, average power, and start time) without having any connection to the users, contrary to that verified in the Van Kriekinge et al. [81] study. The results indicated that certain locations have more characteristic clusters than others, supporting the development of load demand forecasting models. However, the study utilized a limited dataset of 5918 charging events collected over only 10 months, making the results unreliable in an ever-changing world such as electric mobility.

Working with a large dataset from metropolitan areas of the Netherlands, Helmus et al. [84] carried out a two-step, bottom-up data clustering approach that first employs GMM to cluster charging sessions and then portfolios of charging sessions per user using K-Medoids (comparable to K-means clustering). The study considers starting time, connection duration, the distance between two sessions, and hours between sessions as features. From the first step, thirteen clusters were found: 7 types of daytime charging sessions (4 short, 3 of medium duration) and 6 types of nighttime charging sessions. The second step resulted in nine distinct clusters: 3 clusters contained daytime chargers, 3 nighttime chargers, and the remaining 3 featured unusual users. The study is detailed, yet perhaps too complex. It requires careful reading and prior knowledge of some of the methods used.

On the other hand, Märtz et al. [85] claim they utilized the most comprehensive (private) dataset on charging patterns from an EV perspective known in the literature, containing approximately 21000 BMW i3 BEVs and about 2.6 million charging processes during one year (2019). The authors conducted GMM clustering on the EV charging behavior, using plug-in time and duration as features, and identified seven distinct clusters: 3 overnight and 4 daytime. Furthermore, the authors performed a supplementary analysis using K-means clustering to detect EV users transitioning between clusters. The methodology and decisions made throughout the manuscript are justified, enhancing the reader's comprehension of all the steps taken and providing an excellent visualization of the clusters identified, making this analysis one of the most comprehensive in the literature. The authors also highlighted the potential flexibility of the EV charging processes. Similarly, Singh et al. [86] utilized GMM clustering to classify EV charging behavior and forecasted energy demand using regression models such as Random Forest and LSBoost. The authors reasoned that GMM is advantageous over K-means because it accounts for data point variance. They also utilized plug-in time and duration fields, determining the optimal number of clusters using Akaike's Information Criteria (AIC) [87] and Bayesian Information Criterion (BIC) [88], a comparable approach to Märtz et al. [85].

One of the most interesting EV charging analyses was conducted by Shahriar and Al-Ali [78]. This study on real public EV charging activity during the COVID-19 pandemic performed cluster analysis with K-means, Hierarchical Clustering, and GMM to identify similar groups of charging behavior based on vehicle arrival and departure times. K-means produced the best results, followed by Hierarchical Clustering, according to the metrics Silhouette [65], Davies-Bouldin [66], and Calinski-Harabasz [67]. The authors only discovered three clusters corresponding to the knee of the elbow method curve. One of the study's drawbacks is that it only employed a single method for establishing how many clusters were appropriate for the data. Another drawback is that the authors only employed two features to group the data into clusters: arrival and departure times. They mentioned that adding more features may cause the results to vary, resulting in fewer generic clusters.

Bayram et al. [89] conducted detailed research of the first publicly available AC charging sessions in the UK over four months. This study focused on key features such as utilization rates, arrival and departure times, energy delivered, and overstay duration. To analyze the data, the authors applied the DBSCAN algorithm, which enabled the identification of patterns in charging behavior based on arrival and departure times. This approach highlighted distinct clusters, providing valuable insights into charging patterns and station utilization. Similarly, Sadeghianpourhamami et al. [90] employed DBSCAN to study a large dataset of EV charging sessions in the Netherlands. This analysis, which included over 390000 sessions, aimed to quantify the flexibility of EV charging behavior by clustering the sessions into three groups. These clusters were characterized by different behaviors, such as charging near home, near work, or parking to charge, and were further analyzed for their impact on the grid, particularly in terms of load flattening and renewable energy integration.

To address the existing gaps in the literature on charging profiles, Forte et al. [91] applied three clustering methods from different fields, namely K-means, Gaussian Mixture Model, and Hierarchical Clustering, conducting a benchmark analysis with the silhouette, Davies-Bouldin, and Calinski-Harabasz scores. Using open (Caltech dataset) and private (Greek dataset), focusing on the features energy delivered, connection time (plug-in time), and parking time (sojourn time), the authors concluded that the best method for either dataset was K-means. The study applied charging profiles to characterize the flexibility of charging processes, finding that, in Caltech, most chargers can benefit from around a 40% reduction in charging power. In contrast, most users in Greece can achieve around an 80% reduction, due to the characteristics of fast public EVSEs. These insights provide valuable empirical data for power system operators and charging infrastructure managers, aiding better planning and improving grid integration of EVs.

Kim et al. [92] aimed to understand the operational characteristics of EVs to enhance the electrical grid's stability and reduce battery degradation. Different from the previously mentioned studies, the authors characterized EV charging profiles as a joint probability distribution of the start and end states of charge. Using a Hierarchical Clustering algorithm based on the Jensen-Shannon distance, they validated their findings with Silhouette, Calinski-Harabasz, and Davies-Bouldin indexes. The results indicated that most users analyzed in South Korea do not recharge their batteries to 100%, but 80%. This outcome presents an opportunity to optimize charging structures and develop personalized services for battery management.

Another interesting investigation is presented by Ahmed et al. [93], who introduced a study into the impact of EV penetration in load profiling of domestic consumers by performing clustering. Consumers classified into different classes before adopting an EV tend to fall into the same class after the incorporation of EV charging. The authors employed K-means with data from only 10 EVs, making the analysis unreliable, with further in-depth work required.

Unlike previous research that aimed to find user behavior to adapt the power grid system better, Hu et al. [94] conducted a study to identify and categorize typical EV users for marketing purposes. By extending the conventional RFM model to the RFMLT model, the authors were able to cluster EV users using a two-stage clustering technique that combines the DBSCAN algorithm and the K-means algorithm. The traditional RFM model is a well-known database marketing technique that can extract the most valuable information from users with fewer indicators. After analyzing several clustering techniques, the proposed strategy was more reliable than other methods. The findings showed that six groups could be formed from the 7426 EV users: "high-value users", "key users to maintain", "key users to develop", "potential users", "new users", and "lost users".

**Table 3** presents a summary of the most relevant information of the aforementioned studies.

Table 3: Summary of the EV user behavior & charging profiles papers reviewed.

| Study | Brief summary | Clustering method | Dataset | Conclusions |
|---|---|---|---|---|
| Shen et al. [77] USA & Canada 2020 | To manage (dis)charging behavior of EVs in the smart grid, proposes a communication network for analysis and prediction of user behavior | HITL-based K-means clustering and K-NN algorithm for prediction | ACN dataset, from Caltech University Campus | Identified 2 clusters of stable and predictable users, but the third cluster was found to be unexpected users |

