

Article

A Study of Spatiotemporal Distribution of Mobility-On-Demand in Generating Pick-Up/Drop-Offs Location Placement

Ryan K. Gunawan¹ and Susilawati^{1,2,*}

¹ Department of Civil Engineering, School of Engineering, Monash University Malaysia, Subang Jaya 46500, Malaysia; Ryan.Gunawan@monash.edu

² Advanced Engineering Platform, School of Engineering, Monash University Malaysia, Subang Jaya 46500, Malaysia

* Correspondence: susilawati@monash.edu; Tel.: +603-5514-4490

Abstract: The location placement of pick-up/drop-offs of ride hailing usually only considers spatial distribution within a certain area. Previous studies often mapped out the potential hotspots for pick-up/drop-offs, benefitting the ride-hailing company and not considering the passengers. Therefore, in this study, we incorporated the spatiotemporal distribution of mobility-on-demand on generating pick-up/drop-off location placement using a genetic algorithm considering the walking distance and minimum demand data served within the radius. The data collected are analyzed through the clustering method, and the resulting cluster centers are fed into the location optimization algorithm. The genetic algorithm is used to optimize the location placement with the consideration of the user's walking distance and demand. The algorithm is able to find an appropriate placement and determine reliable pick-up/drop-off stations within the study area based on observed spatiotemporal demand despite the difference in demand distribution through different time periods.

Keywords: location placement; pick-up/drop-off hotspots; genetic algorithm



Citation: Gunawan, R.K.; Susilawati. A Study of Spatiotemporal Distribution of Mobility-On-Demand in Generating Pick-Up/Drop-Offs Location Placement. *Smart Cities* 2021, 4, 746–765. <https://doi.org/10.3390/smartcities4020038>

Academic Editor: Frank Witlox

Received: 31 March 2021

Accepted: 12 May 2021

Published: 17 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The traditional method of transport data often used manual traffic survey to identify and analyze traffic patterns. This method suffers from a lack of reliability and accessibility as the amount of data collected is often limited and inadequate. However, the rapid technological development in navigation and communication systems such as GPS, mobile phones and Wi-Fi, etc., has allowed the convenience and availability of detecting large-scale individual travel data and subsequently recording vehicle trajectories [1]. Various data such as vehicle trajectories, origin–destination, pick-up, drop-off locations, trip length, and duration can be retrieved and analyzed.

The convenience of location-aware technologies was previously implemented widely in hired mobility such as taxi services. Mobility services navigation data proved to be a significant source in exploring various traffic research subjects such as travel patterns and characteristics and traffic models. Yuan et al. [2] utilized historical taxi trajectories and used a variance–entropy clustering approach to estimate travel time between locations. Alexander et al. [3] investigated the effects of ridesharing adoption the network-wide traffic congestion by analyzing origin–destination trips from both auto and non-auto travelers that were extracted from mobile phone data. The results suggested that a significant reduction in total vehicle and total travel time were obtained when 50% of the drivers adopted the ridesharing scheme. Qu et al. [4] developed an adaptive routing method using a probabilistic network model taking into consideration between supply and demand. Using the availability of large data, some researchers analyzed the characteristics of urban mobility to understand overall traffic mobility [5] better and subsequently provide insights on urban planning [6]. Some research focuses specifically on defining traffic congestion patterns [7,8].

Furthermore, the advancement of technologies has brought out the flexibility of mobility-on-demand to customer's fingertips through the transportation network companies (TNCs), such as Uber and Grab, by conveniently requesting trips through a personal mobile phone instead of traditional taxi-hailing from the sidewalk. These trips have become the bulk of current transportation services, and records of these trips are stored in the TNC database. Hence, analyzing these data rather than the conventional taxi trips have become essential in visualizing and analyzing traffic demand.

Meanwhile, several kinds of research have also incorporated taxi data in determining pick-up and drop-off (PUDO) zones to integrate taxi operation management better and incorporate the user's travel pattern analysis. Different clustering methods are often used in determining such hotspots considering the spatiotemporal distribution of taxi traffic data [9–11]. However, all of these studies do not consider possible station placement within the area and their possible advantages. By incorporating station placement, the operators will be able to identify prominent PUDO spots and possibly decrease PUDO occurrences in random places. In place of this, optimizing PUDO stations can potentially be deployed to minimize random on-road PUDO events and decrease traffic congestion [12]. By analyzing current PUDO demand as well as other defining factors such as minimum capacity and walking distance, PUDO stations can be optimized and implemented strategically in improving overall traffic condition in the system.

Therefore, the objective of this study is to identify potential PUDO stations based on the spatial–temporal distribution of traffic patterns in Jakarta, Indonesia using the ride hailing trajectories data provided by Grab-Posisi [13]. Grab-Posisi is Southeast Asia's first GPS trajectory dataset. The data were collected in April 2019 with a 1 s sampling rate that consisted of ride hailing's GPS trajectory data from both developed countries (Singapore) and developing countries (Jakarta, Indonesia).

The PUDO location optimization algorithm was developed using a genetic algorithm and the optimum walking distance and capacity were set as constraints. Analysis of spatial–temporal distribution was carried out to observe the PUDO distribution within the study area. Spatial filtering of the provided area was conducted based on the travel demand distribution. Based on the spatio-temporal travel demand, three PUDO location optimization algorithms were developed; morning peak, off-peak, and evening peak. Spatial clustering was conducted for the initial pre-processing of each PUDO location optimization and the resulting analysis was fed into the genetic algorithm where optimal PUDO locations were then determined by incorporating certain placement requirements. Thus, this paper contributes to the following areas:

- The proposed location placement using a genetic algorithm was able to find a prominent location for pick-up/drop-offs for the convenience of both the ride-hailing company and potential riders.
- The proposed algorithm used a refined fitness function that consists of the combination of applied constraints of distance and demand coverage by modifying the potential interaction of the gravity model.
- The algorithm was able to adapt to find an appropriate placement based on observed demand that assisted in determining the optimum pick-up and drop-off (PUDO) zones to improve taxi and ride hailing operation management.

The remainder of this article is structured as follows. In Section 2, reviews of multiple studies are examined and described for the motivation of this study. In Section 3, details of research methodology are presented. The results and conclusion are presented in Sections 4 and 5, respectively.

2. Literature Review

2.1. PUDO Location

Several approaches that consider the different types of features and constraints to improve ridesharing and ride-hailing operations have been introduced. These approaches focused on improving the routing strategy given various routing, time, and capacity

constraints [14,15] in order to optimize operational objectives and quality-related objectives [16]. Operational objectives aim to optimize system-wide operating costs, including minimizing vehicle kilometer travel and travel time, minimizing the fleet of the vehicle while maximizing the number of services, and others. Quality-related objectives are about enhancing the quality of provided services that include minimizing waiting time [16].

It is quite clear that the existing studies on ridesharing focused on the routing strategy, with little attention paid to optimize the PUDO points that are responsive to the spatial and temporal demand distribution [12]. The PUDO optimization is required to avoid unnecessary detours, particularly when the trip is shared with other riders. While in the traditional door-to-door scheme, the vehicle is often required to pick up the passengers at certain locations for which the vehicle needs to pass through congested and narrow streets that deviate it from the shortest and fastest paths [17]. It is quite often that the detours to pick up other riders increase the travel time and cost. As such, there is a critical need to optimize the PUDO location for which all possible pick-up or drop-off points within a walkable walking distance were aggregated into one location. Hence, the riders are required to walk a very short distance to reach the optimized PUDO location.

This optimization method considers both the rider and vehicle assignments that lead to a complex mathematical task involving the NP-hard problem [18]. As such, although numerous implementation strategies in determining pick-up and drop-off (PUDO) zones to improve taxi, ridesharing and ride hailing operation management, PUDO station location placement is not explored extensively [19]. Many studies tend to propose a pre-determined set of “meeting points” to which all passengers have to walk [15,17,20–22]. This is because optimizing placement constraints (i.e., minimum capacity, walking radius, the distance towards adjacent facilities, etc.) stretches the complexity of the PUDO location optimization [15,21]. Czioska, P. et al. [19] introduced a GIS-based method to identify potential meeting points and proposed assessment criteria to rate the suggested meeting points. The assessment was carried out by combining the rider’s stated preferences on meeting point characteristics collected using an online questionnaire and evaluating available facilities around the suggested meeting points, including parking places, seating, shelter, and light. Zheng, Y. et al. [23] proposed a meeting point strategy that required riders to suggest their pick up and dropped off (PUDO) points to improve flex-route transit service. These points can be at their reserved PUDO stops or the meeting points. The benefit of introducing this strategy was evaluated using a memetic algorithm. The results showed that the willingness to walk to meeting points reduced the rejection rate and benefitted both riders and operators. To protect the rider’s location privacy, Aïvodji, U.M. et al. [24] integrated the privacy-preserving algorithm with a multimodal shortest path algorithm. The proposed algorithm was able to produce solutions that mutually benefitted the drivers and riders and able to protect the riders’ location information. Qian, X. et al. [12] designed an incentive strategy for taxi group rides using a heuristic algorithm that solved a dynamic ride-matching problem that significantly increases the total saved travel mileage considering the penalty values for walking extra miles to the meeting points and waiting additional minutes.

