International Journal of Geo-Information

MDPI

Article

Discovering Electric Vehicle Charging Locations Based on Clustering Techniques Applied to Vehicular Mobility Datasets

Elmer Magsino [1,*], Francis Miguel M. Espiritu [2] and Kerwin D. Go [2]

[1] Physics and Engineering Department, Faculty of Applied and Technical Studies, University of the Fraser Valley, 33844 King Rd., Abbotsford, BC V2S 7M7, Canada
[2] Department of Electronics and Computer Engineering, Gokongwei College of Engineering, De La Salle University, 2401 Taft Ave., Malate, Manila 1004, Philippines
* Correspondence: elmer.magsino@ufv.ca

Abstract: With the proliferation of vehicular mobility traces because of inexpensive on-board sensors and smartphones, utilizing them to further understand road movements have become easily accessible. These huge numbers of vehicular traces can be utilized to determine where to enhance road infrastructures such as the deployment of electric vehicle (EV) charging stations. As more EVs are plying today's roads, the driving anxiety is minimized with the presence of sufficient charging stations. By correctly extracting the various transportation parameters from a given dataset, one can design an adequate and adaptive EV charging network that can provide comfort and convenience for the movement of people and goods from one point to another. In this study, we determined the possible EV charging station locations based on an urban city's vehicular capacity distribution obtained from taxi and ride-hailing mobility GPS traces. To achieve this, we first transformed the dynamic vehicular environment based on vehicular capacity into its equivalent urban single snapshot. We then obtained the various traffic zone distributions by initially utilizing *k*-means clustering to allow flexibility in the total number of wanted traffic zones in each dataset. In each traffic zone, iterative clustering techniques employing Density-based Spatial Clustering of Applications with Noise (DBSCAN) or clustering by fast search and find of density peaks (CFS) revealed various area separation where EV chargers were needed. Finally, to find the exact location of the EV charging station, we last ran *k*-means to locate centroids, depending on the constraint on how many EV chargers were needed. Extensive simulations revealed the strengths and weaknesses of the clustering methods when applied to our datasets. We utilized the silhouette and Calinski–Harabasz indices to measure the validity of cluster formations. We also measured the inter-station distances to understand the closeness of the locations of EV chargers. Our study shows how CFS + *k*-means clustering techniques are able to pinpoint EV charger locations. However, when utilizing DBSCAN initially, the results did not present any notable outcome.

Keywords: clustering techniques; vehicular mobility datasets; silhouette and Calinski–Harabasz indices; spatiotemporal stable network characteristics

Citation: Magsino, E.; Espiritu, F.M.M.; Go, K.D. Discovering Electric Vehicle Charging Locations Based on Clustering Techniques Applied to Vehicular Mobility Datasets. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 368. https://doi.org/10.3390/ijgi13100368

Academic Editors: Wolfgang Kainz, Raja Sengupta, Shivanand Balram and Jorge Rocha

Received: 4 September 2024
Revised: 14 October 2024
Accepted: 17 October 2024
Published: 18 October 2024

1. Introduction

The current transportation sector predominantly still relies on fossil fuels as there are more old combustion-engine vehicles plying the road networks. However, the public acceptance of electric vehicles (EVs) as a means to transport goods and services has tremendously turned the fossil fuel-dominant transportation system into a sustainable transport network. EVs are able to address future energy requirements by reducing fossil fuel dependency and utilizing renewable sources like natural gas, wind, solar, hydro, and nuclear fission energy. With more EVs on the road, there is an automatic reduction in carbon dioxide emissions, creating a healthier environment and reducing the effects of climate change. Another factor contributing to the fast acceptance of EVs is the government policies like subsidies and financial incentives [1].

However, one major drawback of driving EVs compared to internal combustion engine (ICE) vehicles is the anxiety of drivers that they may not be able to reach their destination with the current state of charge of their battery [2]. Currently, ICE vehicles running out or low in fuel, would simply look for a petrol station as they are very much distributed spatially, and refueling happens within several minutes. To combat this, some EVs have energy storage that provide high power and has a high energy capacity to reduce charging time [3]. However, in the long run, the most viable solution is the installation of EV charging stations that are accessible to EV users, just like how petrol stations are found along highways and urban roads. The main question related to this deployment is where to install charging stations.

The work in [4] presented a design and technological overview on how to develop and where to locate electric vehicle charging stations. Various considerations were enumerated in their work such as battery swapping, off-grid charging, and charging station reliant on grid power, energy storage, and renewable energy. With the abundance of vehicular mobility data and proliferation of artificial intelligence, other studies have published works focusing on using machine learning for EV charging behaviors [5], characterizing EV mobility pattern based on their trajectories [6], and extracting travel patterns in an urban location [7].

The work in [8] focused on the allocation of charging stations in pre-existing gasoline stations, as they would be intuitively visited by their patrons. Among these existing stations, the study selected those that had a collective number of parking events to draw traffic that would allow them to stay as their vehicles were charging in hour-intervals. On the other hand, the works in [9,10] used taxi mobility traces to represent the relationship of re-charging and travel behavior of possible EV users. Thus, the studies allowed for more insights in the relationship between battery range, charging station allocation, and electrification rate of vehicle miles traveled. In addition to stating that it is most likely that public fleets will be one of the first to integrate EV technology, it can be seen that big data methodologies have been proven to provide sufficient insight into EV user behavior and decision-making. Therefore, the aforementioned ridesharing traces have been used to emulate real-world driving profiles in order to make a more realistic strategy for EV charging infrastructure deployment.

In this study, we formulate a deployment strategy based on traffic zones that are characterized by a vehicular attribute or their combination. Figure 1 illustrates the workflow of the study undertaken. Given a vehicular mobility dataset and after preprocessing, we chose a network feature (e.g., vehicular capacity, speed, environment information) to characterize a given urban map. Dataset preprocessing was performed to ensure mobility feature uniformity. Realizing that the vehicular traces varied both in time and space, we calculated its spatiotemporal stable network characteristic first. This method provided a single snapshot of the urban map under study. From this stable snapshot, we then explored various clustering techniques to reveal areas exhibiting the same network behavior.

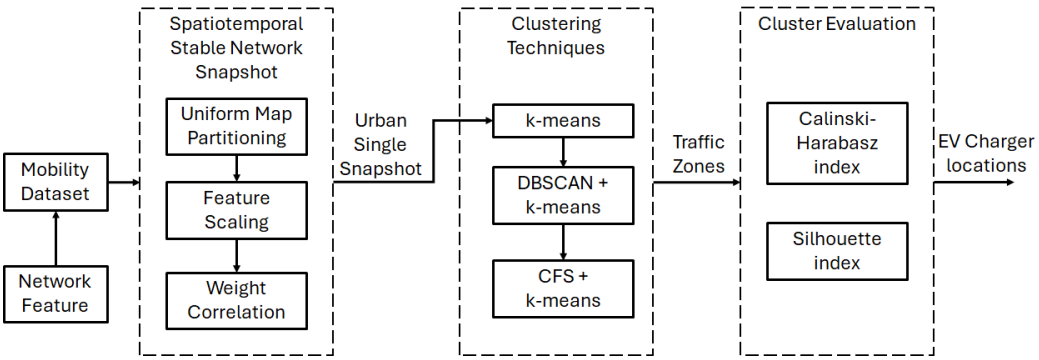

Figure 1. Block diagram of the study.

In addition to formulating a sustainable EV charger deployment strategy, this study also aimed to introduce a profit density-based spatiotemporal analysis. Spatiotemporal studies regarding electric vehicles have already been conducted, most commonly with the prediction of charging demand based on different variables. For example, charging patterns and power demand were predicted using a spatiotemporal model based off Markov chain modeling [11], Monte Carlo simulation [12], and origin–destination analysis [13,14]. Ref. [15] derived a model using the fluid dynamic traffic model and the $M/M/s$ queuing theory to capture spatiotemporal dynamics of charging demand. Another paper by [16] used discrete event simulation (DES) to determine the effects of increasing EV supply equipment on charging peak demand. Finally, a similar paper by [17] used spatiotemporal modeling for EV charging demand and its effects on peak household electrical load. As a result, a spatiotemporal methodology was successfully integrated into the study; however, it was based off a profit-density analysis of various parameters such as vehicular flow and traffic volume.

In this work, we clustered spatiotemporal urban mobility traces of taxis and ride-hailing services from the cities of Beijing, Singapore, and Jakarta to formulate strategic EV charging infrastructure plans. Our work was limited to studying urban and rural places which could be characterized by their rich empirical vehicular mobility datasets. Places with sparse vehicular representation were excluded in this study. Since the features of the available datasets were varied, we only chose the common characteristics for comparison and contrast provided by our methodologies utilizing various clustering techniques. The major contributions of this work are enumerated below.

1. From the mobility datasets, we determine its spatiotemporal network characteristics to statically represent a dynamic vehicular environment into a singular map. This spatiotemporal stable map captures the time- and space-varying characteristics of the entire urban setup, such as transportation operation dynamics, vehicular distribution, and interactions.
2. We utilize three unsupervised clustering algorithms and their combinations to determine the urban traffic zones. These are regions that share mobility characteristics' similarity based on vehicular capacity and can be extended to other network characteristics such as vehicular speed or their combination.
3. Extensive simulations are carried out on three empirical mobility datasets to determine where EV charging stations are to be installed. We utilize the silhouette and Calinski–Harabasz indices to validate our traffic zone formations and effectively, the locations of the EV chargers.

The paper is outlined as follows. Section 2 provides the materials and methodologies used in this study. Section 3 discusses our simulation results while the discussion of our method's performance is found in Section 4. Finally, the conclusion and possible improvements for future work are tackled in Section 5.