**Table 3 cont.**

| Study | Brief summary | Clustering method | Dataset | Conclusions |
|---|---|---|---|---|
| Xiong et al. [80] Los Angeles, USA 2018 | Proposes an EV user behavior technique, using unsupervised and deep learning techniques, applied to historical EV data to make the day-ahead parking and charging prediction | K-means for clustering, multilayer perceptron for classification | More than 4 years data of the UCLA SMERC smart charging network infrastructure | Identified 4 clusters, with 3 relatively predictive behavior, but one cluster represented random traveling schedule and energy consumption |
| Kriekinge et al. [81] Brussels, Belgium 2023 | Proposes a methodology to simulate charging demand for different EV driver types. The identification of similar profiles is performed using clustering | K-means for clustering and KDE to better capture details for the simulation stage | 8 755 private EV charging sessions (Jul 2018 - Jan 2022) | Identified 5 clusters, with distinct and different characteristics, showing good clustering results |
| Gerossier et al. [82] Texas, USA 2019 | Models the consumption profile of EVs from raw power measurements. The charging habits model is then used for forecasting short-term, one day ahead, to long term (2030) | Hierarchical Clustering with Ward's method | 46 private EV charging data recorded every minute of the year 2015 in Texas | Identified 4 clusters. Simulating the projected demand in 2030, it appears that the growth in EVs will have little effect on the load curve's shape |
| Capeletti et al. [83] Brazil 2024 | Analyzes the recharging sessions of highway public fast chargers to investigate the behavior of users when requiring fast power | Hierarchical Clustering and KDE to examine the profiles | 5918 charging events from 10 public fast EVSEs in Brazil (Aug 2023 - Jun 2024) | 5 clusters distributed across the 10 locations. More data is required to validate the results |
| Helmus et al. [84] Amsterdam, Netherlands 2020 | Provides a realistic analysis of charging behavior and EV user types based on clustering, differing from the typical literature that seems over-simplified | Gaussian Mixture Models for clustering and Partition Around Medoids to find portfolios of charging sessions per user | 5.82 million charging transactions (January 2017- March 2019) from the Dutch metropolitan area | 13 clusters were found: 7 types of daytime charging sessions (4 short, 3 medium duration) and 6 types of overnight charging sessions |
| Märtz et al. [85] Germany 2022 | Investigates the possibility of identifying different clusters of EV charging processes, validating the results against synthetic load profiles and the original data | Gaussian Mixture Models and K-means | 2.6 million private charging processes of 21000 BMW's i3 model from 2019 in Germany | High number of charging opportunities during day, as well as user exchange between charging clusters, to reduce localized energy demand. Found 7 clusters |
| Singh et al. [86] Ottawa, Canada 2022 | Proposes a smart EV charging strategy that incorporates user charging behaviors to optimize charging schedules and reduce costs | Gaussian Mixture Models, Random Forest, LSBoost | ElaadNL, Dutch smart charging dataset, 10000 sessions | Flexible smart charging outperforms baseline scheduling in reducing charging costs and managing load shifting |
| Shahriar and Al-Ali [78] UAE 2022 | Investigates the impacts of COVID-19 on EV charging behavior by analyzing the charging activity during the pandemic | K-means, Hierarchical Clustering, and Gaussian Mixture Models | ACN dataset, from Caltech University Campus | Identified 3 groups of charging behavior. The best clustering was obtained using K-means followed by Hierarchical Clustering |
| Bayram et al. [89] UK 2023 | Conducts a statistical analysis of public AC EV charging sessions to characterize usage patterns and identify distinct behavioral clusters using unsupervised learning | DBSCAN | 12000 EV charging sessions from 595 public AC chargers (7 kW and 22 kW) across the UK | Identifies three clusters (short overnight, daytime, long overnight charging) and shows significant opportunities for smart charging to reduce peak loads by 30% |

**Table 3 cont.**

| Study | Brief summary | Clustering method | Dataset | Conclusions |
|---|---|---|---|---|
| Sadeghia. et al. [90] Netherlands 2018 | Analyzes the flexibility in EV charging behavior to optimize load flattening and balancing, using real-world charging data from the Netherlands | DBSCAN | Private ElaadNL, with 387524 EV charging sessions from public EVSEs (2014-2015) | Identifies 3 behavioral clusters (home, work, park), highlighting significant flexibility in charging times that can be leveraged for demand response strategies |
| Forte et al. [91] Portugal & Greece 2025 | Performs a benchmark analysis of different clustering methods to find the best charging profiles to characterize the flexibility of EVs and help integrate them into the grid | K-means, Hierarchical Clustering, and Gaussian Mixture Models | ACN dataset, from Caltech University Campus, and private data from public EVSEs in Greece (2021-2023) | K-means performed better, with the 8 ACN cluster sessions mostly longer and more adaptable, while the 10 Greek cluster sessions are shorter and less flexible |
| Kim et al. [92] S. Korea 2024 | Aims to understand the operational characteristics of EVs to improve grid stability and mitigate battery degradation | Hierarchical Clustering | Private charging records from 499 EVs, during 1 year in South Korea | Identified 7 distinct profiles, with patterns ranging from frequent full charges to strategic partial charges to preserve battery health |
| Ahmed et al. [93] Pakistan & S. Korea 2021 | Investigates the impacts of inserting EVs into smart meter data at the household level by performing clustering | K-means | With 30 minutes resolution, 30 EV charging profiles from the NREL (USA) dataset | Consumers who were classified into different classes before introducing EVs tend to fall into the same classes after adopting EVs |
| Hu et al. [94] China 2022 | The goal was to identify and categorize typical EV users for marketing purposes, by extending the conventional RFM model to the RFMLT | DBSCAN and K-means | 228440 charging records of 7426 EV users, from Internet of Vehicles Platform in 2019 | Identified 6 clusters: high value users, key users to maintain, key users to develop, potential users, new users and lost users |

### *5.2. EVSE Accessibility and Location*

Given the previously mentioned studies, it may seem that clustering is only applied to EV charging data. However, the focus of the literature goes beyond EV charging patterns [95]. As EV sales increase, the location and positioning of EVSEs may become an issue [96], which clustering can help to address.

Carlton and Sultana [97] perform spatial clustering of public EVSEs to analyze the characteristics of their land use and how these characteristics impact EVSE accessibility. The authors applied DBSCAN to identify spatially clustered Level-1, Level-2, and DC fast charging infrastructures in the Chicago Metropolitan Area. The results indicate that access to EVSEs is unequal between suburban and urban neighborhoods, bringing social inequalities into view and preventing the widespread adoption of EVs.

Finding the most appropriate location to place EVSEs is a big problem. Few studies have used clustering methods to locate EVSEs; most research papers concentrate on building charging station placement models and positioning charging stations based on various objective functions and restrictions [98]. To demonstrate that clustering can produce more accurate and understandable results, Li et al. [99] proposed a broadly applicable technique for finding EVSE locations in Qingdao, China, based on multi-type clustering in response to the expanding demand for electric taxis. Electric taxi parking information and charging needs are derived from the extensive GPS trajectories of gasoline taxis. To find the ideal site for the charging station, the study area of Qingdao is subjected to multiple same-type clustering and multi-type clustering methods. K-means and Agglomerative Hierarchical Clustering were methods utilized for this comparative analysis, demonstrating that the positioning results of the multi-type clustering are more credible. Similarly, Shukla et al. [100] aimed to reduce social costs by minimizing the distance traveled for charging EVs. The number of EVSEs to be located was determined by calculating EVs' energy demand and the EVSEs' service radius. K-means and fuzzy C-means [101] are used to find the location of EVSEs, with the location of stations given by the centroid of the clusters and the number of clusters corresponding to the number of stations needed in the area under study.

Sánchez et al. [102] proposed a mixed-integer linear programming (MILP) model to solve the electric location routing problem with time windows (E-LRPTW), integrating the optimization of EV routes with the best placement of EVSEs. The model aimed to minimize the distance traveled, the number of EVs used, and the number of EVSE locations by incorporating constraints such as state of charge, freight, battery capacities, and customer time windows. To enhance computational efficiency, the authors employed K-means to group drivers and determine potential sites for placing the EVSEs. Depending on the instance, the clustering algorithm identified two to five clusters, each representing a specific geographical area with a centroid as a candidate location for a recharging station. This approach significantly reduced the number of binary variables, allowing the model to optimize EV routing scenarios effectively and solve larger instances in a reasonable time.