Several studies introduced methods on identifying hotspots using a modified clustering algorithm. Lee et al. [7] used k-mean clustering to identify pick-up hotspots in the Jeju city area based on the spatiotemporal distribution of taxi mobility history. The author clustered the taxi pick-up pattern based on spatial distribution, then a time-dependent pick-up pattern was conducted to incorporate temporal distribution. The study then was able to suggest a recommended pick-up area for different time periods. Zhang et al. [11] used pre-processing hierarchical k-means clustering based on historical spatiotemporal pick-up demand and categorized them into different time periods and areas. The results were then ranked based on the pick-up demand within each cluster and recommended to the taxi driver. The study provided high precision in taxi pick-up suggestions with low computational time and recommended pick-up points within a couple of seconds due to the advantage of data pre-processing. In the study, the authors only considered the spatial and temporal distribution of PUDO hotspots and were not able to pinpoint the

exact location for PUDO stations to be developed. One could argue that PUDO stations could be developed inside the resulting hotspots. However, placement constraints needed to be considered when constructing a new structure, such as a number of demand data actually served within the radius of PUDO stations. Another study uses DBSCAN to eliminate noises and categorize clusters of arbitrary shapes and analyze both pick-up and drop-off demands [9]. The authors only consider the distance between demand points to create the hotspots. It is found that the distribution between pick-up and drop-off demand provides similar cluster center results, which proves the relationship between people's travel behavior and land use characteristics. The authors however, only considered the distribution of PUDO within the study area over time period. They did not provide any other constraints in the placement of PUDO spots.

2.2. Location Optimization Using Genetic Algorithm

Estimating the appropriate location through a series of constraints is inherently a facility location problem. The simple definition of a facility location problem is to find locations for a new facility such that the transportation cost is minimized between the facility and customers [25]. Other than transportation cost, the problem can grow in complexity considering other factors such as distance between facilities, demands, capacitated or incapacitated, etc. Due to the complexities of the problem, the heuristics method often used to provide a good enough solution while considering various variables and constraints. One of the examples of location optimization is the electric charging station placement. Mehar and Senouci [26] pre-processed the traffic in the study area and computed hierarchical clustering to group the electric car usage and subsequently calculate the energy demand. They considered area traffic density, construction, transportation costs, and charging capacity in their algorithm.

In addition to that, location optimization is inherently an NP-hard problem, and as such, a heuristics approach is preferable [27]. Heuristics or metaheuristic method provides a reasonable solution within the search space through limited runtimes, such as simulated annealing (SA), swarm optimization (SO), genetic algorithm, and others. A genetic algorithm is a stochastic search process through combinatorial evolutionary-inspired techniques that enables one to model appropriate constraints. Some studies have used the genetic algorithm in defining location placement, and currently, with the rise of the electric vehicles, the optimum placement of charging stations is being studied [26,28]. Mehar and Senouci [26] proposed an optimized genetic algorithm by introducing a permutation operator to prevent premature convergence. As a result, the algorithm was able to suggest location placement efficiently in terms of time and optimality. Chen et al. [28] incorporated density function to improve the genetic algorithm by considering cost, capacity and coverage. The resulting algorithm was able to provide an optimal number of stations needed and their locations considering user's demands and convenience. Both of these studies are specifically designed for the location optimization of the electric vehicle charging station. This study differs from those mentioned due to the presence of the pre-processing of data through spatial-temporal filtering and clustering as well as the introduction of parameters specific in determining the construction of PUDO stations; in this case, walking distance and minimum demand data served within service radius. Furthermore, the potential interaction gravity model is introduced as a novel approach in quantifying the relationship between the PUDO station and demand points.

2.3. First-Mile and Last-Mile

Current public transportation issues such as first-mile/last-mile also provide motivation for the need in location placement of PUDO stations [29]. Multiple concepts have been researched in addressing FMLM issues, including the use of personal rapid transit (PRT) [30] and car-sharing programs [31]. Although there seem to be merits in implementing those methods, due to the high investment cost for PRT [32] and other limitations, there is still a need for a better alternative in solving this issue. Another study proposed the use of bicycling to tackle FMLM issues and provides a method to assist

the bike/pedestrian network planning for better transit point-to-point mobility [33]. This method only works when the city infrastructure allows for flexibility in additional network planning for bikes/pedestrians. Due to limitations found, studies on FMLM of public transport are still occurring, and this study provides the potential application of PUDO stations where they can provide seamless passenger transfers to public transport, hence improving FMLM connectivity.

3. Research Methodology

This research aims to determine the optimal PUDO station's location, considering distance and capacity constraints, based on traffic demand's spatial-temporal pattern. The optimization of PUDO stations allows the minimization of customer's walking distance and possibly reducing the number of PUDO occurrences that may lead to longer travel time. By minimizing walking distance, user convenience is achieved, and the optimum placement of a PUDO station could reduce the number of random PUDOs that will lead to longer travel time.

A genetic algorithm was used to incorporate constraints to the placement problem, allowing greater flexibility in modifying the model constraints and producing optimum solutions by carrying out a global stochastic search process. This study incorporated a modified potential interaction gravity model to quantify the interaction between PUDO stations and demand. Firstly, travel data were extracted and analyzed based on spatial and temporal filtration using the combinatorial software application of RStudio and QGIS [34,35]. The data used in this study were provided by Grab and the travel data location provided was in Jakarta, Indonesia. The resulting data created the PUDO trip data within the study area as well as clustering centers for each different trip data. The clustering results were imported to the genetic algorithm as initial input, and optimization was carried out by incorporating walking distance requirements and minimum capacities through a potential interaction gravity model between PUDO stations and assigned demands.

In this study, capacity refers to the demand data served by PUDO stations within its restricted walking distance radius. Unlike traditional traffic demand data collection that were collected by conducting an origin-destination traffic survey, travel demand data analysis is conducted based on quantitative spatio-temporal ride hailing data provided by Grab. This allows for greater precision of the origin-destination location and higher accuracy of traffic demand analysis. Data extraction and analysis were conducted with the use of statistical computing and graphics application, R software and QGIS, an open-source geographic information system for analysis and viewing. Then, the PUDO locations' algorithm was developed using a genetic algorithm coded in Python coding language.

3.1. Data Extraction

The ride hailing-based trips data generated from Grab were used in this study [13]. The data consist of 2 weeks (08/04/2019–22/04/2019) of ride hailing-based trips data within Greater Jakarta and the surrounding area. This corresponds to more than 50,000 trips within a 650 km² area. Private and personal information is omitted from the data to protect ride hailing user consumer rights. Only necessary trip information was recorded into the file such as trip ID, vehicle mode, trip device used, timestamp, latitude, longitude, speed, bearing and accuracy. This information was recorded in the Parquet file and imported to RStudio for analysis. The data consist of the point of trajectory of recorded trips. These trips were separated based on ID, and PUDO details were obtained based on minimum and maximum trajectory time recorded for each trip ID. Then, the data were further filtered based on the determined trip period (weekdays or weekends and morning peak, off-peak or evening peak). Finally, spatial filtration was performed in QGIS based on the study area.

Figure 1 shows the temporal trip distribution throughout the day. To protect the data confidentiality, only the relative frequency of trip demand is presented in this paper. As shown in Figure 1, the temporal trip demand is fairly distributed within the conventional working time period. It is at its minimum between 12 A.M. and 6 P.M.; however, the trend continues upward from then on until the peak at between 3 P.M. to 6 P.M.

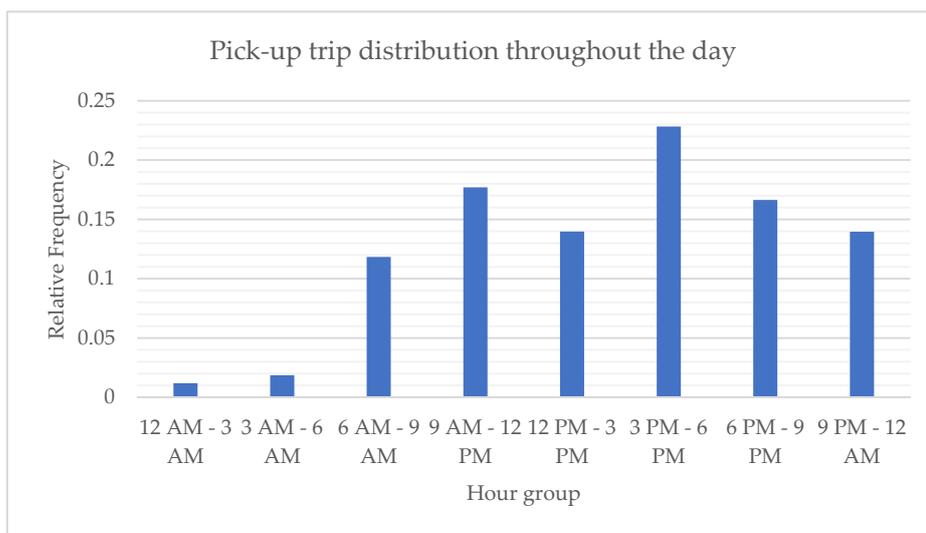


Figure 1. Relative frequency of pick-up trip throughout the day.

3.2. Chosen Study Area and Data

To account for the worst case scenario when trip demand is high, two highest trips in weekdays and weekends were further used to evaluate the effect of temporal trip distribution on optimum PUDO locations. In this case, the chosen data for weekdays are 16/04/2019 and 18/04/2019. For simplification, these two chosen weekday data sets are named data set 1 and data set 2, respectively, while the weekends are 13/04/2019 and 20/04/2019, which are named data set 3 and data set 4, respectively. The temporal distribution can be seen in Figure 2.