2. Review of Related Literature

Much research has been conducted regarding the problem and issues in the determination and installation of EV chargers' possible locations. A comprehensive review in [18] categorized the EV charger deployment problem as a subset of the facility location problem where a certain set of criteria is optimized such as traffic volume and activity, user behavior, operating and capital expenditures, installation cost, and urban points of interest.

One of the earlier studies involved the analysis of flow capturing location–allocation models (FCLM) that determined a predetermined number of charging station locations such that the demand coverage was maximized [19]. In [20], the current petrol stations situated in five urban environments were studied in their simulation tool and utilized as places for deploying fast chargers based on population and road traffic. As the coverage diameter was increased from 5 to 10 km, cities exhibited a wide range of demand coverage, i.e., 53–99%, where a varying number of charging stations, based on traffic and population, were needed to achieve that coverage, i.e., 2–47 chargers. The same background was used

in [21], but the authors utilized linear integer programming to optimize their deployment objective. In another paper, ref. [22] incorporated the dynamic charging demand based on origin–destination (OD) data of 15 min and shorter intervals. The study effectively estimated the spatiotemporal distribution of the city under study. Another work [23] utilized a modified pagerank to study OD data and estimate spatial charging demand distribution. In [24], the deployment of EV chargers focused more on the use of lands belonging to a particular city and maximizing charging profit, which attracted or deterred EV users. In [25], the authors placed their EV chargers at parking stations found in dense residential areas with constraints focusing on the uncertainty of the demand and its impact on the distribution grid.

Another criterion used in the deployment of EV chargers focused on the installation of varying levels of EV chargers. Ideally, level 3 chargers must be installed across an urban location because they take less time to fill up the battery. However, this plan is not economical because it can ruin the electric grid capacity of a place because of the high energy consumption introduced by level 3 chargers [26]. In [27], the power quality of the electric grid was analyzed as the main component where to place EV chargers. Uncoordinated charging among EVs creates challenges and issues for the current power distribution setup [28].

Ref. [29] introduced a flow refueling location model to reflect the real-world movements of vehicles, as their trajectory and decision-making are often influenced by the need to refuel. Anticipating the possibility of vehicles having to stop at more than one station along its path highlighted the necessity of taking into consideration vehicle user behavior, decision-making, and refueling habits. These involved the observation of driver travel patterns and rules followed within a certain observation window. Once applied to the allocation of EV charging stations, such methodologies do not only take into account the range of adjustments individual drivers make in a day-to-day drive but also the spontaneity of deviating from set behaviors once presented with unforeseen circumstances (detours, heavy traffic, and vehicular accidents). For example, ref. [21] found that the installation of charging stations in public parking lots had the potential to significantly increase the mileage from fuel to electricity. Ref. [30] attempted to determine the location of hydrogen refueling stations by proposing a mixed-integer routing model that took into consideration a set of predefined decision makers. It was found that the only downside to such an approach was that addressing multiple possibilities from a series of potential decisions to be made by the driver could greatly increase the processing time. Other studies [31,32] aimed to maximize social welfare opportunities by studying a network equilibrium, in which EV driver decisions are dependent on the availability of stations and the necessity of charging the car. Other studies involved looking at areas of interest to determine key location selection factors, while taking into consideration spontaneous EV driver behavior and decision-making processes. For example, ref. [33] found that transport hubs, parking spaces, and urban points of interest were important aspects to station selection through the use of a spatial model function and stakeholder interviews. Other studies regarding this topic even took an optimization approach, such as [34]. Here, a gray wolf optimization algorithm based off a Markov chain network model was proposed with the objective to maximize net profit in order to determine charging station locations. Similar studies with optimization methods involved integer linear programming [35], mixed-integer linear programming [36], and genetic algorithms [37].

Different from the mentioned research studies above, this work incorporates all mobility traces of all trajectories in a dataset to determine traffic zones where all viable locations of a limited number of EV chargers can be deployed. By utilizing clustering techniques, we leverage their capabilities to form traffic zones even if the areas are shared by multiple boundaries set by their geographical areas. We focus also on the deployment of charging stations that offer free charging to the public. This work can also be categorized as a subset of the maximum coverage problem since it is guaranteed through clustering that all mobility points belong to one coverage area.

3. Spatiotemporal Mobility Datasets

In this section, we describe the mobility datasets utilized in this study. There were three datasets employed with two different saving formats and varying mobility features.

Three mobility datasets from Beijing, Singapore, and Jakarta were used in this study to determine where the EV chargers should be deployed in each city. The Singapore (SIN) and Jakarta (JKT) Global Positioning Systems (GPS) traces had the same storage format that is different from how the Beijing (BJG) mobility traces were stored. The SIN and JKT traces were already preprocessed and unwanted noise was removed [38], unlike the BJG dataset that still contained noise and deviations in both latitude and longitude [39]. Table 1 summarizes the dataset characteristics. The JKT traces were composed of cars and motorcycles trajectories. We excluded those motorcycle traces in our study.

Table 1. Mobility dataset attributes

Urban City Characteristic	**BJG**	**JKT**	**SIN**
Total area (in $\approx$ km^2)	51	51	51
Number of days of data	7	14	14
Total number of recorded time each day (in hours)	24	24	24
Total number of vehicles	24,845	28,000	16,174
Daily average number of trips	79,012	4000	2000
Sampling time T_S (in second)	10	1	1

The SIN and JKT datasets had the following features: Trajectory ID, Driving Mode, OS Name, Timestamp, Lat, Lon, Speed, Bearing, and Accuracy. These labels are defined below.

1. Trajectory ID identifies the journey made by a transportation mode and has a string format.
2. Driving Mod indicates whether a 'car' or a 'motorcycle' made the trajectory.
3. OS Name categorizes whether an IOS or an Android smartphone captured the trajectory.
4. Timestamp—time, measured in Coordinated Universal Time, when the location is sensed.
5. Lat and Lon gives the current vehicular location in terms of latitude and longitude coordinates.
6. Speed—(in meter/sec) instantaneous speed of the vehicle.
7. Bearing—(in degrees) measurement relative to the true north.
8. Accuracy is OS-dependent (IOS or Android) and provides the accuracy with respect to the horizontal plane.

On the other hand, the BJG dataset only had four labels as enumerated below.

1. Trajectory ID identifies the journey made by a taxi.
2. Timestamp—time, measured in multiples of its sampling time, when the location is sensed.
3. Lat and Lon gives the current vehicular location in terms of latitude and longitude coordinates.

Given the differences in the dataset's features and characteristics, we anticipated that various mobility traces datasets would have their own collection and storage formats. Ideally, the target mobility dataset should contain the timestamped trajectories of a unique vehicle identified through its identification number. To ensure that this work was flexible for application on other available datasets, a preprocessing procedure was implemented. The taxi traces were formatted to achieve feature uniformity. The SIN and JKT datasets were sorted by trip ID and then by timestamp. Afterward, all pertinent traces from the three sets were stored into two arrays, namely, origin-destination trace array and the taxi trip trace array. These arrays helped in easily identifying multiple trajectories at a time.

The origin–destination (OD) array was a summary of a taxi's starting point and its end point for a particular trip, both denoted by the longitude, latitude, timestamp, and taxi ID. On the other hand, the taxi trip trace array contained the traces of a whole trip, τ_v of vehicle v. We defined $\tau_v = \left(\lambda_1, \lambda_2, \ldots, \lambda_k, \ldots, \lambda_{N-1}, \lambda_N\right)$, where λ_1 and λ_N are the origin and destination coordinates of the recorded trip, and λ_k are the intermediate coordinates of the trajectory [40].

One final note regarding our datasets is that suburban vehicular networks are sparse in nature because more vehicles tend to pick up and drop off passengers in places where commercial and industrial areas are found. We see in Section 5.1 that the urban road networks are easily visualized at the urban map's center because of the area's vehicular density. As vehicles move away from the center, roads are scattered and less formed.

4. Methodology

In this section, we present the procedures utilized in determining the strategic locations where the EV charging stations should be deployed. We first illustrate how to transform the dynamic urban mobility dataset into a single snapshot to characterize its urban map. Finally, we discuss how we used three clustering algorithms to determine the city's traffic zones and the locations of EV chargers.

4.1. Spatiotemporal Stable Network Characteristics of an Urban Map

The acquired empirical mobility datasets varied both in time and space; thus, there was a need to capture the city's dynamic behavior in one snapshot that would describe its movements. To transform the spatiotemporal variations of a dataset, we employed the spatiotemporal stable network characteristics concept presented in [41].

An urban map that comprised dynamic and static data and vehicular GPS traces, as shown in Figure 2, was uniformly partitioned into $N \times N$ map grids, $g_{p,q}$. We defined a utility function, $\zeta_{p,q}$, at time t to characterize each map grid, $g_{p,q}$. The indexes p and $q \in \{1, \ldots N\}$. We defined $\zeta_{p,q}$ as $E\left[\eta_{p,q}\right]$ or the expectation of $\left[\eta_{p,q}\right]$.