Sun [103] introduces a novel methodology for optimizing public EVSE deployment in urban neighborhoods, focusing on high-traffic areas. The approach integrates several phases, including data preprocessing, DBSCAN clustering, trajectory-road map matching, dynamic road segmentation, and optimal station placement using a 0/1 knapsack problem. By analyzing EV taxi GPS data from one day within a 4.5 square kilometer area, the study identifies optimal locations for charging stations. The DBSCAN clustering resulted in 256 distinct clusters, each representing areas with dense EV traffic. These clusters were further refined using trajectory-road map matching to ensure alignment with the road network, thereby avoiding impractical locations. The methodology proved more efficient than existing infrastructure, achieving higher coverage rates with fewer charging stations, thus providing a scalable solution for enhancing urban EV charging networks.

Furthermore, Kalakanti and Rao [104] addressed two main problems related to EVSE: the EVSE location problem and the EVSE need estimation problem. This work investigated different explainable solutions based on machine learning and simulation. For the problem of EVSE location, the authors utilized a geolocation dataset of EV households (of Austin, USA, and a greenfield area of Bengaluru, India) to perform a comprehensive analysis with different classes of clustering methods, namely K-means, GMM, OPTICS [105] (similar to DBSCAN but more flexible and scalable), and Spectral clustering. The results were compared in two planning areas: the Austin area, with the existing EVSE location data (to show the improvement over the existing setup), and a greenfield area like Bengaluru, where synthetic EVSE data were used. Silhouette coefficient, Calinski-Harabasz index, and Davies-Bouldin index were the chosen metrics to evaluate the clustering results. The final results guide urban planners in making better EVSE placement.

**Table 4** presents a summary of the most relevant information of the aforementioned studies.

Table 4: Summary of the EVSE Accessibility and Location papers reviewed.

| Study | Brief summary | Clustering method | Dataset | Conclusions |
|---|---|---|---|---|
| Carlton et al. [97]<br>North Carolina, USA<br>2022 | Performs spatial clustering of public EVSE to analyses their associated land use tendency, and how these can impact EVSE accessibility | Hierarchical Clustering based on DBSCAN | Public EVSE location data from the Alternative Fuel Data Center (AFDC) | Majority of level 2 EVSE, only 26% of clusters with mixed land uses (residential, commercial and recreational) |
| Li et al. [99]<br>China<br>2022 | Proves that clustering can be used to find optimal EVSE locations, specially important for the growing market of the electric taxis in China | K-means and Hierarchical Clustering | Obtained from "Gao De Map" | The optimal location result of the multi-type clustering is more plausible than that of the same-type clustering |
| Shukla et al. [100]<br>India<br>2016 | Intends to solve the optimal location problem of EVSEs by minimizing the distance traveled for charging EVs | K-means and fuzzy C-means | Estimate the EV no. by evaluating the households, modifying the IEEE 123 bus distribution system | The clustering methods performed better when compared with random placement of EVSEs |

**Table 4 cont.**

| Study | Brief summary | Clustering method | Dataset | Conclusions |
|---|---|---|---|---|
| Sánchez et al. [102] São Paulo, Brazil 2022 | Proposes a model for optimizing EV routing and the placement of EVSEs, using clustering to reduce computational complexity | MILP with K-means | Simulated data based on previous studies | The model reduces computation time significantly while optimizing routes and EVSE locations |
| Sun [103] Undisclosed city, 2023 | Optimizes the placement of public EVSEs in an urban area using real EV taxi trajectory data. High-density traffic areas are obtained trough clustering | DBSCAN | 101600 EV taxi trajectories, GPS data from October 2016 | Methodology achieves higher coverage rates with fewer charging stations compared to existing infrastructure |
| Kalakanti and Rao [104] India 2022 | Aims to solve two problems: the EVSE placement and the EVSE need estimation, to guide the urban planners in making better EVSE placement | K-means, GMM, OPTICS, Spectral Clustering, and other ML methods | Austin Charging Station Network real geolocations | K-means and GMM consistently yielded the best results, with OPTICS and Spectral Clustering often wrong or nonsense |

### 5.3. *Overview of Analyzed Applications per Clustering Method*

To provide a more comprehensive overview of the clustering methods used in EV data analysis, Table 5 summarizes the reviewed studies according to the clustering technique adopted and the primary objective. All the studies listed identified EV user behavior/charging profiles or EVSE accessibility through clustering. However, while some articles focused exclusively on uncovering these patterns, others utilized the identified profiles to support additional objectives, such as scheduling, forecasting, grid impact assessment, flexibility quantification, or market segmentation.

Table 5: Summary of reviewed applications per clustering method.

| | K-means | Hierarchical Clustering | GMM | DBSCAN | Fuzzy C-means | Spectral Clustering | OPTICS |
|---|---|---|---|---|---|---|---|
| Scheduling | [77] | | [86] | | | | |
| Forecasting/Classification Models | [80], [85] | [82] | [85] | | | | |
| User Behavior Charging Analysis | [78] | [78], [83], [92] | [78], [84] | | | | |
| Grid Impact Assessment | [81], [93] | | | [89] | | | |
| Flexibility Quantification | [91] | [91] | [91] | [90] | | | |
| Market Segmentation | [94] | | | [94] | | | |
| EVSE Placement/Accessibility | [99], [100], [102], [104] | [97], [99] | [104] | [97], [103] | [100] | [104] | [104] |

As revealed in Table 5, K-means and Hierarchical Clustering are among the most commonly employed methods across various applications, particularly in EVSE planning and user behavior analysis. GMM and DBSCAN appear less frequently, revealing that density-based methods are underexploited in EV data. However, GMM is more competitive in studies focused on characterizing user behavior and charging profiles, while DBSCAN reveals its strengths in EVSE placement. More advanced methods, such as OPTICS, spectral clustering, and fuzzy C-means, are less commonly used, although they have emerged in studies targeting EVSE placement and accessibility.

This distribution highlights the popularity and limitations of traditional clustering techniques and the potential for the adoption of more refined algorithms. The diversity of methods also reflects the variation in data availability and research goals, reinforcing the need for methodological transparency and comparative studies.

### 5.4. *Other Applications*

In addition to EV user behavior, charging profiles, or EVSE accessibility, clustering can be an extra tool to achieve different goals. In fact, the idea of applying clustering to help machine learning methods forecast charging habits has received a lot of attention in the literature.

The study by Nespoli et al. [107] aimed to create a complete representation of the power required for the charging station through an EV charging session forecasting method that was both accurate and detailed. The main areas of study were the forecast and reconstruction of the aggregated power profile, but clustering represented a key step in obtaining the forecast results. The authors used OPTICS to obtain EV user behavior. To do that, they chose the start charging time and the energy used throughout the charging period as features. From this, each charging session was predicted with a trio of parameters: the "arrival time", the "charging duration", and the "average power" expected during the process. Since the paper's main goal was forecasting rather than clustering, the clustering step was not as thoroughly discussed as the rest of the work. Nonetheless, it is a great example of how clustering can be used in EV data analysis.