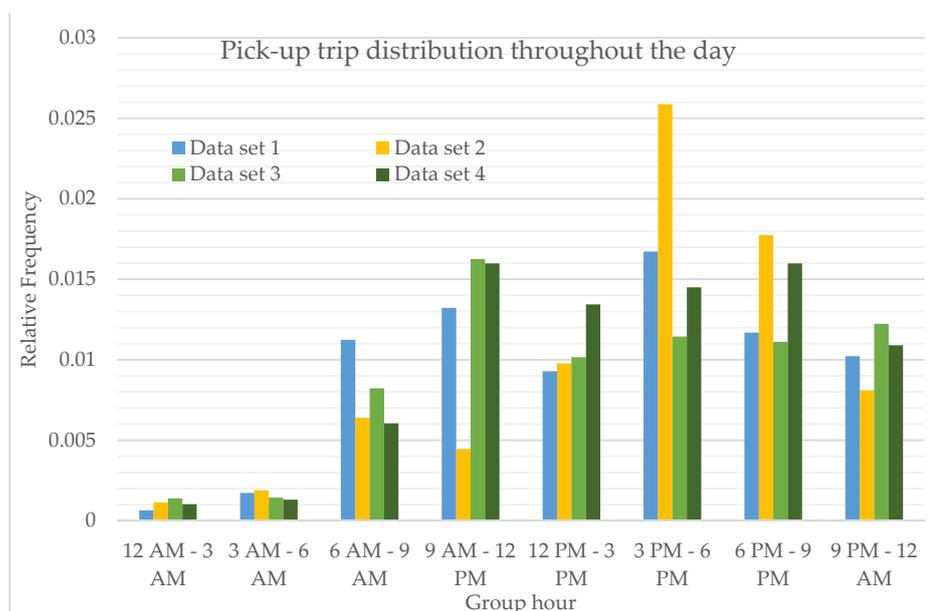


Figure 2. Relative frequency of pick-up trip distribution of data set 1, data set 2; pick-up data set 3 and data set 4.

The temporal distribution of the chosen dates are distinctively different from each other. For the weekdays, it can be seen the pick-up distribution on data set 1 is similar towards total hourly distribution shown in Figure 1, unlike the distribution shown on data set 2. The cause for this distinction on data set 2 may be due to the presence of a public holiday on 19-04-2019 (Good Friday). Since 19-04-2019 is a public holiday, some

people might decide to take a leave day on Thursday 18-04-2019 so they may enjoy a long weekend, causing low trip data recorded on the morning peak of 18-04-2019. High trip data recorded on the evening peak of data set 2 may be due to a similar reason of a public holiday resulting in people starting to go out for recreational activities. The pick-up distribution on weekends (data set 3 and data set 4) differs very much from each other as, similar to the explanation for the distinct temporal pick-up distribution on data set 2, 20-04-2019 is still part of the long weekend caused by the public holiday on 19-04-2019, susceptibly making people travel more in the afternoon for leisure activities.

Due to the high irregularity of the temporal pick-up distribution, the study area chosen for this study was located in the combined area of Central and South Jakarta with the time period selected based on previously selected weekday and weekend dates. A further breakdown of trip distribution was carried out based on the morning peak, off-peak and evening peak. The morning peak hours were selected to be from 6 A.M. to 9 A.M. The evening peak period was selected to be from 4 P.M. to 8 P.M., while the off-peak period was from 11 A.M. to 2 P.M. (based on the lunch period).

3.3. Genetic Algorithm

The genetic algorithm is a metaheuristic search algorithm inspired by biological evolution [36,37]. The algorithm candidates go through a series of evolutionary processes (selection, crossover, mutation) to converge the algorithm in finding a near-optimum solution. Another important aspect of the algorithm is the fitness value. Each candidate solution has a specific determinant value that represents their overall strength in the evolutionary algorithm. The identification of their overall strength in the population is based on a specific set of requirements determined previously during problem identification. Based on these parametric requirements, a suitable fitness function can be developed that appropriately assesses the candidate's capability to solve the problem into a single quantitative value, named fitness value. Choosing and developing appropriate fitness functions is critical, as a well-constructed fitness function may substantially increase the chance of finding a solution and reaching higher coverage.

3.3.1. Chromosome Representation

In this work, stations were encoded as coordinates (latitude, longitude) and taken as a gene within a chromosome. The length of a chromosome represented the maximum number of PUDO stations to be generated for the given study area.

3.3.2. Initial Chromosome and Population

The initial chromosome was set to be the result of different clustering methods applied to the study area. By applying this result, the algorithm was expected to converge faster to the optimum solution while also comparing each initial configuration. The initial representation of the trip demand can be visualized using the clustering technique; hence, spatial clustering techniques via k-means and k-medoid were used to represent the initial chromosome. The resulting clustering are shown in Figure 3 for the data set 3, while others are presented in Appendix A. A random initial chromosome was also created for comparison purposes.

3.3.3. Fitness Function

The fitness function consists of the combination of applied constraints of distance and coverage as well as a modified potential interaction gravity model. The potential interaction gravity model is one of three basic spatial interaction models that measure interaction between locations. In this particular model, the interaction of the proposed station needed to be measured against overall demand points within the coverage of that station. The accumulation of demand points assigned to the proposed station was the final fitness value of that station. The algorithm's objective was to locate the appropriate installation of PUDO stations that maximized the demand points in the study area while considering the walking distance limit and minimum demand point serve per station. Based on those considerations, assumptions were created and listed as follows:

1. A maximum walking distance of 500 m is established. The 500 m walking distance was suggested a comfortable walking distance in ridesharing studies. If the distance between the demand point and the nearest PUDO station is above the established limit, the demand point will not be serviced.
2. Minimum demand points serviced by a PUDO station is 5 points in order for the station to be constructed. The minimum of 5 demand point has been selected to ensure that the proposed method can pool and serve the given number of passengers while maintaining the minimum fleet size [38].
3. The distribution of a certain time period only allows a small number of stations to be developed. Taking into consideration the size of the study area as well as PUDO distributions, the assumption of maximum 25 stations constructed is implemented in this study.
4. Not all demand points need to be serviced. The goal is to maximize the demand points, while considering the above constraints.

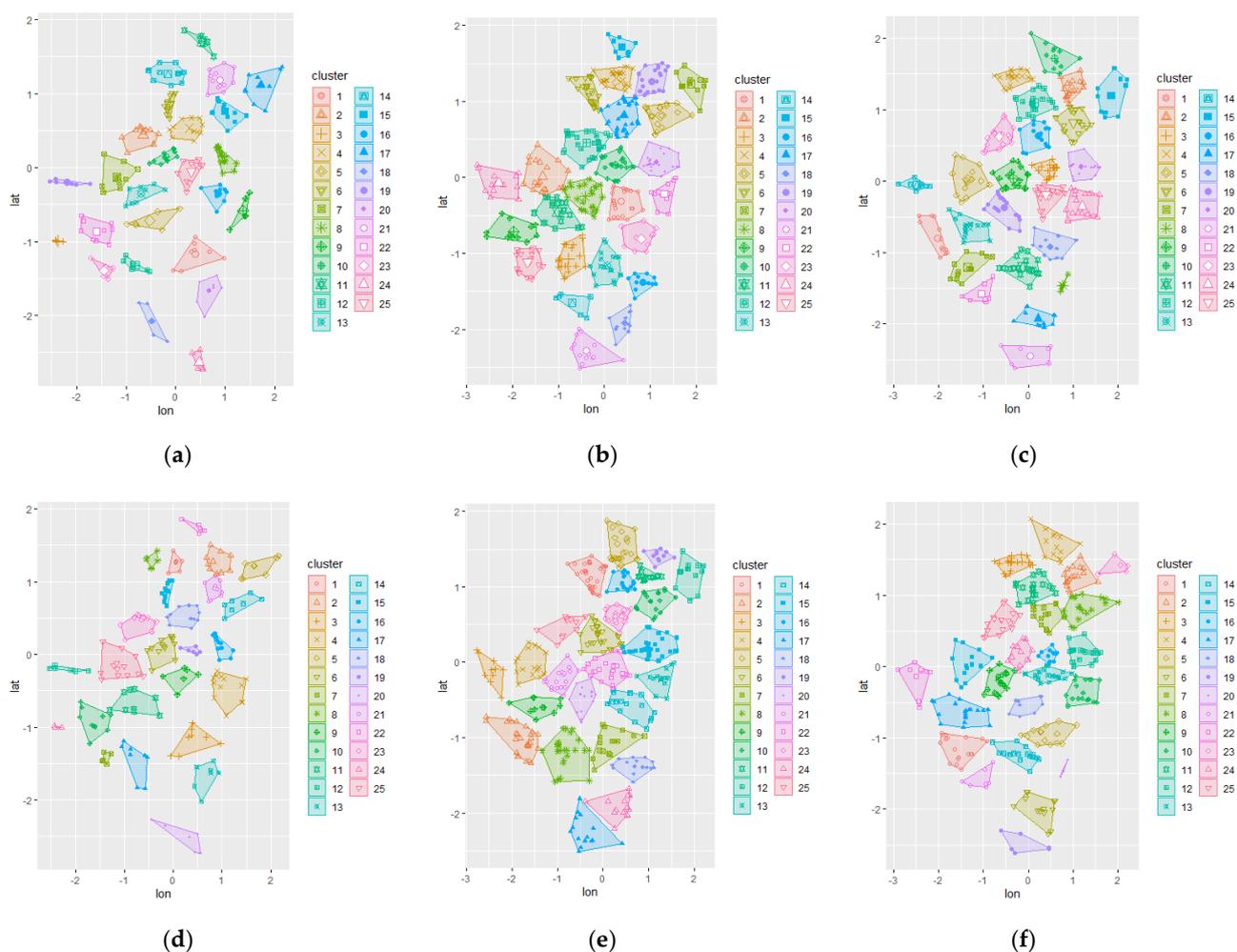


Figure 3. Clustering result for data set 3. (a) K-means cluster result during morning peak; (b) K-means cluster result during off-peak; (c) K-means cluster result during evening peak; (d) K-medoid cluster result during morning peak; (e) K-medoid cluster result during off-peak; (f) K-medoid cluster result during evening peak.