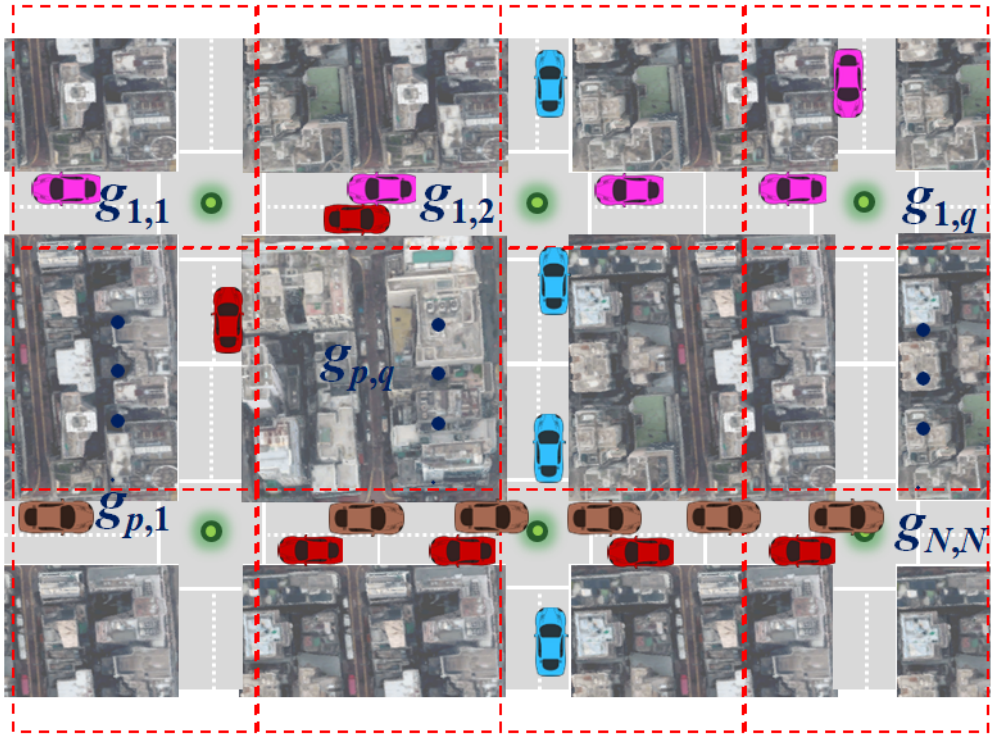

Figure 2. The urban map is uniformly partitioned to reveal different $g_{p,q}$ and its utility network parameters at sampling time $t = iT_S$. Vehicles of the same color represent their respective trajectories.

$\eta_{p,q}$ characterized the grid's spatial network characteristic at time t, e.g., environment data, vehicular capacity, and density. There was an assumed network characteristics independence. With this, we also developed a utility function that was a combination of any of these $\eta_{p,q}$'s. In this work, $\eta_{p,q} = c_{p,q}$, the vehicular capacity.

Given our spatiotemporal datasets, we derived a single urban city snapshot by utilizing all sampling times, iT_S, where $i \in \{1, 2, \ldots, I\}$, $I = \frac{24 \times 60}{T_S}$. T_S is the sampling time. Equation (1) establishes the map grid $g_{p,q}$'s spatiotemporal stable network characteristic, $\zeta_{p,q,STS}$.

$$\zeta_{p,q,STS} = \sum_{i=0}^{I} \alpha(iT_S)\omega(iT_S), \tag{1}$$

where

$$\alpha(iT_S) = \frac{\zeta(iT_S) - \min\left[\zeta(iT_S)\right]}{\max\left[\zeta(iT_S)\right] - \min\left[\zeta(iT_S)\right]}$$
$$\omega(iT_S) = \frac{\zeta(iT_S)}{\max\left[\zeta(i = 0, \ldots, IT_S)\right]}.$$

$\alpha(iT_S)$ and $\omega(iT_S)$ are the feature scaling parameter and weight at time $t = iT_S$ correlating all the $\zeta_{p,q}$'s, respectively. Given $N = 20$, an illustrative example of how (1) converts a dynamic spatiotemporal dataset into a single urban vehicular snapshot is shown in Figure 3.

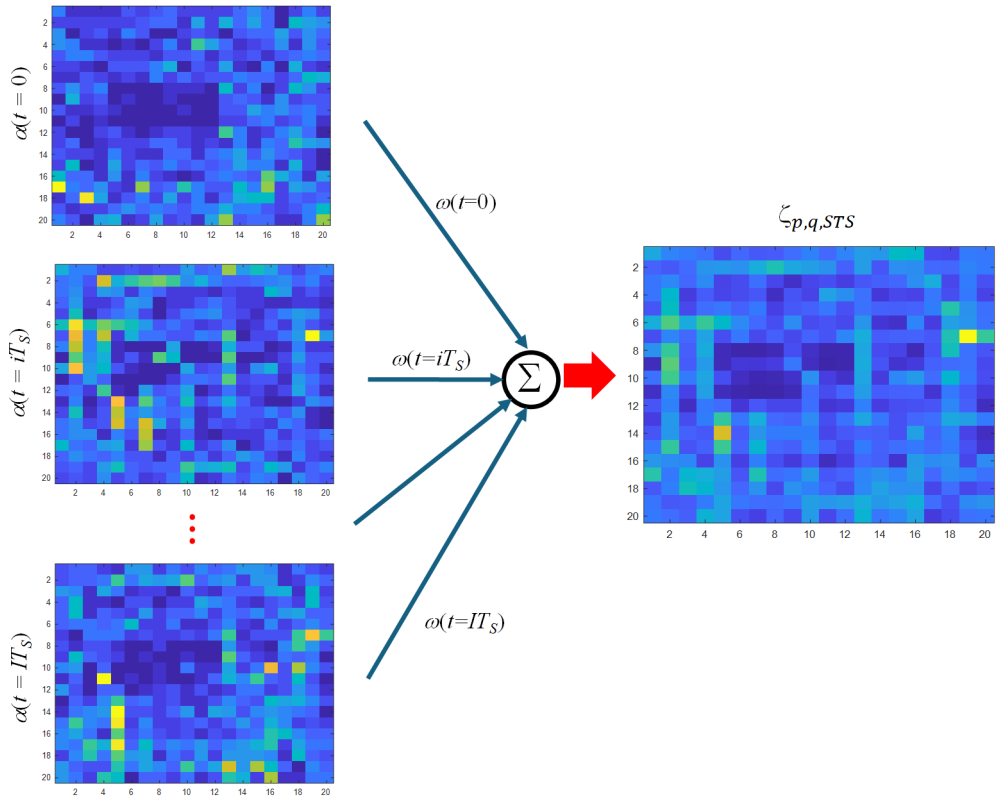

Figure 3. Using vehicular capacity, the dynamic urban vehicular map is transformed into a snapshot where darker colors represent low vehicular capacity and lighter colors show places with high vehicular capacity [41].

4.2. Clustering Algorithms to Determine Traffic Zones and EV Charging Locations

The spatiotemporal stable network characteristics of an urban map were represented as a single snapshot obtained from processing the mobility dataset. From this spatiotemporal urban map, we performed three iterative clustering algorithms to obtain vehicular traffic zones for possible EV charging locations. The three unsupervised machine learning algorithms employed were (1) *k*-means clustering, (2) Density-based Spatial Clustering of Applications with Noise (DBSCAN) [42] plus *k*-means, and (3) clustering by fast search and find of density peaks (CFS) plus *k*-means [43]. We describe each process below and state the procedures we undertook.

4.2.1. $k-$Means Clustering for Non-Uniform Partitions

To adapt to the boundary definitions of a given urban city based on available mobility datasets, we employed *k*-means clustering to generate *k* traffic zones across the city. In an actual implementation, cities can be divided according to natural boundaries set by the researcher/deploying agency. In this study, we left those conditions out and focused on available vehicular GPS traces. In summary, *k*-means places a point to the nearest cluster center following a distance metric [44]. Below is the procedure of utilizing *k*-means for initially partitioning a map under study.

1. The number of *k* clusters is experimentally varied from 50 to 500 with an interval of 50. Each cluster becomes a traffic zone and is labeled as Z_k.
2. Each Z_k obtained from every simulation run for a defined *k* is validated by the silhouette index analysis.

Figure 4 shows an example of how *k*-means clustering is able to divide an urban map's GPS traces in (a) to four traffic zones in (b) following the procedure stated above, when $k = 4$.

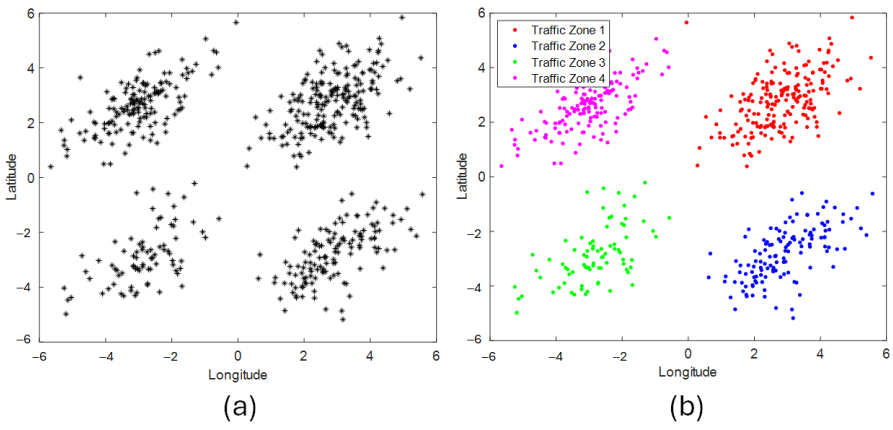

Figure 4. Example of $k-$means clustering applied to randomly generated GPS data (**a**). When $k = 4$, the four traffic zones are shown in (**b**). Z_1 is represented by the magenta color (**upper left**). Z_2 is represented by the red color (**upper right**). Z_3 is represented by the green color (**lower left**). Z_4 is represented by the blue color (**lower right**).

One may argue that a simple uniform partitioning of an urban map avoids computational complexities. However, employing *k*-means allowed us to handle noisy datasets, such as the BJG dataset, and avoided visual inspections of traces at the boundaries. Also, uniform partitioning greatly depended on how high the resolution was, which was addressed by varying the *k* value.