A more detailed and complex approach was performed by Shahriar et al. [108] and Crozier et al. [109]. Shahriar et al. [108] proposed the usage of historical charging data in conjunction with weather, traffic, and events data to predict EV session duration and energy consumption using popular machine learning algorithms, including Random Forest, Support Vector Machine, eXtreme Gradient Boosting, and deep Neural Networks. On the other hand, by combining information from two different data sources (travel survey data and vehicle usage data), Crozier et al. [109] employed a stochastic model to assess EV charging and predict its effects on the energy network in the UK. The model is based on conditional probability distributions and incorporates random variables for vehicle usage, stage of charge, time, and type of day. The survey dataset was clustered using K-means, and the discrete probability distributions were created using the EV experiment data. The study underlined the importance of effectively understanding the diversity of EV charging demands among consumers.

To better comprehend the process of forecasting EV charging, Zhu et al. [110] presented a comparative study of deep-learning methodologies to forecast EVs' short-term charging (for monitoring and controlling applications). Several deep-learning-based methods were analyzed, using real-world data from EVSEs to compare their performance.

Based on actual fine-grained data, Sun et al. [111] completed a study to comprehend the behavior of an EV battery. The objective was to identify profiles in time series data of battery charging rates that might be utilized to forecast charging behavior. The authors performed alternating minimization [112] on an optimization function, comparing the results with Euclidean distance, DTW, and a modified version of Euclidean distance, which yielded the best results. This study demonstrates applications of EV time series clustering.

An alternative method to machine learning algorithms is simulation, a technique widely used by the great majority of researchers due to the lack of EV real-world open data [68]. Several simulation models have been put out to address this issue. Zhang et al. [113] investigated charging profiles of EVs, presenting a sophisticated simulation method that takes people's demographic and social characteristics into account. Numerous articles were written in the field of EV charging analysis based on simulated data or information owned by private companies, which is rarely made available. In the absence of real data, one approach is to use synthetic charging profiles derived from the driving behavior of conventional vehicle users, like Schäuble et al. accomplished in [114].

## 6. Recommendations and Future Work

As EVs become increasingly integrated into power systems and urban environments, a deeper understanding and evolution of clustering methods is essential to ensure their practical utility, robustness, and scalability [25]. This section reflects on the current limitations observed in the literature, explores how clustering techniques are already being applied in electric mobility, and outlines possible improvements and future work directions.

### *6.1. Limitations of Current Clustering Studies*

Clustering has emerged as a valuable tool for EV user and charging behavior, yet several limitations persist in existing studies that impede the generalization and scalability of their findings

Firstly, one of the most pressing issues is the limited availability of high-quality, publicly accessible datasets. As previously mentioned in the review, many analyses rely on data from private companies, rarely made available to the research community. Additionally, the currently available datasets typically focus on limited geographic regions (primarily in the United States), which challenges the extrapolation of results to other contexts, particularly in Europe and developing countries.

Another critical limitation is the restricted dimensionality of the data used. Most studies apply clustering algorithms to datasets with only a few features, often limited to charging session start times, durations, and energy consumption. This narrow scope prevents the discovery of more nuanced behavioral patterns. In particular, contextual variables such as weather conditions, traffic congestion, or local area events (often used by forecasting studies) are rarely considered. Integrating these variables could significantly improve cluster interpretability and model performance.

Finally, the diversity of clustering techniques remains limited. Classical approaches like K-means or Hierarchical Clustering dominate the literature, but these are not always suitable for capturing dynamic and heterogeneous EV charging behavior. The limited number of comparative studies and benchmarks further complicates the clustering performance evaluation across research works.

### *6.2. Practical Applications of EV Clustering*

Despite the limitations outlined earlier, clustering techniques have revealed strong potential for informing practical applications within the EV ecosystem by characterizing user/charging behavior and supporting operational and strategic decisions:

- **Charging Scheduling:** Clustering can group users with similar charging patterns and preferences, enabling the design of intelligent scheduling algorithms. These can optimize charging sessions to avoid grid congestion or to align charging demand with renewable generation (Shen et al. [77], Singh et al. [86]).
- **Forecasting and Classification Models:** Clustering provides a valuable preprocessing step for machine learning models, improving the accuracy of forecasting EV demand or classifying users by their charging behavior or energy flexibility (Xiong et al. [80], Gerossier et al. [82], Märtz et al. [85]).
- **User Behavior and Charging Process Analysis:** By analyzing clustered behavior, researchers can gain insights into the diversity of user needs and charging habits. This enables more accurate user segmentation and facilitates the personalization of EV services (Capeletti et al. [83], Helmus et al. [84], Shahriar and Al-Ali [78], Kim et al. [92]).
- **Grid Impact Assessment:** Clustering also plays a key role in evaluating the impact of EV charging on the electrical grid. By distinguishing high-demand clusters and their temporal patterns, system operators can anticipate stress points and plan reinforcement actions accordingly (Kriekinge et al. [81], Bayram et al. [89], Ahmed et al. [93]).
- **Flexibility Quantification:** Clustered profiles enable a more accurate estimation of the flexibility potential from EV users, which is particularly important for demand response programs, where it is crucial to identify which users or usage patterns are best suited for temporal load shifting (Sadeghianpourhamami et al. [90], Forte et al. [91]).
- **Dynamic Pricing and Market Segmentation:** Grouping users into behavior-based clusters supports the development of differentiated tariff structures and allows market actors to target specific segments with incentives for off-peak charging, supporting more efficient grid operation (von Bonin et al. [39], Hu et al. [94]).
- **EVSE Infrastructure Planning and Accessibility:** Clustering helps uncover spatial and temporal charging demand patterns, informing the placement and sizing of new EVSEs. This contributes to more efficient infrastructure deployment and improves accessibility (Carlton and Sultana [20], Li et al. [99], Shukla et al. [100], Sanchez et al. [102], Sun [103], Kalakanti and Rao [104]).

### *6.3. Future Research Directions*

To fully leverage the potential of clustering in electric mobility, future research must embrace more integrated, context-aware methodological approaches in algorithm design and data processing. Addressing these challenges facilitates a more effective application of clustering techniques and reveals several promising and underexplored areas for future academic and practical research.

#### 6.3.1. Integrating Contextual and Spatial Dimensions

Clustering methods should increasingly consider contextual and spatial variables to improve the interpretability and relevance of identified patterns. Context-aware clustering can help explain temporal fluctuations in charging behaviors by considering external factors such as weather, public holidays, or traffic [108]. Furthermore, future studies could combine quantitative clustering analysis with qualitative methods, such as user surveys, to validate whether the identified charging and behavior clusters truly reflect real-world usage patterns and needs.

However, an underexplored yet impactful direction lies in incorporating spatial context, which is essential for supporting decision-making in EVSE accessibility and location planning. Factors such as urban density, proximity to transport corridors, availability of EVSEs, and socio-economic characteristics significantly impact how, when, and where vehicles are recharged. By embedding these geographical attributes into clustering models, it is possible to generate more accurate and actionable profiles for ESMs, urban planners, utilities, and policymakers, enabling comparisons across areas and supporting the identification of region-specific needs to maximize effectiveness and user satisfaction.

#### 6.3.2. Methodological Innovations in Unsupervised Learning

Traditional clustering algorithms have provided valuable insights in previous studies (recall Section 5). However, they may fall short in capturing the complexity of high-dimensional, non-linear, or overlapping usage patterns, especially when combining various features. It is expected that access to big data by smart meters will imply a need for future research to explore alternative and more advanced unsupervised learning techniques, such as spectral clustering, fuzzy C-means, or OPTICS, which can better manage non-convex cluster shapes and uncertainty in behavior. Additionally, hybrid approaches that combine classical clustering with deep learning models (e.g., autoencoders [115] or self-organizing maps [106]) have shown promising results in several scientific domains and could improve the detection of complex patterns in large-scale EV data.