Therefore, the fitness function of a single station/gene can be determined using Equation (1) subjected to predetermined constraints mentioned previously. Table 1 shows the notations used in the mathematical formulation.

$$\max f_j = \sum_{i=1}^{i=m} \frac{W_{ij}}{d_{ij}^{\beta}}, x_j = 1 \tag{1}$$

Subject to

$$z_{ij} = \begin{cases} 1 & \text{if } d_{ij} \leq 500m \\ 0 & \text{otherwise} \end{cases}, \forall i \in I, j \in J \quad (2)$$

$$x_j = \begin{cases} 1 & \text{if } \sum_{i=1}^{i=m} z_{ij} \geq 5 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Once the evaluation of each station was computed, the fitness function of one chromosome was simply then the summation of the fitness function of the individual station/gene.

Table 1. Description of notations.

Notation	Description
I	a set of PUDO demand points, $\{1, \dots, m\}$.
J	a set of PUDO stations, $\{1, \dots, n\}$.
f_j	fitness function of station j .
W_{ij}	weight of demand point assigned demand point i to station j .
β	parameter of transport friction related to efficiency of transport between two locations. Assumption is 0.2 to decrease the influence of walking distance towards fitness value.
d_{ij}	distance of demand point i to the nearest station j , determined using haversine formula
x_j	binary variable for building station j .
z_{ij}	binary variable of assigning demand point i . to station j .

3.3.4. Selection Procedure

After evaluating the fitness value, the chromosomes inside the population were sorted in descending order according to the fitness value. The particular reason for this stage is to partition them into subpopulations: elite chromosomes and non-elite. The elite chromosomes were retained for the next generation, and the remaining chromosomes needed for the next generation were determined either through crossover or mutation procedure.

3.3.5. Crossover Procedure

A crossover operator was initiated between arbitrarily chosen elite and non-elite chromosomes, as determined previously, to produce the subsequent generation offspring. Figure 4 shows the elitist chromosome selection and the crossover procedure. Uniform crossover was chosen for this study as it generally produces better recombination compared to traditional one-point and two-point crossover [39]. When the crossover procedure occurred, each gene was evaluated and had a 50% chance of having crossover until the procedure reached the last gene and produced two offspring for the next generation.

3.3.6. Mutation Procedure

The mutation operator was initiated based on the probability of mutation rate. A high mutation rate of 0.3 was chosen to encourage stochastic search in the study space while preserving the algorithm's ability to converge to the optimum result. The selection of a chromosome was determined using tournament selection, where only the best chromosome prevailed and was selected for mutation. The tournament participant pool size was selected randomly based on the predetermined tournament size, which was 40% out of the current generation population. Once the selection chromosome was determined, and mutation occurred, each gene had a 50% chance of being mutated until the procedure reached the last gene and produced one offspring for the next generation.

3.3.7. Termination Condition

The algorithm was run for a series number of generations to ensure convergence of an optimum solution. The number of generations used in this study started from 0 (initial population) to 100 with the increment of 5 generations.

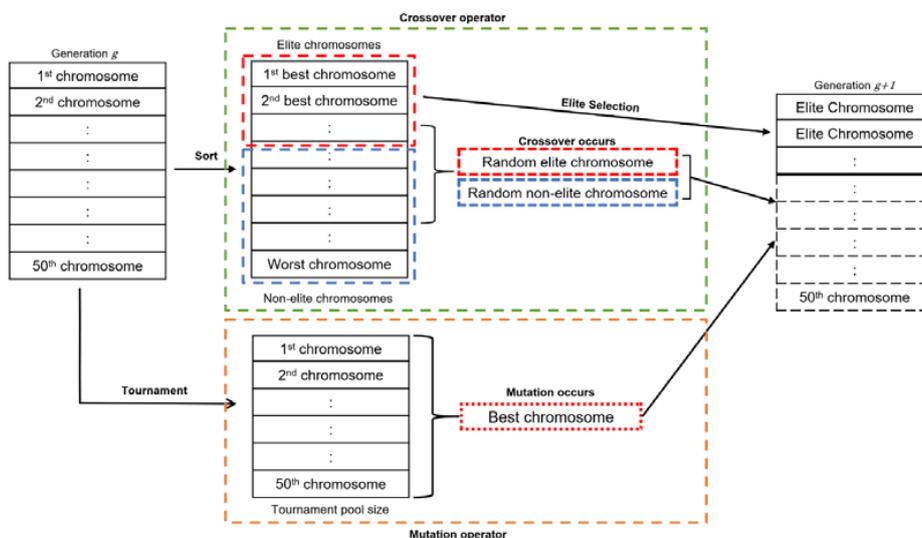


Figure 4. The evolutionary process of the genetic algorithm used in this study.

3.4. Scenario Development

The data collected from local TNCs were filtered based on the determined temporal distribution (weekdays/weekends and morning peak/off-peak/evening peak) using RStudio. The resulting data were transferred to QGIS for spatial filtering based on Central and South Jakarta’s spatial area. Then, the data were further analyzed by means of the clustering method. K-means and k-medoid clustering was used for all scenarios, resulting in 25 clusters for each scenario. These cluster centers were then fed to the genetic algorithm as initial chromosome representation (initial PUDO stations) for further optimization based on the potential interaction gravity model between stations and demand as well as incorporating the distance requirement and minimum capacities.

4. Results

Figure 5 shows the stations generated for different dates and different time periods in comparison with the total demand analyzed during those periods. A clear distinction can be seen in all the graphs, in that when the demand is low, the algorithm generates fewer stations to meet the demands effectively. Other factors may play into mainly how distributed the raw PUDO demand is in the study space. For all different initializations (k-means, k-medoid, and random), there seems to be not much difference in the number of generated stations as well. More distinction between all initialization is shown in their resulted fitness value.

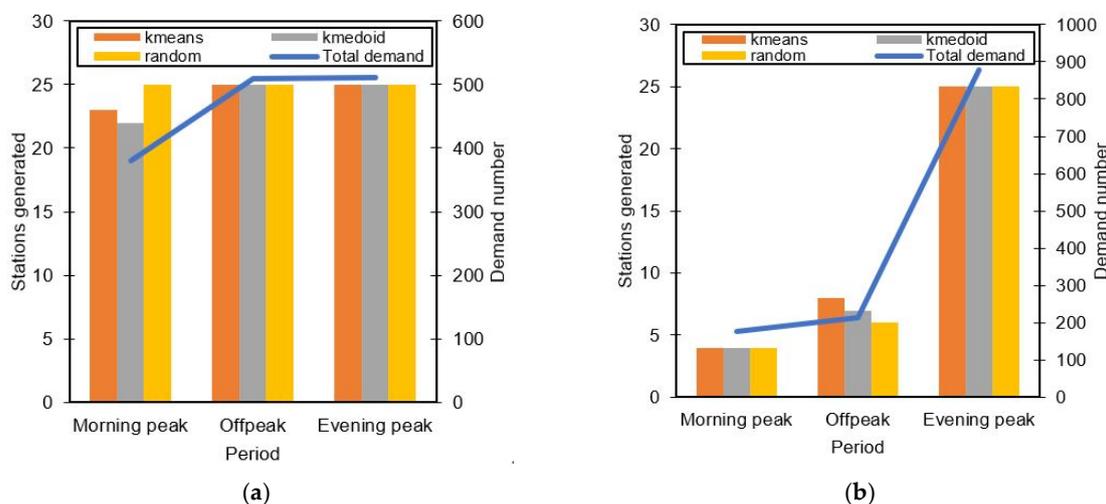


Figure 5. Cont.

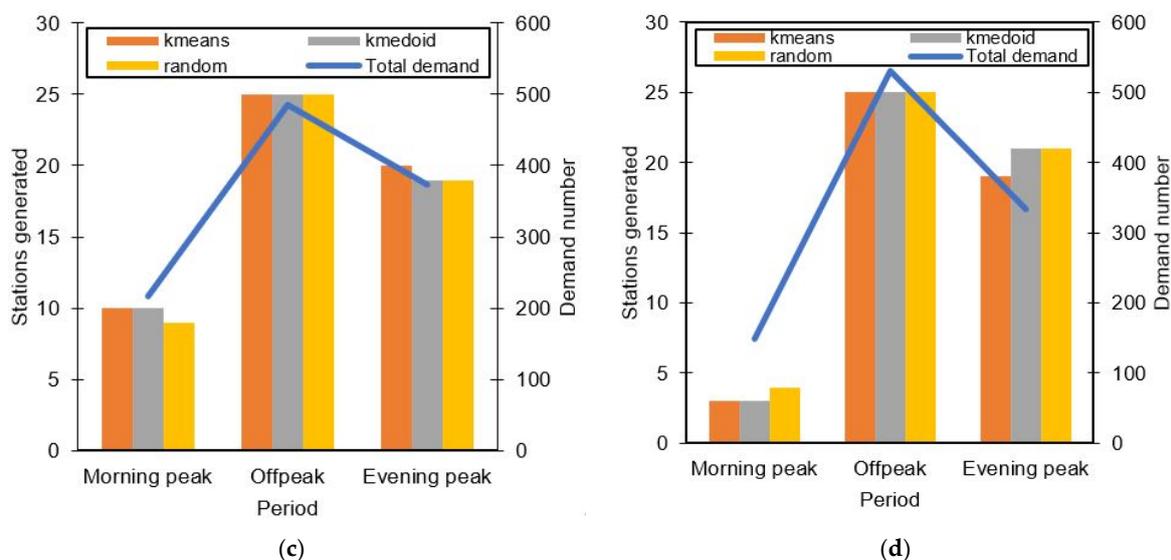


Figure 5. Comparison between stations generated with demand number. (a) Data set 1; (b) data set 2; (c) data set 3; (d) data set 4.