4.2.2. DBSCAN Clustering Plus *k*-Means Clustering

For each traffic zone, Z_k, we utilized DBSCAN to recognize non-spherical clusters. Effectively, partitions formed by DBSCAN exposed the road network. The following procedure summarizes how we performed DBSCAN for each Z_k.

1. We defined the radius between two points, ϵ, and minimum number of points to form clusters, m. Each DBSCAN cluster was labeled Z_{k_d}, in Z_k. ϵ and m were two parameters that were initialized and experimented to capture the least number of clusters within the initial traffic zones Z_k. We targeted the least number of EV charging locations to minimize cost deployment of uniform chargers, but it could be extended depending on the optimization cost.
2. From each Z_{k_d}, we re-ran the *k*-means clustering to partition and eventually determine the EV charging locations, providing the location of the possible EV charging stations through the center of each formed cluster.

In an actual implementation, the researcher/deploying agency may set the desired ϵ and m values to produce their desired partitions. Figure 5 is an example of how the procedure was implemented. DBSCAN clustered the original data in Figure 4 with $\epsilon = 1$ and $m = 20$, producing only three viable groups. Each DBSCAN cluster was further clustered into four sub-clusters using *k*-means clustering to determine the smaller traffic zones within the DBSCAN-based traffic zones.

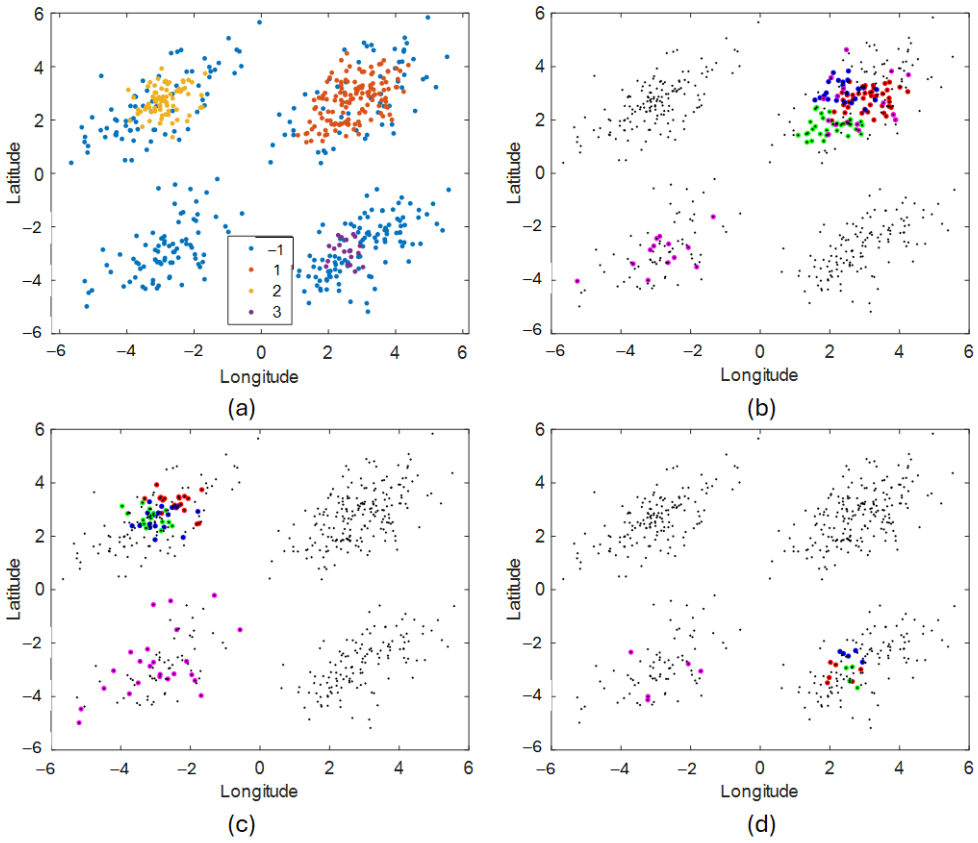

Figure 5. (**a**) Example in Figure 4a clustered using DBSCAN producing three clusters with many outliers represented by −1. (**b**–**d**) are the *k*-means clustering results when $k = 4$ for each DBSCAN cluster. These were superimposed onto the original data to show how DBSCAN performed reduction in the original dataset. Colors represent the cluster determined by DBSCAN in (**a**) and $k-$means clustering in (**b**–**d**).

4.2.3. CFS Clustering Plus *k*-Means Clustering

Another way to group data points for each traffic zone, Z_k, in a non-spherical clustering is through CFS. CFS defines clusters as the peaks in the density of points and determines the cluster centroid based on the outliers when the minimum distance is plotted as a function of density. The following procedure summarizes how we performed CFS for each Z_k.

1. For each point in Z_k, we determined the local density (number of points around it).
2. For those high-density points, we computed the distances of all involved points, then took the minimum value.
3. We plotted the minimum distance as a function of density. The outliers in that graph determined the cluster centers and the number of clusters.

From the CFS decision graph, we decided how many outliers would be considered to determine the number of partitions to be analyzed. We then re-ran the *k*-means clustering to determine the centroid locations. Like the DBSCAN approach, the center of each formed cluster served as our possible EV charging locations. Figure 6 illustrates how the original data in Figure 4a were partitioned using CFS. Notice that there are four outliers in Figure 6b. Out of these four, we only chose the two most outlying (found in the upper right). These clusters are shown with a blue diamond Figure 6c.

Notice the differences in results outputted by DBSCAN and CFS. To determine the EV locations, as much as possible, these two clustering techniques must provide the same traffic zone results.

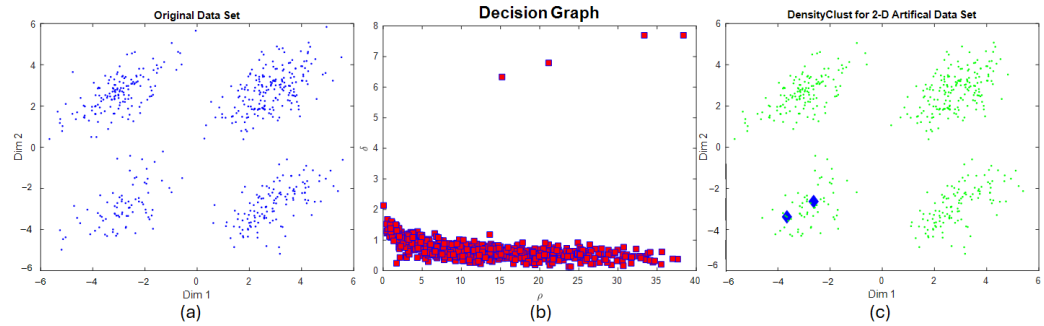

Figure 6. Example in Figure 4a clustered using CFS having the two outliers as the main traffic zone within the four clusters derived from $k-$means clustering. (**a**) Original Data, (**b**) Distance vs. Density plot, and (**c**) two chosen outliers from (**d**) as the cluster center represented by the blue diamond.

5. Results

We present here our extensive simulation results after applying our proposed methodology to the BJG, JKT, and SIN datasets. For simulation uniformity, we set the sampling time to $T_S = 10$ s, since this was the smallest resolution of the BJG dataset.

In preprocessing the BJG, JKT, and SIN datasets, we only chose the features common to all datasets, i.e., Trajectory ID, timestamp, latitude, and longitude coordinates. The other features found in the JKT and SIN datasets were simply disregarded.

5.1. Dataset Statistics and Formation of Spatiotemporal Stable Urban Maps

The hourly vehicular capacity and average speed of each dataset are shown in Figures 7 and 8, respectively. These data were used in (1) to arrive at the single snapshot of the city's dynamic vehicular environment. JKT and SIN exhibited approximately the same behavior with peak capacities occurring from 9:00 to 11:00 A.M. and then decreasing until there were almost no vehicles from 5:00 to 10:00 P.M. Unlike these two, BJG was almost always busy from 6:00 A.M. to 12:00 midnight. On the other hand, when comparing the three datasets, the average speed showed slow-moving vehicles for BJG, average-moving vehicles for JKT, and fast-moving vehicles for SIN.

From Figures 7 and 8, we then determined the environment's spatiotemporal stable network characteristics. Figures 9 and 10 show the city snapshot when the network

characteristics were vehicular capacity and speed. Notice that the road networks of each urban location is revealed by observing the lighter color sections. Unlike JKT and SIN, the characteristic stability of BJG was scattered, which could be attributed to the presence of noise in that dataset. BJG suffered from missing and deviated mobility traces. We note that JKT and SIN traces were captured by smartphones, while BJG traces were obtained from GPS equipment installed in taxis. The lighter colors of Figures 9 and 10 illustrate the city's most traveled routes or roads and which places were traversed at faster speeds. These findings can be used to identify road congestion and possible locations of EV chargers if based on demand.

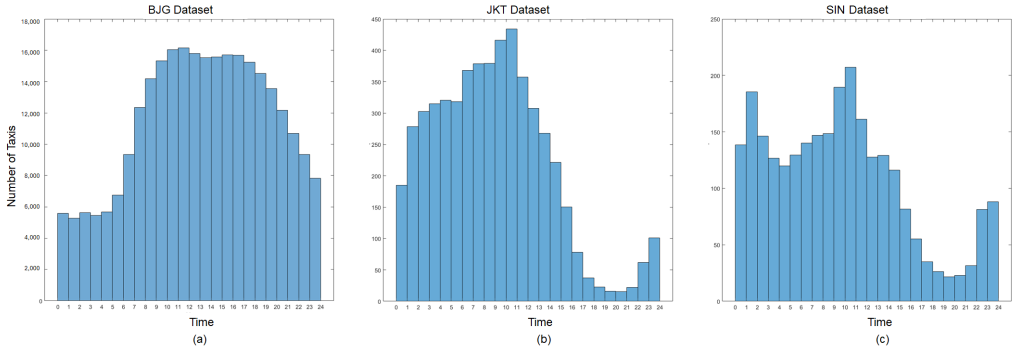

Figure 7. The hourly vehicular capacity of (**a**) BJG, (**b**) JKT, and (**c**) SIN.