#### 6.3.3. Temporal Dynamics and Time Series Clustering

Despite its relevance, as mentioned in Section 3.5, the application of time series clustering remains limited in the current electric mobility literature. Time series profiles are becoming indispensable for demand response programs, time-of-use pricing, and forecasting algorithms, which require time-stamped data. Integrating this technique can uncover hidden structures in complex temporal datasets, offering insights that static clustering methods may overlook [61]. Nevertheless, the adoption of these techniques also raises methodological challenges, such as selecting appropriate distance measures, handling irregular sampling, and ensuring scalability with large datasets [116]. Future research should, therefore, prioritize the development and validation of time series clustering methods that capture the complexity and variability of user behavior. As the availability of fine-grained EV data increases, time series clustering is expected to play an increasingly important role in optimizing electric mobility systems and supporting data-driven policy and operational strategies.

#### 6.3.4. Benchmarking, Reproducibility, and Open Data

To ensure progress in this field is cumulative, future studies must prioritize methodological transparency and comparability. A significant challenge in the current clustering literature on EV data is the lack of standardized benchmarking that allows for the comparison and replication of results. Researchers should consistently report validation metrics (e.g., silhouette score [65], Davies-Bouldin index [66]) and conduct sensitivity analyses to assess the robustness of their outcomes. In addition, as new studies are published, sharing datasets in open repositories would significantly accelerate innovation by allowing other researchers to validate, replicate, and build upon existing findings. Fostering a culture of openness and reproducibility is essential for creating reliable and scalable clustering frameworks that support data-driven decision-making in electric mobility.

## 7. Conclusions

A comprehensive review of clustering applications for electric vehicle (EV) user behavior, EV charging behavior, electric vehicle supply equipment (EVSE) accessibility, and its optimal location has been presented in this paper. It started by providing the background and current state of EVs, an exposition of the main negative impacts of uncoordinated EV charging, followed by a thorough description of the most commonly employed clustering algorithms in the context of EV data analysis, along with various metrics that can be applied to validate the results. A comparative analysis of various studies using different methods, objectives, and datasets was presented, structured by key topics such as EV user behavior and charging profiles, EVSE accessibility, and location strategies. In addition, studies applying clustering for broader objectives, such as forecasting, scheduling, and quantifying flexibility, were also reviewed. Finally, recommendations and future challenges were also discussed. We identified a scarcity of studies, regional bias of open EV datasets, limited implementation of clustering techniques, and the absence of metrics to compare results between articles. Practical use cases for clustering were explored, including its role in assisting charging scheduling algorithms, forecasting and classification models, assessment of grid impact, flexibility quantification, and EVSE planning. Furthermore, future research directions were discussed, from hybrid deep learning and adaptive clustering models to geographical expansion, data availability, and advanced unsupervised learning techniques. This study highlights that clustering remains a powerful but underutilized tool for understanding EV charging behavior and supporting the design of more efficient and user-centered mobility and energy systems. Continued research is needed to improve model adaptability, integrate contextual data, and validate results across diverse geographic regions. These efforts aim to identify valuable information to support energy system modelers, distribution system operators, and urban planners in navigating the transition to a more electrified and sustainable future.

### CRediT authorship contribution statement

**Marcelo Forte:** Writing – original draft, Writing – review & editing, Methodology, Conceptualization. **Cindy P. Guzman:** Writing - review & editing, Supervision. **Lucas Pereira:** Writing – review & editing. **Hugo Morais:** Conceptualization, Writing - review & editing, Supervision.


### Acknowledgements

This document is the results of the research project funded by European Union´s Horizon Europe R&I programme under grant agreement no. 101056765. Views and opinions expressed in this document are however those of the authors only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the grating authority can be held responsible for them. This work was also supported by national funds through FCT, Fundação para a Ciência e a Tecnologia, under project UIDB/50021/2020 (DOI:10.54499/UIDB/50021/2020) and within the scope of the project nº 56 - "ATE", financed by European Funds, namely "Recovery and Resilience Plan - Component 5: Agendas Mobilizadoras para a Inovação Empresarial", included in the NextGenerationEU funding program."

## References


[1] United Nations. (n.d.). SDG Indicators — Sustainable Development Goal Indicators. https://unstats.un.org/sdgs/indicators/indicators-list/

[2] United Nations Climate Change. (n.d.). The Paris Agreement. https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement

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

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

[5] European Environment Agency (2023). Transport and environment report 2022 — Decarbonising road transport: the role of vehicles, fuels and transport demand. https://www.eea.europa.eu/publications/transport-and-environment-report-2022

[6] European Council (2023). Fit for 55: towards more sustainable transport. https://europa.eu/!yfBkpH

[7] Liu, Z., Deng, Z., Davis, S.J., Giron, C., and Ciais, P. (2022). Monitoring global carbon emissions in 2021. *Nature Reviews Earth & Environment*, **3**(4), 217–219. doi:10.1038/s43017-022-00285-w

[8] Pi, Z., Wang, K., Wei, Y.-M., and Huang, Z. (2024). Transitioning from gasoline to electric vehicles: Electrification decision of automakers under purchase and station subsidies. *Transportation Research Part E: Logistics and Transportation Review*, **188**, 103640. doi:10.1016/j.tre.2024.103640

[9] Kane, M. (2022). Global Plug-In Electric Car Sales Increased 61Access here

[10] Lakshmi, R.B. (2023). The Environmental Impact of Battery Production for Electric Vehicles. *Earth.Org*. Access here

[11] Zhang, X., Gao, F., Gong, X., Wang, Z., and Liu, Y. (2018). Comparison of Climate Change Impact Between Power System of Electric Vehicles and Internal Combustion Engine Vehicles. *Advances in Energy and Environmental Materials*, Springer Singapore, 739–747. doi:10.1007/978-981-13-0158-2_75

[12] Yap, K.Y., Chin, H.H., and Klemeš, J.J. (2022). Solar Energy-Powered Battery Electric Vehicle charging stations: Current development and future prospect review. *Renewable and Sustainable Energy Reviews*, **169**, 112862. doi:10.1016/j.rser.2022.112862

[13] IRENA (2023). World Energy Transitions Outlook 2023: 1.5°C Pathway; Preview. https://www.irena.org/Publications/2023/Mar/World-Energy-Transitions-Outlook-2023

[14] Richardson, D.B. (2013). Electric vehicles and the electric grid: A review of modeling approaches, Impacts, and renewable energy integration. *Renewable and Sustainable Energy Reviews*, **19**, 247–254. doi:10.1016/j.rser.2012.11.042

[15] Mureddu, M., Facchini, A., Scala, A., Caldarelli, G., and Damiano, A. (2018). A Complex Network Approach for the Estimation of the Energy Demand of Electric Mobility. *Scientific Reports*, **8**(1), 268. doi:10.1038/s41598-017-17838-5

[16] Jones, C.B., Lave, M., Vining, W., and Garcia, B.M. (2021). Uncontrolled Electric Vehicle Charging Impacts on Distribution Electric Power Systems with Primarily Residential, Commercial or Industrial Loads. *Energies*, **14**(6), 1688. doi:10.3390/en14061688