The number of PUDO stations generated for different initial chromosomes resulted into similar numbers. A randomized initial chromosome seems to be able to provide similar results compared to a k-means or k-medoids initial chromosome. The algorithm was able to adapt well in determining the number of stations based on the prerequisite requirement even though the initial chromosome was randomized thoroughly.

Figures 6–9 show the result of the different initial population’s fitness value for different demand periods. All three initial populations converge rapidly as generation increases, with k-medoid providing the best fitness for most of the period. This result is contributed from the nature of k-medoid, where it partitioned the data using the actual demand data and allows greater interpretability of the clusters’ centers compared to k-means. Although it is worth noting that the difference in final fitness value for all cases, when compared only the initialization, does not differ too much.

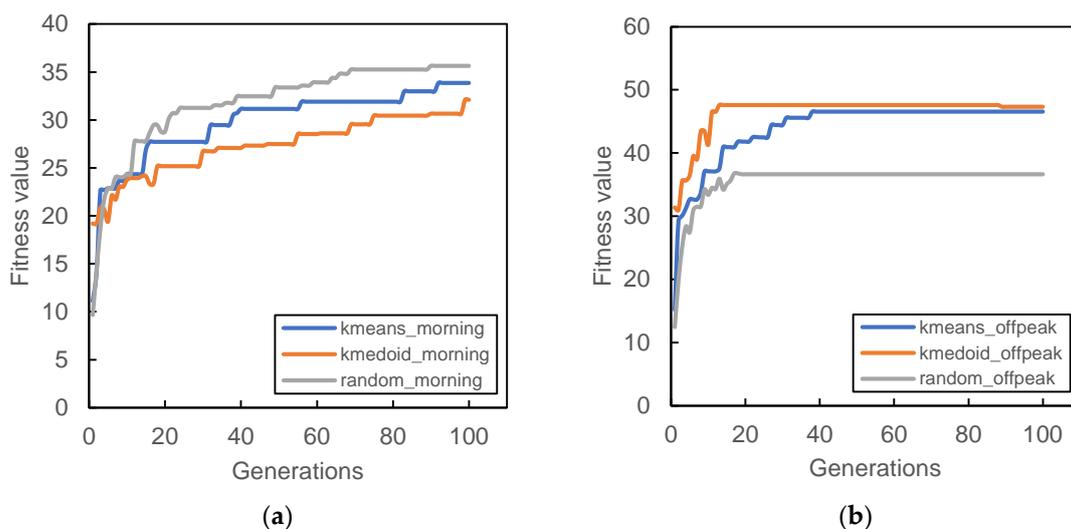
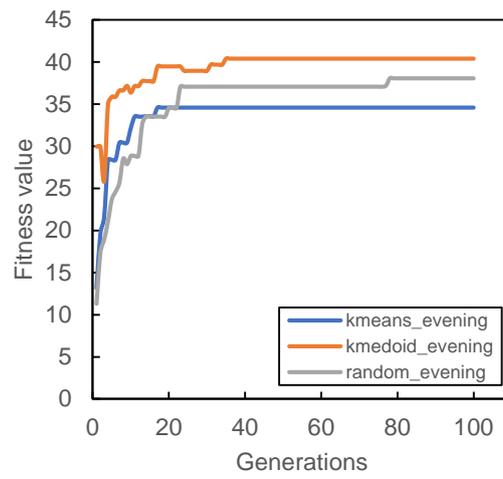
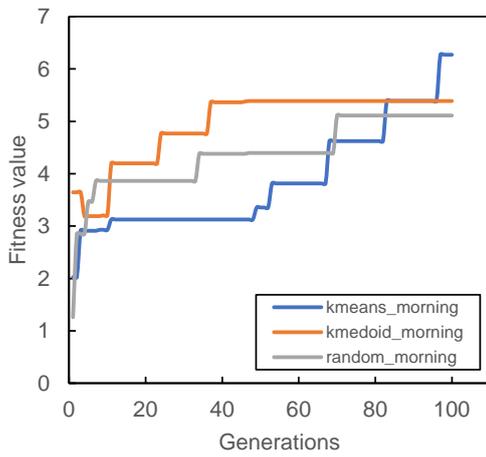


Figure 6. Cont.

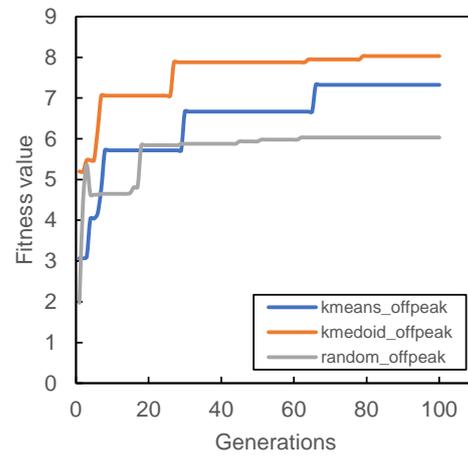


(c)

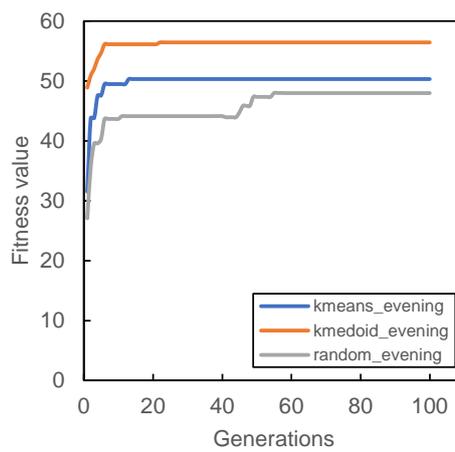
Figure 6. Fitness value—data set 1. (a) Morning peak period; (b) off-peak period; (c) evening period.



(a)

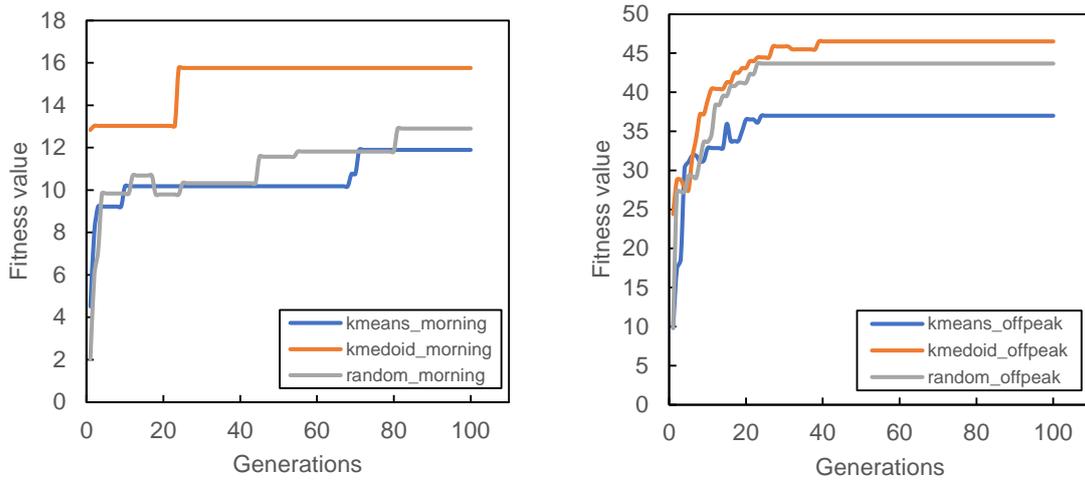


(b)



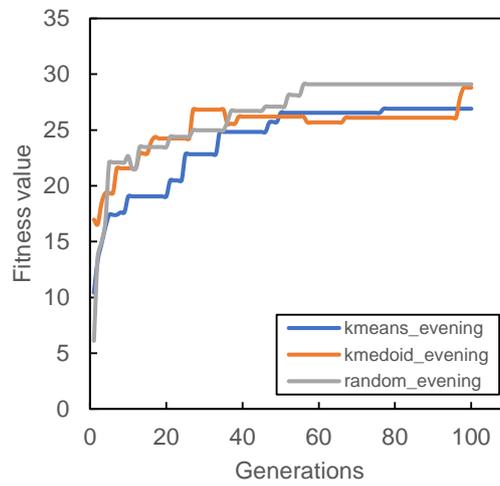
(c)

Figure 7. Fitness value—data set 2. (a) Morning peak period; (b) off-peak period; (c) evening period