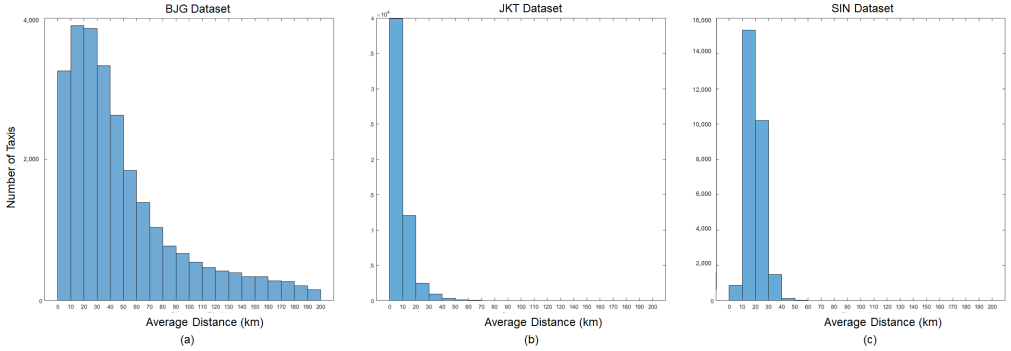

Figure 8. The hourly vehicular speed of (**a**) BJG, (**b**) JKT, and (**c**) SIN.

It is also noticeable that the capacity and speed were complementary of each other. The higher the vehicular capacity, the slower the speed, and vice versa. This was very much observable in the BJG dataset.

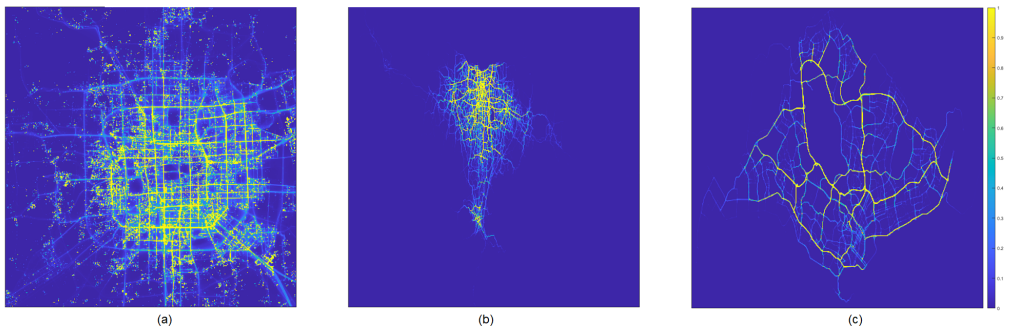

Figure 9. The spatiotemporal stable vehicular capacity network characteristic snapshot of (**a**) BJG, (**b**) JKT, and (**c**) SIN. Lighter colors depict high values when compared to dark colors.

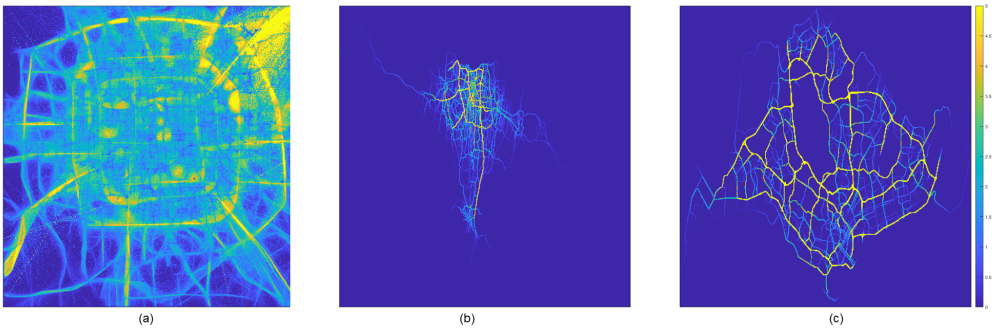

Figure 10. The spatiotemporal stable vehicular speed network characteristic snapshot of (**a**) BJG, (**b**) JKT, and (**c**) SIN. Lighter colors depict high values when compared to dark colors.

5.2. Clustering Results of Spatiotemporal Stable Urban Maps

Between vehicular capacity and speed, we present clustering results obtained from vehicular capacity spatiotemporal stable urban maps. Figure 11 shows the various clusters formed from the BJG, JKT, and SIN datasets when $k = 50, 300, 500$, though $k = 50, 100, \ldots, 500$. As k was increased, we saw smaller cluster sizes and more locations where EV chargers could be deployed. During the implementation, k can be constrained by the number of allowable EV chargers to be installed.

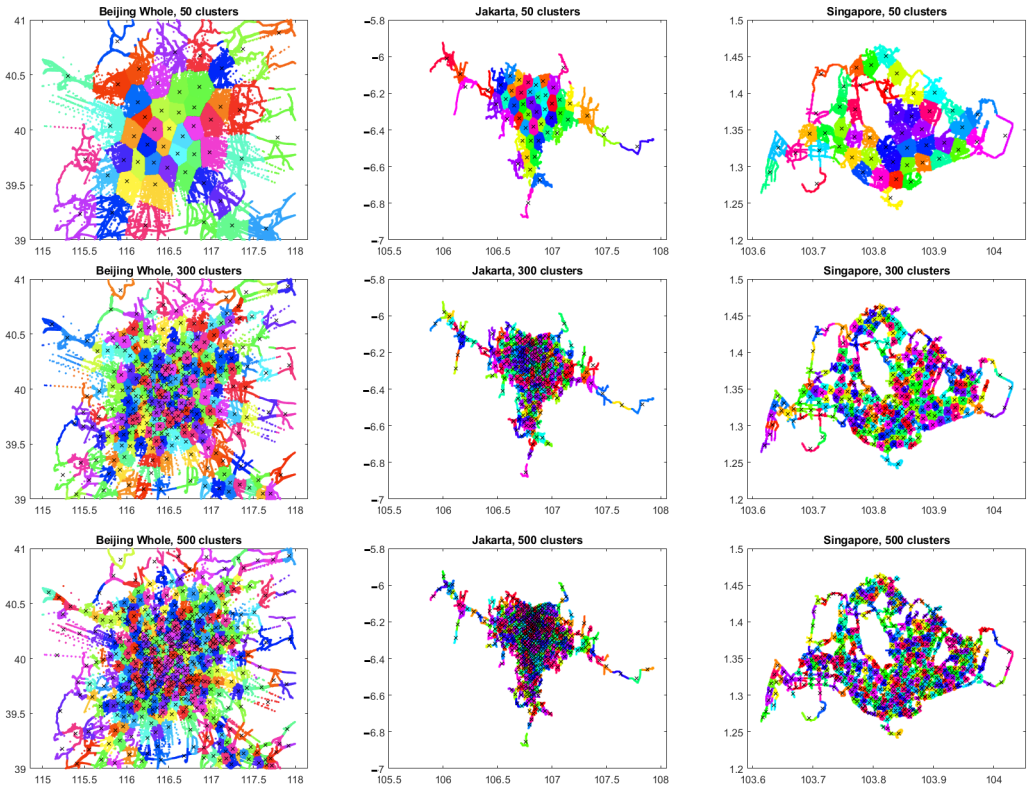

Figure 11. $k-$means clustering results when $k = 50, 300, 500$ for BJG, JKT, and SIN. The first row is $k = 50$, the second row is $k = 300$, and the third row is $k = 500$. The first column is clustering BJG, the second column is clustering JKT, and the third column is clustering SIN. "$\times$" denotes the cluster center of the colored cluster formed by $k-$means.

For each simulation run producing k clusters in each of the datasets, we validated the formation by checking the similarity of points within the cluster and its dissimilarity with the other clusters. This is graphically presented through a silhouette plot [45]. The silhouette

index for each k value for each dataset is shown in Figure 12. A silhouette index approaching one (1) means that the points in a cluster are placed correctly, while a negative one (-1), means otherwise.

Evidently, the JKT and SIN datasets had most silhouette indices approaching 1, while the BJG dataset had almost an equal number of clusters with -1 and 1 indices. The reason behind this result was because at the cluster boundaries of the BJG urban map, there lay GPS traces that were close together, eventually lowering down the silhouette index. This meant that a GPS point could be situated in one of the clusters nearest to it. Empirically, one may interpret close GPS traces as two or more vehicles found on the same highway or road at the same time but traveling in opposite directions and their lanes separated by an island. Unlike BJG, the JKT and SIN datasets did not have this problem because (1) the GPS points were clean and (2) there were not many overlapping traces on roads compared to BJG because of fewer vehicles.