[17] Pamidimukkala, A., Kermanshachi, S., Rosenberger, J.M., and Hladik, G. (2024). Barriers and motivators to the adoption of electric vehicles: A global review. *Green Energy and Intelligent Transportation*, **3**(2), 100153. doi:10.1016/j.geits.2024.100153

[18] Gutiérrez-Lopez, J.B., and Möst, D. (2023). Characterising the flexibility of electric vehicle charging strategies: a systematic review and assessment. *Transport Reviews*, **43**(6), 1237–1262. doi:10.1080/01441647.2023.2217519

[19] Hopkins, E., Potoglou, D., Orford, S., and Cipcigan, L. (2023). Can the equitable roll out of electric vehicle charging infrastructure be achieved? *Renewable and Sustainable Energy Reviews*, **182**, 113398. doi:10.1016/j.rser.2023.113398

[20] Carlton, G., and Sultana, S. (2023). Transport equity considerations in electric vehicle charging research: a scoping review. *Transport Reviews*, **43**(3), 330–355. doi:10.1080/01441647.2022.2109775

[21] Shafiei, M., and Ghasemi-Marzbali, A. (2022). Fast-charging station for electric vehicles, challenges and issues: A comprehensive review. *Journal of Energy Storage*, **49**, 104136. doi:10.1016/j.est.2022.104136

[22] Al-Ogaili, T.H., Rahmat, N.A., Ramasamy, A.K., Marsadek, M.B., Faisal, M., and Hannan, M.A. (2019). Review on Scheduling, Clustering, and Forecasting Strategies for Controlling Electric Vehicle Charging: Challenges and Recommendations. *IEEE Access*, **7**, 128353–128371. doi:10.1109/ACCESS.2019.2939595

[23] Shahriar, S., Al-Ali, A.R., Osman, A.H., Dhou, S., and Nijim, M. (2020). Machine Learning Approaches for EV Charging Behavior: A Review. *IEEE Access*, **8**, 168980–168993. doi:10.1109/ACCESS.2020.3023388

[24] Andrenacci, N., and Valentini, M.P. (2023). A Literature Review on the Charging Behaviour of Private Electric Vehicles. *Applied Sciences*, **13**(23), 12877. doi:10.3390/app132312877

[25] Nazari, M., Hussain, A., and Musilek, P. (2023). Applications of Clustering Methods for Different Aspects of Electric Vehicles. *Electronics*, **12**(4), 790. doi:10.3390/electronics12040790

[26] Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery*, **88**, 105906. doi:10.1016/j.ijsu.2021.105906

[27] Massey, R. (2021). We test a replica of Trouvé's 1881 rechargeable electric vehicle. *This is Money*. https://www.thisismoney.co.uk/money/cars/article-9512103/We-test-replica-Gustave-Trouves-1881-rechargeable-electric

[28] Chandran, M., Palanisamy, K., Benson, D., and Sundaram, S. (2022). A Review on Electric and Fuel Cell Vehicle Anatomy, Technology Evolution and Policy Drivers towards EVs and FCEVs Market Propagation. *The Chemical Record*, **22**(2). https://onlinelibrary.wiley.com/doi/10.1002/tcr.202100235

[29] Office of the Historian (n.d.). Milestones: 1969–1976. *History.state.gov*. https://history.state.gov/milestones/1969-1976/oil-embargo

[30] International Energy Agency (IEA) (2025). Global EV Outlook 2025. Access here

[31] International Energy Agency (IEA) (2024). Stated Policies Scenario (STEPS) – Global Energy and Climate Model. https://www.iea.org/reports/global-energy-and-climate-model/stated-policies-scenario-steps

[32] EV Database (2025). EV Database: Range of full electric vehicles cheatsheet. Access here

[33] European Parliament and Council (2024). Alternative Fuels Infrastructure Regulation. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=LEGISSUM:4702332

[34] European Commission (2023). European Alternative Fuels Observatory. https://alternative-fuels-observatory.ec.europa.eu/
[35] Manzolli, J.A., Trovão, J.P., and Antunes, C.H. (2022). A review of electric bus vehicles research topics – Methods and trends. *Renewable and Sustainable Energy Reviews*, **159**, 112211. doi:10.1016/j.rser.2022.112211
[36] Drive to Zero (2023). The Program. https://globaldrivetozero.org
[37] Kang, Z., Ye, Z., Lam, C.-M., and Hsu, S.-C. (2023). Sustainable electric vehicle charging coordination: Balancing CO2 emission reduction and peak power demand shaving. *Applied Energy*, **349**, 121637. doi:10.1016/j.apenergy.2023.121637
[38] Vimal, K.E.K., Goel, P., Sharma, N., Mathiyazhagan, K., and Luthra, S. (2024). Where there is a will there is a way: A strategy analysis for electric vehicles sales in India. *Transportation Research Part E: Logistics and Transportation Review*, **185**, 103506. doi:10.1016/j.tre.2024.103506
[39] Von Bonin, M., Dörre, E., Al-Khzouz, H., Braun, M., and Zhou, X. (2022). Impact of Dynamic Electricity Tariff and Home PV System Incentives on Electric Vehicle Charging Behavior. *Energies*, **15**(3), 1079. doi:10.3390/en15031079
[40] Guzman, C.P., Montefusco, L., Morais, H., Forte, M., Khajehmahmo, F., Smpoukis, K., Mendek, I., Marentič, T., Matias, S., Silva, T., and Pediaditis, P. (2024). D4.5 Demand Response Programs Design for EVs. Horizon Europe EV4EU project.
[41] Amiruddin, A., Dargaville, R., Liebman, A., and Gawler, R. (2024). Integration of Electric Vehicles and Renewable Energy in Indonesia's Electrical Grid. *Energies*, **17**(9), 2037. doi:10.3390/en17092037
[42] Ruspini, E.H. (1969). A new approach to clustering. *Information and Control*, **15**(1), 22–32. doi:10.1016/S0019-9958(69)90591-9
[43] Zaki, M.J., and Meira Jr, W. (2020). Data Mining and Machine Learning: Fundamental Concepts and Algorithms. *Cambridge University Press*. doi:10.1017/9781108564175
[44] Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O.P., Tiwari, A., Er, M.J., Ding, W., and Lin, C.-T. (2017). A review of clustering techniques and developments. *Neurocomputing*, **267**, 664–681. doi:10.1016/j.neucom.2017.06.053
[45] Amestoy, T. (2022). Clustering basics and a demonstration in clustering infrastructure pathways. https://waterprogramming.wordpress.com/2022/03/16/clustering-basics-and-a-demonstration-in-clustering
[46] Cam, L.M.L., and Neyman, J. (1967). Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability: Weather modification. Google-Books-ID: IC4Ku_7dBFUC, *University of California*.
[47] Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, **39**(1), 1–38. http://www.jstor.org/stable/2984875
[48] Kodinariya, T., and Makwana, P. (2013). Review on Determining of Cluster in K-means Clustering. *International Journal of Advance Research in Computer Science and Management Studies*, **1**, 90–95.
[49] Steinbach, M., Karypis, G., and Kumar, V. (2000). A Comparison of Document Clustering Techniques. *University of Minnesota Digital Conservancy*. http://conservancy.umn.edu/handle/11299/215421
[50] Xu, R., and WunschII, D. (2005). Survey of Clustering Algorithms. *IEEE Transactions on Neural Networks*, **16**(3), 645–678. doi:10.1109/TNN.2005.845141
[51] Sneath, P.H.A., and Sokal, R.R. (1973). Numerical Taxonomy. *W H Freeman & Company*.
[52] Florek, K., Łukaszewicz, J., Perkal, J., Steinhaus, H., and Zubrzycki, S. (1951). Sur la liaison et la division des points d'un ensemble fini. *Colloquium Mathematicum*, **2**(3-4), 282–285. http://eudml.org/doc/209969
[53] Rohlf, F.J. (1982). Single-link clustering algorithms. *Handbook of Statistics*. doi:10.1016/S0169-7161(82)02015-x
[54] Sørenson, T. (1948). A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content. *Biologiske skrifter*. https://www.scirp.org/reference/referencespapers?referenceid=2150293
[55] Sokal, R.R., and Michener, C.D. (1958). A Statistical Method for Evaluating Systematic Relationships. *University of Kansas Science Bulletin*. https://ia600703.us.archive.org/5/items/cbarchive_33927_astatisticalmethodforevaluatin1902/astatisticalmethodforevaluatin1902.pdf
[56] Ward, J.D. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, **58**(301), 236. doi:10.2307/2282967
[57] Alvez, P.B.G. (2011). Inference of a human brain fiber bundle atlas from high angular resolution diffusion imaging. PhD Thesis, *Université Paris Sud - Paris XI*. https://theses.hal.science/tel-00638766
[58] Lance, G.N., and Williams, W.T. (1967). A General Theory of Classificatory Sorting Strategies: 1. Hierarchical Systems. *The Computer Journal*, **9**(4), 373–380. doi:10.1093/comjnl/9.4.373
[59] Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial Databases with Noise. *Knowledge Discovery and Data Mining*, 226–231. https://www.aaai.org/Papers/KDD/1996/KDD96-037.pdf
[60] Jia, H., Ding, S., Xu, X., and Nie, R. (2014). The latest research progress on spectral clustering. *Neural Computing and Applications*, **24**(7-8), 1477–1486. doi:10.1007/s00521-013-1439-2
[61] Alqahtani, A., Ali, M., Xie, X., and Jones, M.W. (2021). Deep Time-Series Clustering: A Review. *Electronics*, **10**(23), 3001. doi:10.3390/electronics10233001
[62] Berndt, D.J., and Clifford, J. (1994). Using Dynamic Time Warping to Find Patterns in Time Series. *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*, 359–370. https://dl.acm.org/doi/10.5555/3000850.3000887
[63] Wen, L., Zhou, K., Yang, S., and Li, L. (2018). Compression of smart meter big data: A survey. *Renewable and Sustainable Energy Reviews*, **91**, 59–69. doi:10.1016/j.rser.2018.03.088
[64] Warren Liao, T. (2005). Clustering of time series data—a survey. *Pattern Recognition*, **38**(11), 1857–1874. doi:10.1016/j.patcog.2005.01.025
[65] Rousseeuw, P.J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, **20**, 53–65. doi:10.1016/0377-0427(87)90125-7
[66] Davies, D.L., and Bouldin, D.W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-1**(2), 224–227. doi:10.1109/TPAMI.1979.4766909
[67] Calinski, T., and Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods*, **3**(1), 1–27. doi:10.1080/03610927408827101
[68] Amara-Ouali, Y., Goude, Y., Massart, P., Poggi, J.-M., and Yan, H. (2021). A Review of Electric Vehicle Load Open Data and Models. *Energies*, **14**(8), 2233. doi:10.3390/en14082233