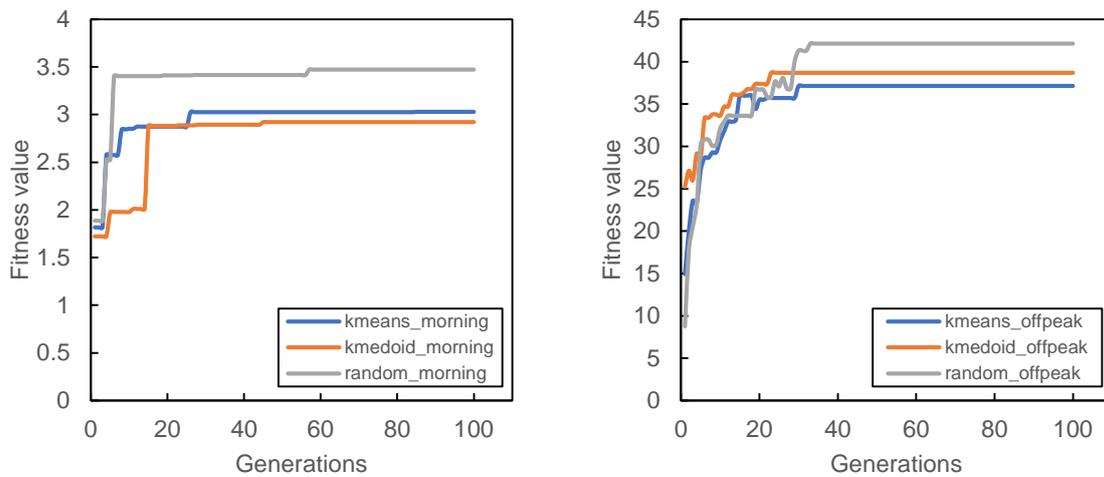
(a)

(b)



(c)

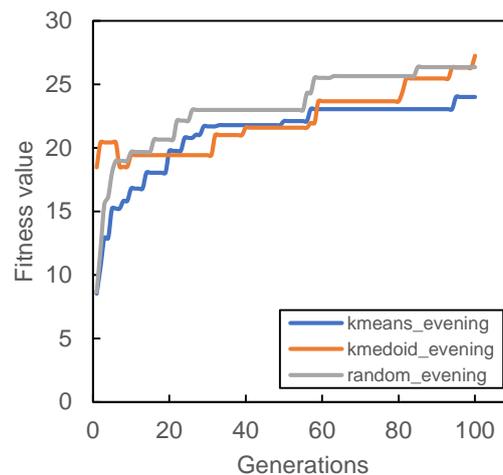
Figure 8. Fitness value—data set 3. (a) Morning peak period; (b) off-peak period; (c) evening period.



(a)

(b)

Figure 9. Cont.



(c)

Figure 9. Fitness value—data set 4. (a) Morning peak period; (b) off-peak period; (c) evening period.

Randomized initial population performances are inconsistent throughout all case studies. In some cases, it performs better than other initial populations (Figures 7a and 9a,b) while in some cases, it provides the worst fitness value (Figures 7b and 8c). The stochastic nature of the station search process may contribute to such results, along with its inability to search all the required spaces in time. The mutation process can only occur within the probability limit set at the beginning of the algorithm, and the limitation of the number of generations also prevented the algorithm from converging to better solutions.

It is also worth noting that some cases provide relatively lower fitness values compared to others; for example, Figures 8a and 9a. The particular reason for that occurrence is connected to the total demand analyzed within that period. Lower demands were recorded during that time period; hence, the algorithm was not able to search more PUDO stations that could cater to the demand, which is reflected in Figure 5c,d during the morning peak where a only few PUDO stations are generated.

A visualization of PUDO stations in the study area can be seen in Figure 10. Only k-medoids are shown here since it generally provides consistent good results. Close observation from all the figures shows that the distribution of those PUDO stations often overlapped regardless of the time period. For example, the highlighted areas 1, 2 and 3 of data set 1 and data set 3 suggest similar trends of PUDO points. Figure 10 also shows that the PUDO points were consistently scarce at the bottom part of the area and clustered at the upper part of the study area. This variation in the spatial distribution of PUDO points is because the upper part of the study area is the central business district, while the bottom part of the study is mainly residential areas. As such, as shown in data set 3 and data set 4, more PUDOs were observed in the bottom part of the study area, which indicates that more trips were made in the residential areas during the weekends, while during the weekdays, PUDO points tend to be located at the upper part of the study area. This finding suggests the similar demand hotspot on some areas throughout the time period. This similarity in demand pattern can also be seen across the observed days.

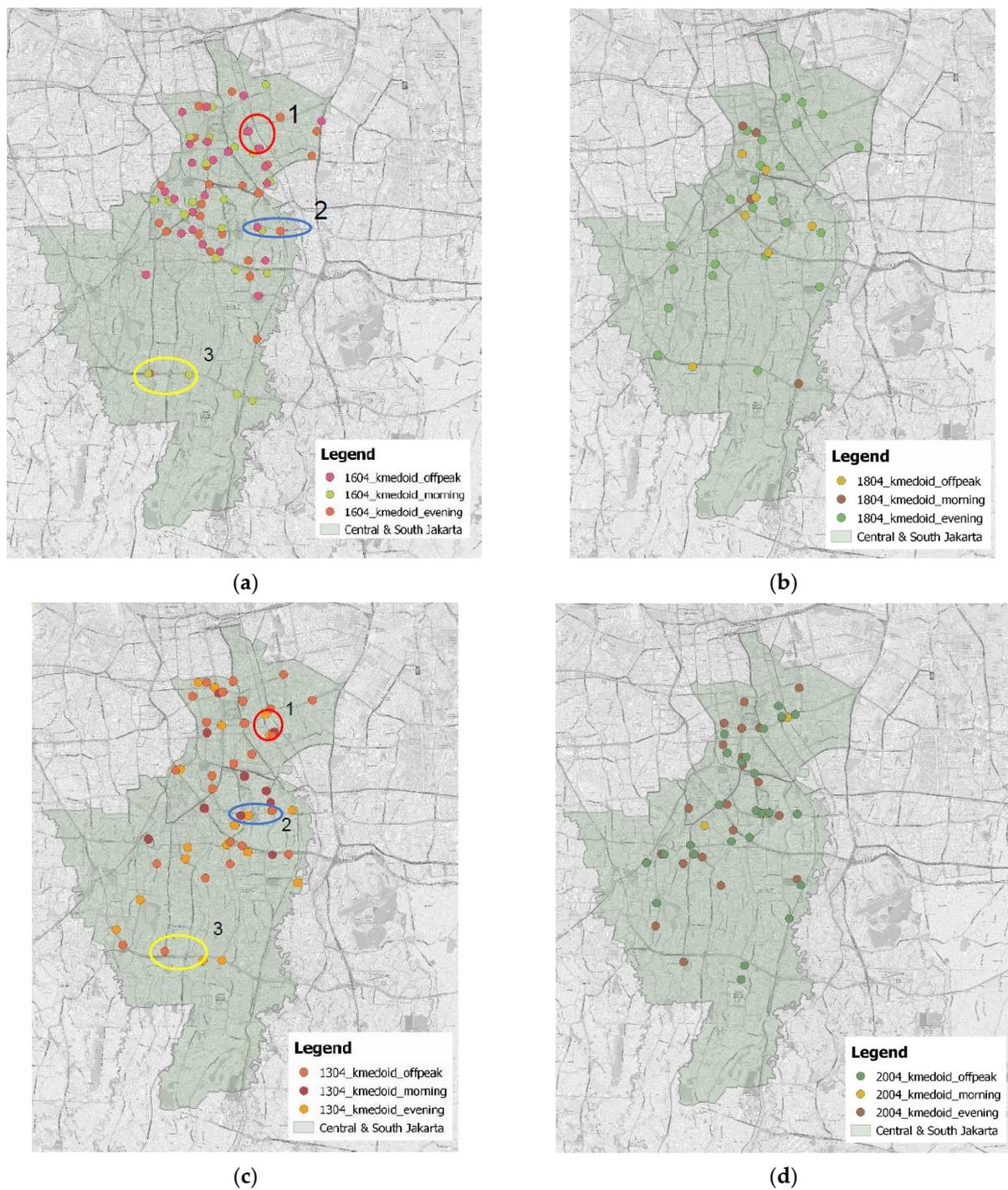


Figure 10. PUDO stations distribution for different dates and time periods. (a) Data set 1; (b) data set 2. (c) Data set 3; (d) data set 4.

5. Conclusions and Future Directions

This paper addressed the use of spatial-temporal ride hailing trip data as a potential determination of PUDO stations within a study area. The distribution of traffic demand within the targeted study area can be carried out based on ride hailing data available from local TNCs. The advancement of mobility-on-demand has proven to be invaluable for traffic engineers in developing overall traffic distribution in the area. For further clarification, the provided data were used in developing appropriate PUDO stations within certain constraints. A genetic algorithm was used in identifying the optimum locations

and incorporating distance and minimum capacity constraints. Unlike the traditional fitness function that used basic functions such as the sum of a set of calculated parameters related to the problem domain, in this study, a refined fitness function was developed. It consists of the combination of the applied constraints of distance and demand coverage by modifying potential interaction gravity model. This refined fitness function takes into consideration the walkable walking distance and the minimum demand that can be served by the suggested PUDO points. These constraints benefit both riders and ride-hailing operators, particularly in maximizing the total saved travel time and travel cost.