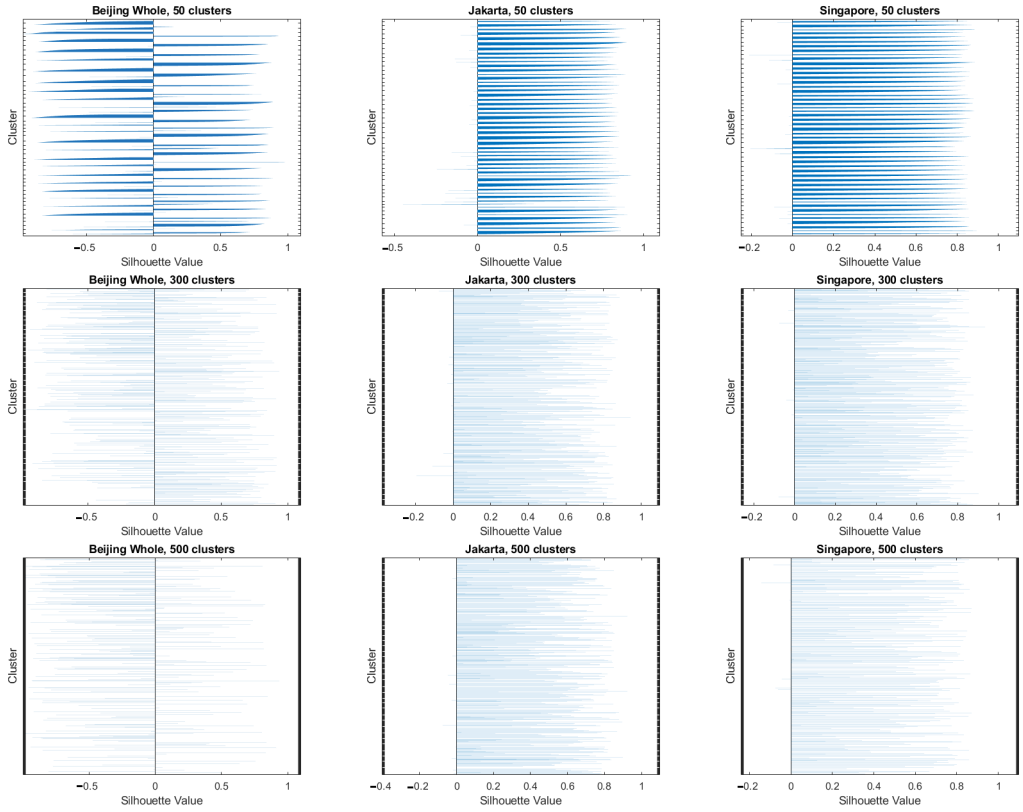

Figure 12. Silhouette evaluation after performing k-means clustering when $k = 50, 300, 500$ for BJG, JKT, and SIN. The first row is $k = 50$, the second row is $k = 300$, and the third row is $k = 500$. The first column is clustering BJG, the second column is clustering JKT, and the third column is clustering SIN.

For all datasets, non-polygonal clusters were formed at the city boundaries and were equivalent to the road configurations at those locations. This is due to the fact that there were fewer vehicular trips happening at the urban map edges. This could be remedied by limiting the boundaries on purpose.

6. Discussion

In this section, we discuss how to obtain the EV charging station locations based on the clusters formed in the previous section.

6.1. Evaluating k-Means Clusters Through Distance Metric

The inter-cluster distances when $k = 50, 500$ are depicted in Figure 13. As k increases, the distribution illustrates a clearer picture on how far each station is from the rest.

All datasets exhibited a distance distribution skewed to the left, signifying a shorter travel time between two clusters. This is beneficial for road travelers when they need to go to the next EV charging station because the current location is already full or there is now a queue of vehicles waiting to charge their vehicle. Since SIN was the smallest city among the three datasets, it was evident that cluster centroids were nearer to each other when compared to the other two, especially in BJG, which was the biggest city among the three.

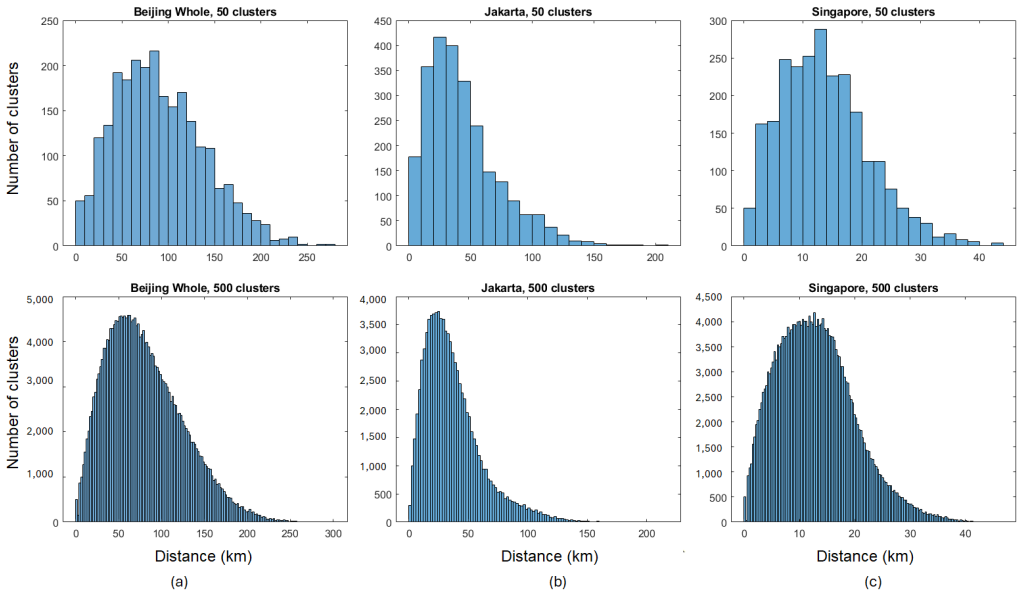

Figure 13. Inter-cluster distances when $k = 50$ (first row) and $k = 500$ (second row) for (**a**) BJG (first column), (**b**) JKT (second column), and (**c**) SIN (third column).

6.2. EV Charging Locations Based on CFS

From the set of formed clusters, Z_k, for each mobility dataset from Figure 11, we selected partition Z_{30} from the SIN dataset to evaluate and determine EV charger locations using CFS. Other partitions could also be chosen, depending on what the researcher/deploying agency wants to analyze.

The upper right of Figure 14 shows the separation of partition $Z_k = 30$ into two clusters when we decided to choose the two outliers (encircled square points) found in the decision graph of the upper left part of Figure 14. Notice that in the decision graph, the majority of square points were found near zero. Those above $\delta > 0.002$ could be considered possible further clusters in partition $Z_k = 30$. The more outliers we selected, the smaller the regions partition $Z_k = 30$ would provide. From these two clusters, we could decide how many EV charging stations to deploy based on available constraints set by the city administration. In our example, we ran k-means clustering to locate four places in each cluster. The EV charging locations were situated at positions labeled with "$\times$" and highlighted by the black circle.

The silhouette index obtained from these formed clusters was calculated to be above 0.8, which was a good indication of cluster formation. The inter-cluster distances for the two clusters divided further into four regions were eight (8) and five (5) km, respectively. In reality, an EV can always go to the area's designated EV charging station to charge, provided there is a vacant slot for charging. We highlight that determining the number of available chargers in an EV charging station was beyond the scope of this research.

Finally, we used the Calinski–Harabasz (CH) index [46], shown in Figure 15, to determine the maximum number of EV charging stations that could be deployed in a given cluster of partition 30, Z_{30}. The maximum allowable number of EV chargers that could be deployed in the two clusters were 20 and 15, respectively, as depicted by the highest CH value.

One important observation revealed by the lower right section of Figure 14 is the possibility of having EV chargers inaccessible to roads or situated in private spaces. Such scenario would require more than engineering design and need government intervention.

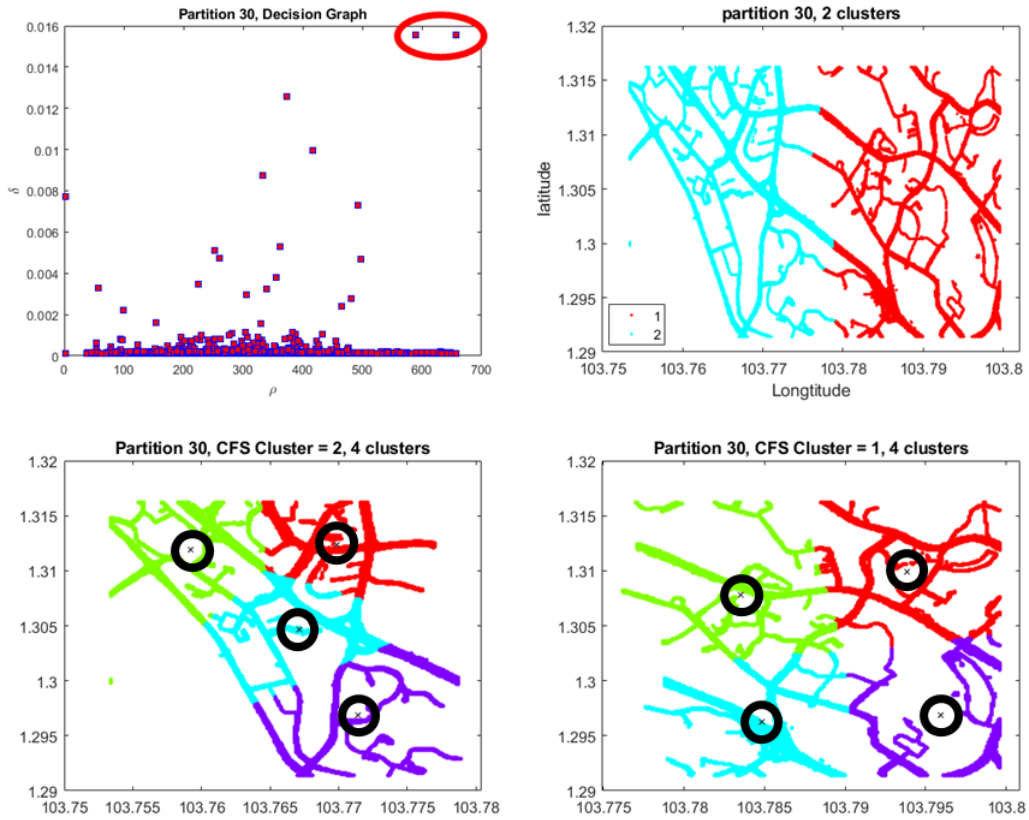

Figure 14. Determining the locations of EV chargers from partition 30 of the SIN dataset calculated by CFS. The upper left shows the decision graph, the upper right shows the clusters when two outliers were chosen, represented by cyan and red cluster groups, and the second row shows clusters 1 and 2 further divided into four subareas, represented by four different colors, using *k*-means. "×" denotes EV charger locations.