[69] Calearo, L., Marinelli, M., and Ziras, C. (2021). A review of data sources for electric vehicle integration studies. *Renewable and Sustainable Energy Reviews*, **151**, 111518. doi:10.1016/j.rser.2021.111518
[70] Lee, Z.J., Li, T., and Low, S.H. (2024). ACN-Data – A Public EV Charging Dataset. https://github.com/tongxin-li/ACN-Data-Static Access here
[71] Lee, Z.J., Li, T., and Low, S.H. (2019). ACN-Data: Analysis and Applications of an Open EV Charging Dataset. *Proceedings of the Tenth ACM International Conference on Future Energy Systems*, 139–149. doi:10.1145/3307772.3328313
[72] City of Colorado (2023). Electric Vehicle Charging Station Energy Consumption in the city of Boulder, Colorado. Access here
[73] City of Palo Alto (2020). Electric Vehicle Charging Station Usage (July 2011 - Dec 2020). Access here
[74] Elaad NL (2019). Elaad NL EV Charging Platform. Access here
[75] National Grid Electricity Distribution, UK (n.d.). National Grid - Electric Nation Data. Access here
[76] Electric Nation (2016). Smart Charging Project — Electric Nation. https://electricnation.org.uk/resources/smart-charging-project/
[77] Shen, Y., Fang, W., Ye, F., and Kadoch, M. (2020). EV Charging Behavior Analysis Using Hybrid Intelligence for 5G Smart Grid. *Electronics*, **9**(1), 80. doi:10.3390/electronics9010080
[78] Shahriar, S., and Al-Ali, A.R. (2022). Impacts of COVID-19 on Electric Vehicle Charging Behavior: Data Analytics, Visualization, and Clustering. *Applied System Innovation*, **5**(1), 12. doi:10.3390/asi5010012
[79] Xu, C., Behrens, P., Gasper, P., Smith, K., Hu, M., Tukker, A., and Steubing, B. (2023). Electric vehicle batteries alone could satisfy short-term grid storage demand by as early as 2030. *Nature Communications*, **14**(1), 119. doi:10.1038/s41467-022-35393-0
[80] Xiong, Y., Wang, B., Chu, C.-C., and Gadh, R. (2018). Electric Vehicle Driver Clustering using Statistical Model and Machine Learning. *2018 IEEE Power & Energy Society General Meeting (PESGM)*, 1–5. doi:10.1109/PESGM.2018.8586132
[81] Van Kriekinge, G., De Cauwer, C., Sapountzoglou, N., Coosemans, T., and Messagie, M. (2023). Electric Vehicle Charging Sessions Generator Based on Clustered Driver Behaviors. *World Electric Vehicle Journal*, **14**(2), 37. doi:10.3390/wevj14020037
[82] Gerossier, A., Girard, R., and Kariniotakis, G. (2019). Modeling and Forecasting Electric Vehicle Consumption Profiles. *Energies*, **12**(7), 1341. doi:10.3390/en12071341
[83] Capeletti, M.B., Hammerschmitt, B.K., Silva, L.N.F.D., Knak Neto, N., Passinato Sausen, J., Barriquello, C.H., and Abaide, A.D.R. (2024). User Behavior in Fast Charging of Electric Vehicles: An Analysis of Parameters and Clustering. *Energies*, **17**(19), 4850. doi:10.3390/en17194850
[84] Helmus, J.R., Lees, M.H., and van den Hoed, R. (2020). A data driven typology of electric vehicle user types and charging sessions. *Transportation Research Part C: Emerging Technologies*, **115**, 102637. doi:10.1016/j.trc.2020.102637
[85] Märtz, A., Langenmayr, U., Ried, S., Seddig, K., and Jochem, P. (2022). Charging Behavior of Electric Vehicles: Temporal Clustering Based on Real-World Data. *Energies*, **15**(18), 6575. doi:10.3390/en15186575
[86] Singh, S., Vaidya, B., and Mouftah, H.T. (2022). Smart EV Charging Strategies Based on Charging Behavior. *Frontiers in Energy Research*, **10**, 773440. doi:10.3389/fenrg.2022.773440
[87] Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, **52**(3), 345–370. doi:10.1007/BF02294361
[88] Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, **6**(2). doi:10.1214/aos/1176344136
[89] Bayram, I.S., Saad, A., Sims, R., Babu, A., Edmunds, C., and Galloway, S. (2023). Statistical Characterization of Public AC EV Chargers in the U.K. *IEEE Access*, **11**, 70274–70287. doi:10.1109/ACCESS.2023.3293091
[90] Sadeghianpourhamami, N., Refa, N., Strobbe, M., and Develder, C. (2018). Quantitative analysis of electric vehicle flexibility: A data-driven approach. *International Journal of Electrical Power & Energy Systems*, **95**, 451–462. doi:10.1016/j.ijepes.2017.09.007
[91] Forte, M., Guzman, C.P., Lekidis, A., and Morais, H. (2025). Clustering Methodologies for Flexibility Characterization of Electric Vehicles Supply Equipment. *Green Energy and Intelligent Transportation*, 100304. doi:10.1016/j.geits.2025.100304
[92] Kim, K., Kim, G., Yoo, J., Heo, J., Cho, J., Ryu, S., and Kim, J. (2024). Data-Driven Clustering Analysis for Representative Electric Vehicle Charging Profile in South Korea. *Sensors*, **24**(21), 6800. doi:10.3390/s24216800
[93] Ahmed, S., Khan, Z.A., Gul, N., Kim, J., and Kim, S.M. (2021). Machine Learning-Based Clustering of Load Profiling to Study the Impact of Electric Vehicles on Smart Meter Applications. *2021 Twelfth International Conference on Ubiquitous and Future Networks (ICUFN)*, 444–447. doi:10.1109/ICUFN49451.2021.9528396
[94] Hu, D., Zhou, K., Li, F., and Ma, D. (2022). Electric vehicle user classification and value discovery based on charging big data. *Energy*, **249**, 123698. doi:10.1016/j.energy.2022.123698
[95] Majhi, R.C., Ranjitkar, P., Sheng, M., Covic, G.A., and Wilson, D.J. (2021). A systematic review of charging infrastructure location problem for electric vehicles. *Transport Reviews*, **41**(4), 432–455. doi:10.1080/01441647.2020.1854365
[96] Yi, Z., and Bauer, P.H. (2016). Optimization models for placement of an energy-aware electric vehicle charging infrastructure. *Transportation Research Part E: Logistics and Transportation Review*, **91**, 227–244. doi:10.1016/j.tre.2016.04.013
[97] Carlton, G.J., and Sultana, S. (2022). Electric vehicle charging station accessibility and land use clustering: A case study of the Chicago region. *Journal of Urban Mobility*, **2**, 100019. doi:10.1016/j.urbmob.2022.100019
[98] Lin, Y., Zhang, K., Shen, Z.-J.M., Ye, B., and Miao, L. (2019). Multistage large-scale charging station planning for electric buses considering transportation network and power grid. *Transportation Research Part C: Emerging Technologies*, **107**, 423–443. doi:10.1016/j.trc.2019.08.009
[99] Li, Q., Li, X., Liu, Z., and Qi, Y. (2022). Application of Clustering Algorithms in the Location of Electric Taxi Charging Stations. *Sustainability*, **14**(13), 7566. doi:10.3390/su14137566
[100] Shukla, A., Verma, K., and Kumar, R. (2016). Consumer perspective based placement of electric vehicle charging stations by clustering techniques. *2016 National Power Systems Conference (NPSC)*, 1–6. doi:10.1109/NPSC.2016.7858946
[101] Bezdek, J.C., Ehrlich, R., and Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, **10**(2-3), 191–203. doi:10.1016/0098-3004(84)90020-7
[102] Sánchez, D.G., Tabares, A., Faria, L.T., Rivera, J.C., and Franco, J.F. (2022). A Clustering Approach for the Optimal Siting of Recharging Stations in the Electric Vehicle Routing Problem with Time Windows. *Energies*, **15**(7), 2372. doi:10.3390/en15072372