The findings suggest that the algorithm was able to successfully adapt in finding the appropriate number of stations according to the observed demand. The initial population of the genetic algorithm in this location optimization only provides miniscule contribution to the final fitness value. Subsequently, the location placement of those PUDO stations often overlapped with each other throughout time period, suggesting that the demand hotspots within the study area do not change significantly throughout the day. In addition to that, the proposed algorithm was able to adapt to find appropriate placement based on observed demand that assisted in determining optimum pick-up and drop-off (PUDO) zones to improve taxi and ride hailing operation management.

Further implementation of such a system can be used for the deployment of autonomous vehicle (AV) and shared autonomous vehicle (SAV) systems and their fleet management [16]. Forthcoming mobility-on-demand may incorporate AVs as primary vehicles and utilizing appropriately placed PUDO stations could improve their travel performance while possibly relieving overall traffic issues. Additionally, examination on the effect of such stations towards a possible SAV-PuT system, where a shared autonomous vehicle (SAV) provides FMLM transport connectivity to public transport line (i.e., train, mass rapid transit, and light train), can be carried out as well [40]. Additional recommendations can be considered to further improve this study, that include considering a wider range of constraints to add to the complexity and adaptability of the algorithm. In this study, generated stations are restricted to 25 stations; however, it is possible for the algorithm to create more stations to better serve the total demand. However, it is clear that the selection of the number of stations used in the algorithm plays a critical role in the location optimization algorithm; hence, the future studies can further investigate the effect of increasing or decreasing the number of stations on clustering results and optimum PUDO locations. Due to certain limitations, the current study uses a simplistic haversine formula to determine the distance between demand data and nearest station. The next consideration will be to take real walking distance into consideration with actual infrastructure within the study area. This allows for better accuracy in calculation as well as improvement in the algorithm process. Another limitation shows some of the resulted PUDO stations located too close to each other and subsequently, their area of service overlapped too much towards each other. We need to spread them out more while taking into consideration the catering of more demand.

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

Funding: This research was funded by FRGS/1/2019/TK01/MUSM/03/1, Ministry of Higher Education Malaysia.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: Thank you to Grab, the ride-hailing company who was responsible for providing detailed trip data for the research.

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

Appendix A

Initial PUDO hotspots were analyzed using the clustering method in R Studio. The clustering result of data set 3 is presented in the main report, while the clustering results of other data sets for each three distinct time periods, evening peak, morning peak and the off-peak period, are attached in the appendix. K-means and k-medoids clustering were used in the clustering method.

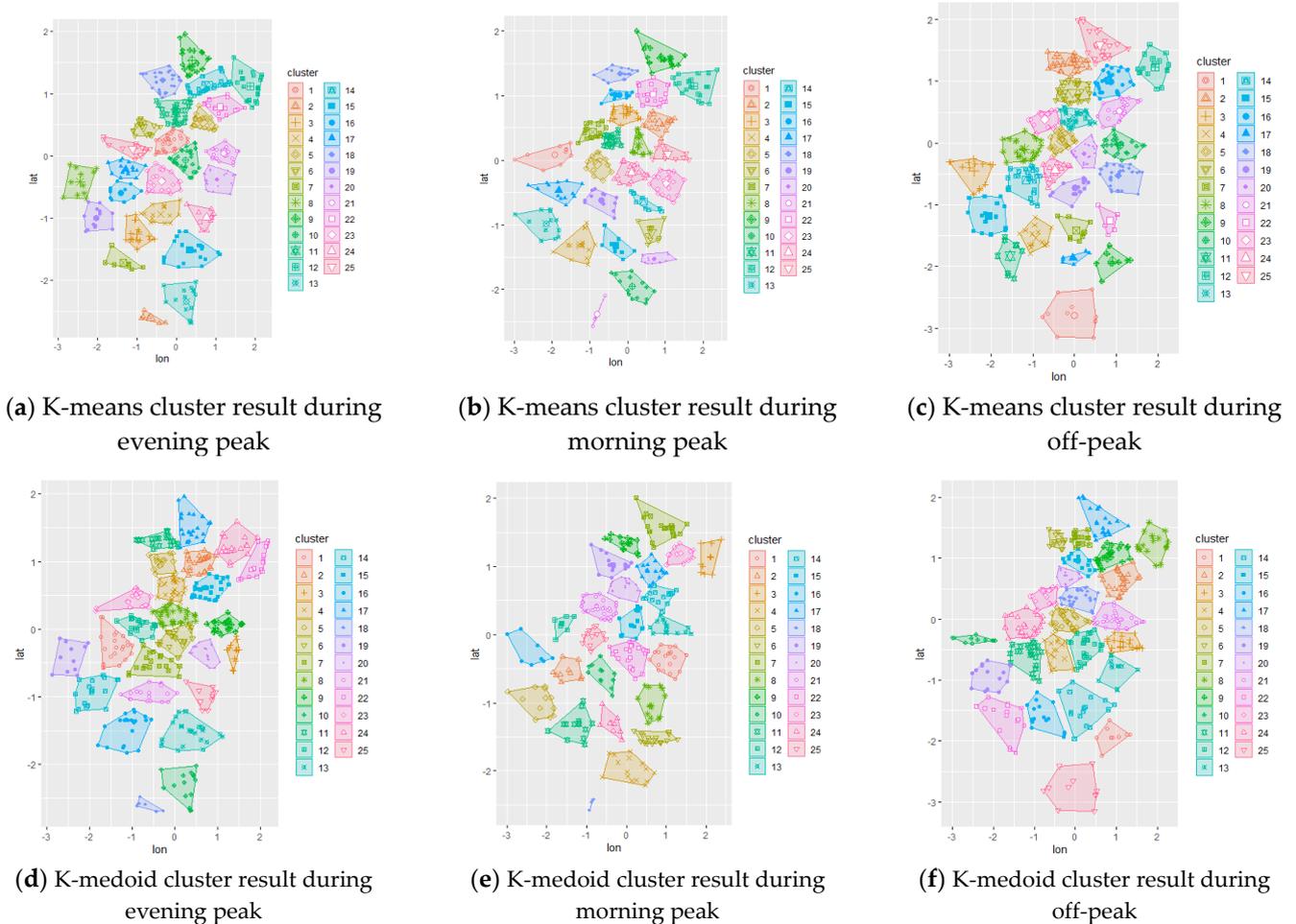


Figure A1. Clustering result for data set 1.

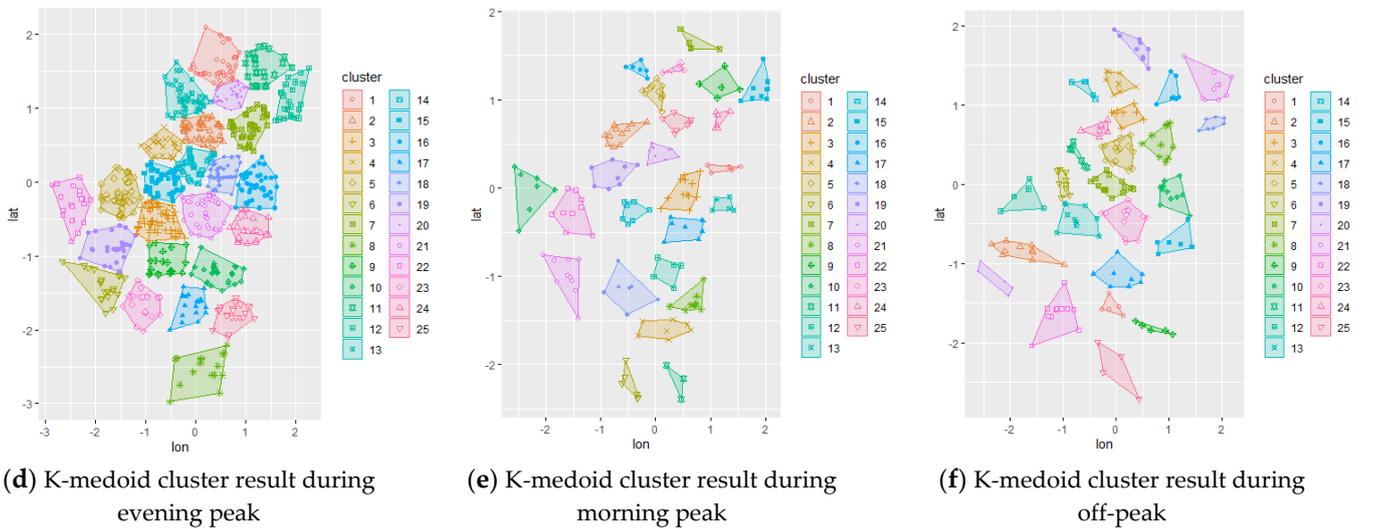
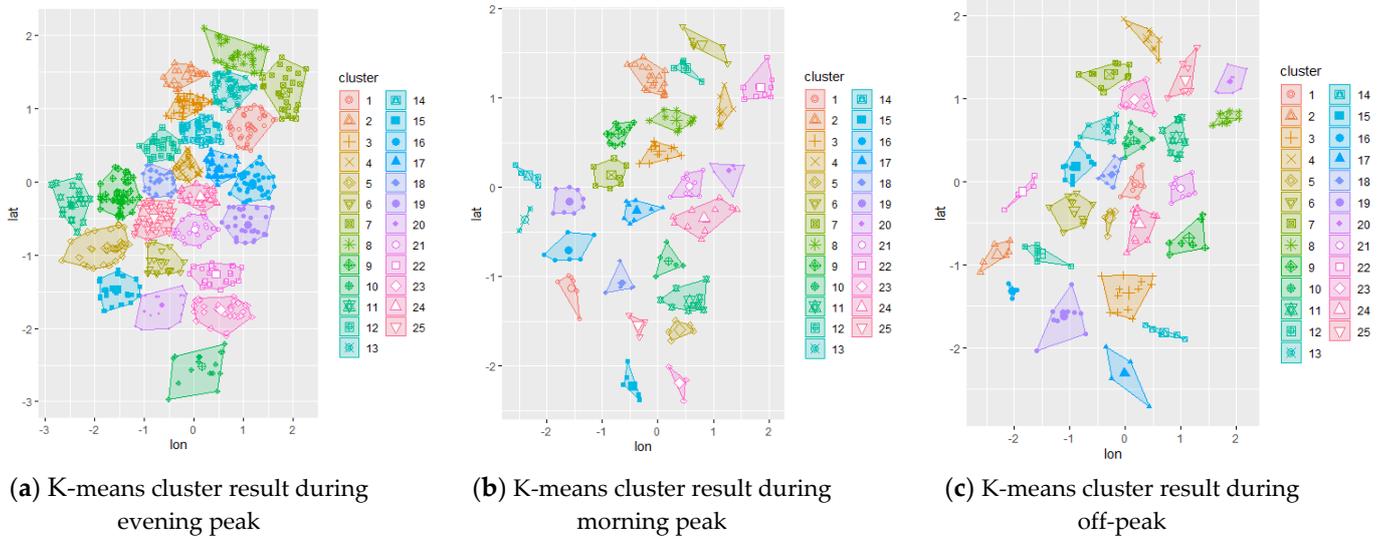