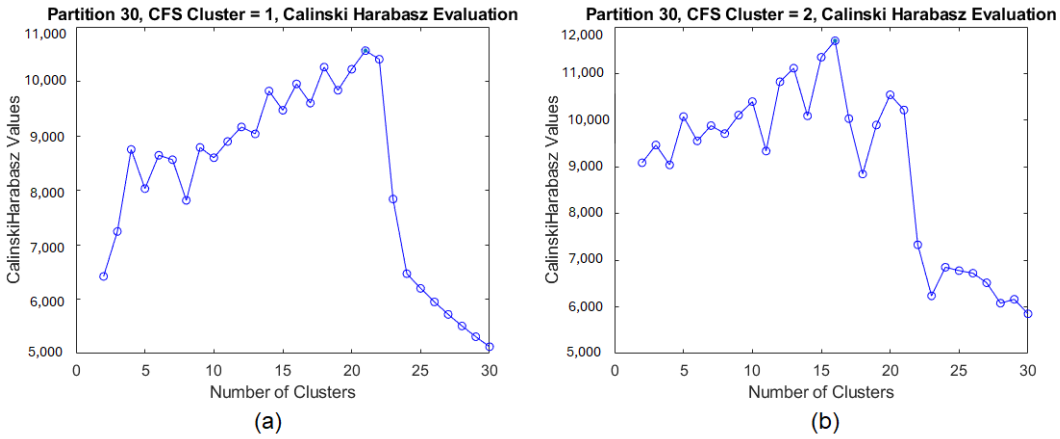

Figure 15. Determining the most number of allowable EV charging stations in clusters (**a**) 1 and (**b**) 2 of partition 30.

6.3. EV Charging Locations Based on DBSCAN

We report here that when using DBSCAN in our datasets, especially in JKT and SIN, we only had one returned cluster regardless of the values that we set for ϵ and m.

This was effectively equal to the cluster formed from the initial *k*-means procedure. For the BJG dataset, DBSCAN was most useful in eliminating mobility traces' outliers such as roads less traveled.

With these outcomes, we implemented a secondary *k*-means clustering in the previous derived *k*-means cluster. We show results and evaluation indices when $k = 50$ and $k = 500$ in Figure 16.

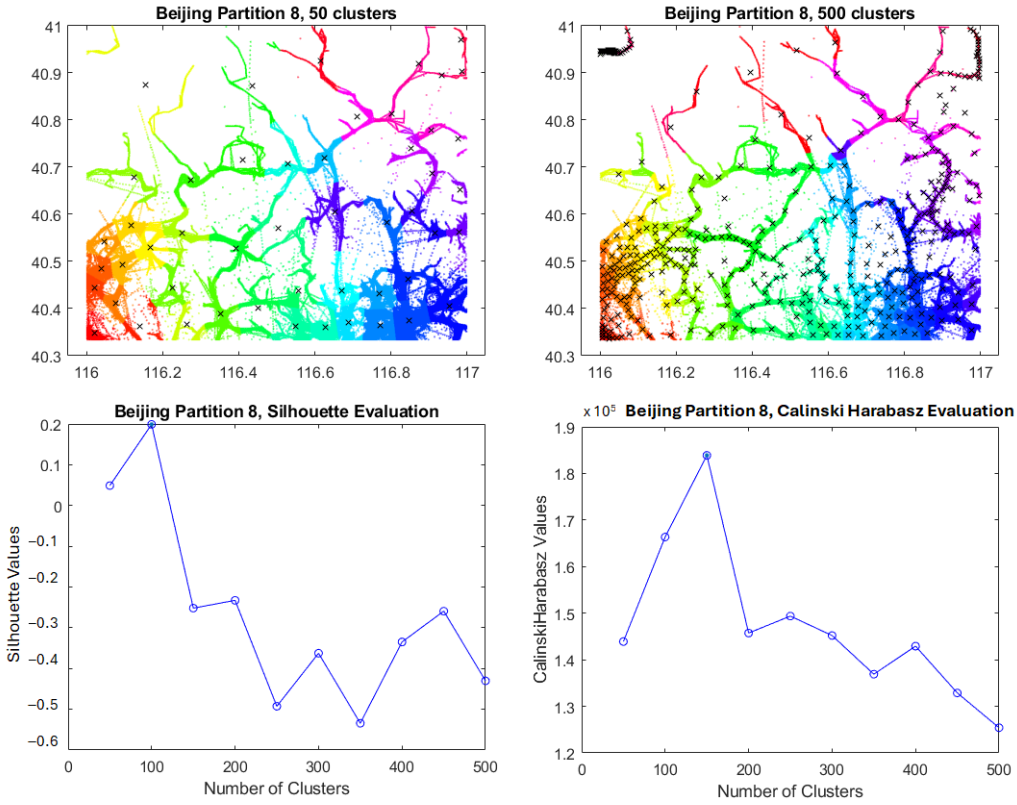

Figure 16. Using a secondary *k*-means partitioning with $k = 50$ (**upper left**, first row) and $k = 500$ (**upper right**, first row) on an initialy *k*-means cluster. Colors represent which cluster a mobility trace belongs. The second row shows the silhouette and Calinski–Harabasz indices to determine the optimal number of clusters in this partition.

Based on the two used evaluation indices, we see that around 100–150 clusters were the optimal number of EV charging locations indicated by the "×" in the first row of Figure 16. As the number of EV charging stations increased, the indices values decreased. This was due to the fact that on less traveled roads, vehicles could accelerate much faster to attain higher speed and reach the next nearest EV charging station. We still reiterate that the noisy BJG dataset contributed to a low silhouette value, and ideally, silhouette values should be positive and almost equal to one.

In summary, our methods provided possible EV charging locations for future EV users based on unsupervised learning clustering techniques, specifically, the CFS + *k*-means combination. The number of installed EV chargers was easily changed by varying the *k* value when *k*-means was re-run and evaluated by the validation indices. Ideally, more chargers are needed to remove driving anxiety when compared to internal combustion engine vehicles. This work also presented the importance of removing mobility traces' outliers, as shown in the BJG dataset. The road networks of BJG were clearly visualized and interpreted because of the initial method of determining the spatiotemporal stable network characteristics. However, the silhouette index revealed that there were still too many outliers that were not removed.

7. Conclusions

In this study, we employed various clustering techniques to determine possible EV charging stations based on empirical vehicular mobility traces traversing an urban road network. To achieve this, we first transformed the spatiotemporal dynamic vehicular environment into a single snapshot characterized by its spatiotemporal stable vehicular capacity characteristics. Given the allowable number of EV chargers to be deployed in an area, we employed clustering by fast search and find of density peaks (CFS) and secondary *k*-means clustering to obtain EV charging locations. Our work also showed how Density-based Spatial Clustering of Applications with Noise (DBSCAN) was unable to further partition a *k*-means cluster; therefore, we resorted to a secondary implementation of *k*-means clustering.

Our validation indices showed that inter-cluster distances fit a log-normal distribution with decreasing EV charger separations as more traffic zones are identified in an urban city. We note, however, that the number of EV chargers and locations can be further optimized by adding constraints such as type of chargers, area energy capacity, or target profit. This optimization study will further be studied in the future.

In our work, DBSCAN revealed the road networks of the mobility datasets, and we just differentiated highly traveled ones from roads less traveled through a color map representation. In the next phase of this research, we will dwell more on how we can better utilize the DBSCAN output from our datasets, especially in creating traffic zones/areas. We would like to devise a systematic approach on how to obtain the DBSCAN parameters from a given dataset and ensure that there are good separations between clusters having the same network characteristics. We will target especially those results that include locations with abundant traces.

Author Contributions: Conceptualization, manuscript writing, dataset gathering, analysis, and supervision, Elmer Magsino; dataset preprocessing, clustering simulation, Francis Miguel M. Espiritu; dataset preprocessing, spatiotemporal stable network characteristics simulation, Kerwin D. Go. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding author.