[103] Sun, Y. (2023). Public Charging Infrastructure Optimization in Urban Neighborhood: Using DBSCAN and Map Matching with EV Trajectories. *2023 International Conference on Electrical, Communication and Computer Engineering (ICECCE)*, 1–6. doi:10.1109/ICECCE61019.2023.10442028

[104] Kalakanti, A.K., and Rao, S. (2022). Charging Station Planning for Electric Vehicles. *Systems*, **10**(1), 6. doi:10.3390/systems10010006

[105] Ankerst, M., Breunig, M.M., Kriegel, H.-P., and Sander, J. (1999). OPTICS: ordering points to identify the clustering structure. *ACM SIGMOD Record*, **28**(2), 49–60. doi:10.1145/304181.304187

[106] Mingoti, S.A., and Lima, J.O. (2006). Comparing SOM neural network with Fuzzy c-means, K-means and traditional hierarchical clustering algorithms. *European Journal of Operational Research*, **174**(3), 1742–1759. https://linkinghub.elsevier.com/retrieve/pii/S0377221705003863

[107] Nespoli, A., Ogliari, E., and Leva, S. (2023). User Behavior Clustering Based Method for EV Charging Forecast. *IEEE Access*, **11**, 6273–6283. doi:10.1109/ACCESS.2023.3235952

[108] Shahriar, S., Al-Ali, A.R., Osman, A.H., Dhou, S., and Nijim, M. (2021). Prediction of EV Charging Behavior Using Machine Learning. *IEEE Access*, **9**, 111576–111586. doi:10.1109/ACCESS.2021.3103119

[109] Crozier, C., Morstyn, T., and McCulloch, M. (2021). Capturing diversity in electric vehicle charging behaviour for network capacity estimation. *Transportation Research Part D: Transport and Environment*, **93**, 102762. doi:10.1016/j.trd.2021.102762

[110] Zhu, J., Yang, Z., Mourshed, M., Guo, Y., Zhou, Y., Chang, Y., Wei, Y., and Feng, S. (2019). Electric Vehicle Charging Load Forecasting: A Comparative Study of Deep Learning Approaches. *Energies*, **12**(14), 2692. doi:10.3390/en12142692

[111] Sun, C., Li, T., Low, S.H., and Li, V.O.K. (2020). Classification of electric vehicle charging time series with selective clustering. *Electric Power Systems Research*, **189**, 106695. doi:10.1016/j.epsr.2020.106695

[112] Liang, M., and Dai, L. (2021). Alternating Minimization Methods for Solving Multilinear Systems. *Mathematical Problems in Engineering*, **2021**, 1–13. doi:10.1155/2021/6629243

[113] Zhang, J., Yan, J., Liu, Y., Zhang, H., and Lv, G. (2020). Daily electric vehicle charging load profiles considering demographics of vehicle users. *Applied Energy*, **274**, 115063. doi:10.1016/j.apenergy.2020.115063

[114] Schäuble, J., Kaschub, T., Ensslen, A., Jochem, P., and Fichtner, W. (2017). Generating electric vehicle load profiles from empirical data of three EV fleets in Southwest Germany. *Journal of Cleaner Production*, **150**, 253–266. doi:10.1016/j.jclepro.2017.02.150

[115] Wu, W., Wang, W., Jia, X., and Feng, X. (2024). Transformer Autoencoder for K-means Efficient clustering. *Engineering Applications of Artificial Intelligence*, **133**, 108612. doi:10.1016/j.engappai.2024.108612

[116] Aghabozorgi, S., Shirkhorshidi, A. S., and Wah, T. Y. (2015). Time-series clustering – A decade review. *Information Systems*, **53**, 16–38. doi:10.1016/j.is.2015.04.007