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

References

1. Irle, R. Global ev Sales for 2021. Volume 2. 2022. Available online: https://ev-volumes.com/news/ev/global-ev-sales-for-2021/#:~:text=Global (accessed on 17 October 2024).
2. Rainieri, G.; Buizza, C.; Ghilardi, A. The psychological, human factors and socio-technical contribution: A systematic review towards range anxiety of battery electric vehicles' drivers. *Transp. Res. Part F Traffic Psychol. Behav.* **2023**, *99*, 52–70. [CrossRef]
3. Verma, S.; Mishra, S.; Gaur, A.; Chowdhury, S.; Mohapatra, S.; Dwivedi, G.; Verma, P. A comprehensive review on energy storage in hybrid electric vehicle. *J. Traffic Transp. Eng. (Engl. Ed.)* **2021**, *8*, 621–637. [CrossRef]
4. Narasipuram, R.P.; Mopidevi, S. A technological overview & design considerations for developing electric vehicle charging stations. *J. Energy Storage* **2021**, *43*, 103225.
5. Shahriar, S.; Al-Ali, A.R.; Osman, A.H.; Dhou, S.; Nijim, M. Machine learning approaches for EV charging behavior: A review. *IEEE Access* **2020**, *8*, 168980–168993. [CrossRef]
6. Yang, X.; Zhuge, C.; Shao, C.; Huang, Y.; Tang, J.H.C.G.; Sun, M.; Wang, P.; Wang, S. Characterizing mobility patterns of private electric vehicle users with trajectory data. *Appl. Energy* **2022**, *321*, 119417. [CrossRef]
7. Brancaccio, G.; Deflorio, F.P. Extracting travel patterns from floating car data to identify electric mobility needs: A case study in a metropolitan area. *Int. J. Sustain. Transp.* **2023**, *17*, 181–197. [CrossRef]
8. Cai, Y.; Zhang, J.; Gu, Q.; Wang, C. An Analytical Framework for Assessing Equity of Access to Public Electric Vehicle Charging Stations: The Case of Shanghai. *Sustainability* **2024**, *16*, 6196. [CrossRef]
9. Li, M.; Jia, Y.; Shen, Z.; He, F. Improving the electrification rate of the vehicle miles traveled in Beijing: A data-driven approach. *Transp. Res. Part A Policy Pract.* **2017**, *97*, 106–120. [CrossRef]
10. Magsino, E.R.; Abello, A.J.; Lalusin, J.M. Taxi hotspots identification through origin and destination analysis of taxi trips using k-means clustering and h-indexing. *J. Phys. Conf. Ser.* **2021**, *1997*, 012006. [CrossRef]

11. Arias, M.B.; Kim, M.; Bae, S. Prediction of electric vehicle charging-power demand in realistic urban traffic networks. *Appl. Energy* **2017**, *195*, 738–753. [CrossRef]
12. Zhang, H.; Nie, G.; Chen, J.; Zhang, J.; Yang, G. Uncertain motion tracking combined markov chain monte carlo and correlation filters. *IEEE Access* **2019**, *7*, 167076–167088. [CrossRef]
13. Pan, M.; Li, W.; Wang, C. Modeling Electric Vehicle Charging Load Using Origin-Destination Data. In Proceedings of the International Conference on Transportation and Development 2024, Atlanta, GA, USA, 15–18 June 2024 ; pp. 265–275.
14. Tan, C.M.S.; Magsino, E.R. Demand-Based Deployment of Electric Vehicle Charging Stations Employing Empirical Mobility Dataset. In *Pervasive Computing and Social Networking: Proceedings of ICPCSN 2022*; Springer: Singapore, 2022; pp. 285–294.
15. Bae, S.; Kwasinski, A. Spatial and temporal model of electric vehicle charging demand. *IEEE Trans. Smart Grid* **2011**, *3*, 394–403. [CrossRef]
16. Lopez, N.S.; Allana, A.; Biona, J.B.M. Modeling electric vehicle charging demand with the effect of increasing evses: A discrete event simulation-based model. *Energies* **2021**, *14*, 3734. [CrossRef]
17. Paevere, P.; Higgins, A.; Ren, Z.; Horn, M.; Grozev, G.; McNamara, C. Spatio-temporal modelling of electric vehicle charging demand and impacts on peak household electrical load. *Sustain. Sci.* **2014**, *9*, 61–76. [CrossRef]
18. Kchaou-Boujelben, M. Charging station location problem: A comprehensive review on models and solution approaches. *Transp. Res. Part C Emerg. Technol.* **2021**, *132*, 103376. [CrossRef]
19. Hodgson, M.J. A flow-capturing location-allocation model. *Geogr. Anal.* **1990**, *22*, 270–279. [CrossRef]
20. Bayram, I.S.; Zafar, U.; Bayhan, S. Could petrol stations play a key role in transportation electrification? A gis-based coverage maximization of fast ev chargers in urban environment. *IEEE Access* **2022**, *10*, 17318–17329. [CrossRef]
21. Ullah, I.; Liu, K.; Layeb, S.B.; Severino, A.; Jamal, A. Optimal Deployment of Electric Vehicles' Fast-Charging Stations. *J. Adv. Transp.* **2023**, *2023*, 6103796. [CrossRef]
22. Zhang, T.Y.; Yang, Y.; Zhu, Y.T.; Yao, E.J.; Wu, K.Q. Deploying public charging stations for battery electric vehicles on the expressway network based on dynamic charging demand. *IEEE Trans. Transp. Electrif.* **2022**, *8*, 2531–2548. [CrossRef]
23. Yi, Z.; Liu, X.C.; Wei, R. Electric vehicle demand estimation and charging station allocation using urban informatics. *Transp. Res. Part D Transp. Environ.* **2022**, *106*, 103264. [CrossRef]
24. Lin, H.; Bian, C.; Wang, Y.; Li, H.; Sun, Q.; Wallin, F. Optimal planning of intra-city public charging stations. *Energy* **2022**, *238*, 121948. [CrossRef]
25. Wang, H.; Meng, Q.; Xiao, L. Electric-vehicle charging facility deployment models for dense-city residential carparks considering demand uncertainty and grid dynamics. *Transp. Res. Part C Emerg. Technol.* **2024**, 104579. [CrossRef]
26. Zeb, M.Z.; Imran, K.; Khattak, A.; Janjua, A.K.; Pal, A.; Nadeem, M.; Zhang, J.; Khan, S. Optimal placement of electric vehicle charging stations in the active distribution network. *IEEE Access* **2020**, *8*, 68124–68134. [CrossRef]
27. Archana, A.; Rajeev, T. EV charging station allocation in a distribution network based on power quality. In Proceedings of the 2021 IEEE Power and Energy Conference at Illinois (PECI), Urbana, IL, USA, 1–2 April 2021; pp. 1–6.
28. Konstantinidis, G.; Karapidakis, E.; Paspatis, A. Mitigating the impact of an official PEV charger deployment plan on an urban grid. *Energies* **2022**, *15*, 1321. [CrossRef]
29. Hosseini, M.; MirHassani, S.A. A heuristic algorithm for optimal location of flow-refueling capacitated stations. *Int. Trans. Oper. Res.* **2017**, *24*, 1377–1403. [CrossRef]
30. Lee, J.B. Review on environmentally-friendly vehicle charging station location and size problems. *KSCE J. Civ. Eng.* **2022**, *26*, 3741–3751. [CrossRef]
31. Zhou, M.; Long, P.; Kong, N.; Zhao, L.; Jia, F.; Campy, K.S. Characterizing the motivational mechanism behind taxi driver's adoption of electric vehicles for living: Insights from China. *Transp. Res. Part A Policy Pract.* **2021**, *144*, 134–152. [CrossRef]
32. Pan, L.; Yao, E.; Yang, Y.; Zhang, R. A location model for electric vehicle (EV) public charging stations based on drivers' existing activities. *Sustain. Cities Soc.* **2020**, *59*, 102192. [CrossRef]
33. Karolemeas, C.; Tsigdinos, S.; Tzouras, P.G.; Nikitas, A.; Bakogiannis, E. Determining electric vehicle charging station location suitability: A qualitative study of greek stakeholders employing thematic analysis and analytical hierarchy process. *Sustainability* **2021**, *13*, 2298. [CrossRef]
34. Shabbar, R.; Kasasbeh, A.; Ahmed, M.M. Charging station allocation for electric vehicle network using stochastic modeling and grey wolf optimization. *Sustainability* **2021**, *13*, 3314. [CrossRef]
35. Jegham, M.A.; Layeb, S.B. Locating and sizing electric vehicles charging stations in new markets. *Int. J. Appl. Decis. Sci.* **2020**, *13*, 153–182. [CrossRef]
36. Anjos, M.F.; Gendron, B.; Joyce-Moniz, M. Increasing electric vehicle adoption through the optimal deployment of fast-charging stations for local and long-distance travel. *Eur. J. Oper. Res.* **2020**, *285*, 263–278. [CrossRef]
37. Moreno-Carbonell, S.; Sánchez-Úbeda, E.F.; Muñoz, A. Rethinking weather station selection for electric load forecasting using genetic algorithms. *Int. J. Forecast.* **2020**, *36*, 695–712. [CrossRef]
38. Obias, K.C.U.; Magsino, E.R. Extracting vehicular mobility dynamics from taxi fleet trajectories. In Proceedings of the 2021 IEEE 9th Region 10 Humanitarian Technology Conference (R10-HTC), Bangalore, India, 30 September–2 October 2021; pp. 1–6.
39. Ho, I.W.H.; Chau, S.C.K.; Magsino, E.R.; Jia, K. Efficient 3D road map data exchange for intelligent vehicles in vehicular fog networks. *IEEE Trans. Veh. Technol.* **2019**, *69*, 3151–3165. [CrossRef]

40. Magsino, E.R. Investigating the operational dynamics of different transportation modes in vehicular networks. In Proceedings of the 2021 International Conference on Information Technology (ICIT), Amman, Jordan, 14–15 July 2021; pp. 246–250.
41. Magsino, E.R.; Ho, I.W.H. An enhanced information sharing roadside unit allocation scheme for vehicular networks. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 15462–15475. [CrossRef]
42. Schubert, E.; Sander, J.; Ester, M.; Kriegel, H.P.; Xu, X. DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. *ACM Trans. Database Syst. (TODS)* **2017**, *42*, 1–21. [CrossRef]
43. Rodriguez, A.; Laio, A. Clustering by fast search and find of density peaks. *Science* **2014**, *344*, 1492–1496. [CrossRef]
44. Dixit, A. Fast kmeans Algorithm Code. MATLAB Central File Exchange. Available online: https://www.mathworks.com/matlabcentral/fileexchange/44598-fast-kmeans-algorithm-code (accessed on 17 October 2024).
45. Shutaywi, M.; Kachouie, N.N. Silhouette analysis for performance evaluation in machine learning with applications to clustering. *Entropy* **2021**, *23*, 759. [CrossRef]
46. Wang, X.; Xu, Y. An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index. *IOP Conf. Ser. Mater. Sci. Eng.* **2019**, *569*, 052024. [CrossRef]