Figure A2. Clustering result for data set 2.

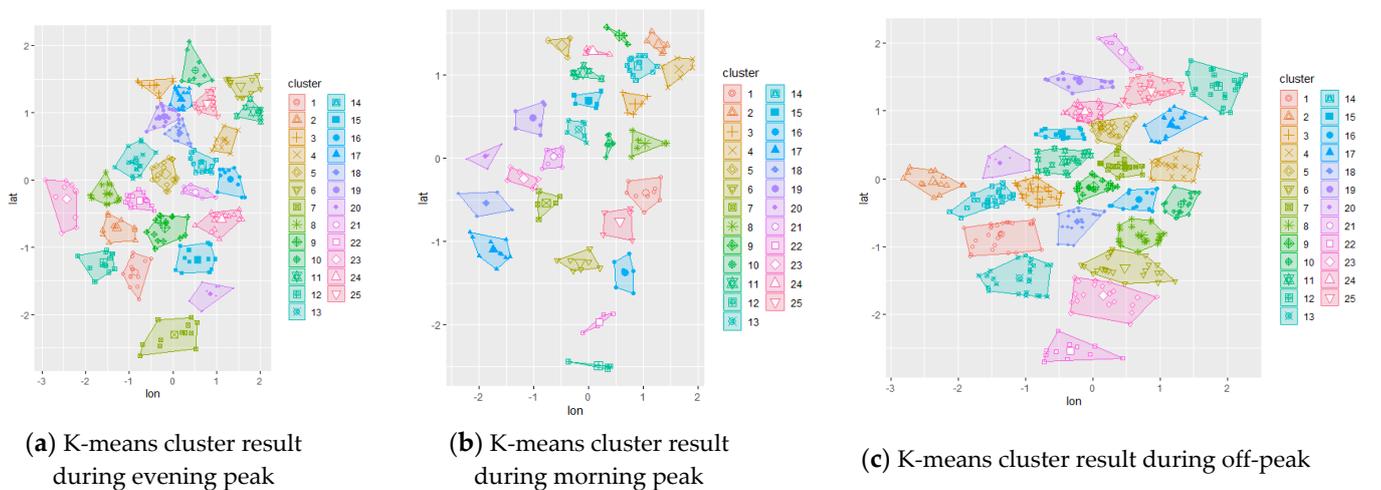


Figure A3. Cont.

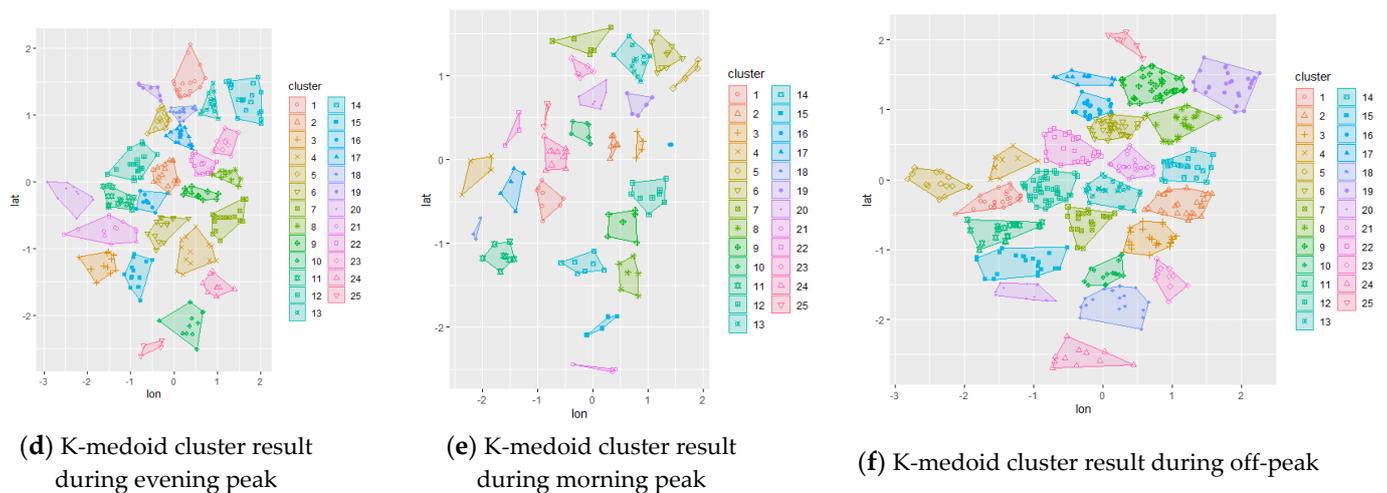


Figure A3. Clustering result for data set 4.

References

- Forghani, M.; Karimipour, F.; Claramunt, C. From cellular positioning data to trajectories: Steps towards a more accurate mobility exploration. *Transp. Res. Part C Emerg. Technol.* **2020**, *117*, 102666. [\[CrossRef\]](#)
- Yuan, J.; Zheng, Y.; Xie, X.; Sun, G. T-Drive: Enhancing Driving Directions with Taxi Drivers' Intelligence. *IEEE Trans. Knowl. Data Eng.* **2013**, *25*, 220–232. [\[CrossRef\]](#)
- Alexander, L.P.; González, M.C. Assessing the Impact of Real-time Ridesharing on Urban Traffic using Mobile Phone Data. In Proceedings of the UrbComp, Sydney, Australia, 10 August 2015.
- Qu, B.; Yang, W.; Cui, G.; Wang, X. Profitable Taxi Travel Route Recommendation Based on Big Taxi Trajectory Data. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 653–668. [\[CrossRef\]](#)
- Veloso, M.; Phithakkitnukoon, S.; Bento, C. Urban mobility study using taxi traces. In Proceedings of the 2011 International Workshop on Trajectory Data Mining and Analysis—TDMA'11, Beijing, China, 18 September 2011; ACM Press: New York, NY, USA, 2011.
- Zheng, Y.; Liu, Y.; Yuan, J.; Xie, X. Urban computing with taxicabs. In Proceedings of the 13th International Conference on Ubiquitous Computing—UbiComp'11, Beijing, China, 17–21 September 2011; ACM Press: New York, NY, USA, 2011.
- Liu, C.; Qin, K.; Kang, C. Exploring time-dependent traffic congestion patterns from taxi trajectory data. In Proceedings of the 2015 2nd IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM), Fuzhou, China, 8–10 July 2015; IEEE: Piscataway Township, NJ, USA, 2015.
- Chen, Z.; Yang, Y.; Huang, L.; Wang, E.; Li, D. Discovering Urban Traffic Congestion Propagation Patterns with Taxi Trajectory Data. *IEEE Access* **2018**, *6*, 69481–69491. [\[CrossRef\]](#)
- Lee, J.; Shin, I.; Park, G.-L. Analysis of the Passenger Pick-Up Pattern for Taxi Location Recommendation. In Proceedings of the 2008 Fourth International Conference on Networked Computing and Advanced Information Management, Gyeongju, Korea, 2–4 September 2008; IEEE: Piscataway Township, NJ, USA, 2008.
- Zhou, D.; Hong, R.; Xia, J. Identification of Taxi Pick-Up and Drop-Off Hotspots Using the Density-Based Spatial Clustering Method. In Proceedings of the CICTP, Shanghai, China, 7–9 July 2017; American Society of Civil Engineers: Reston, VA, USA, 2018.
- Zhang, M.; Liu, J.; Liu, Y.; Hu, Z.; Yi, L. Recommending Pick-up Points for Taxi-drivers Based on Spatio-temporal Clustering. In Proceedings of the 2012 Second International Conference on Cloud and Green Computing, Xiangtan, China, 1–3 November 2012; IEEE: Piscataway Township, NJ, USA, 2012.
- Qian, X.; Zhang, W.; Ukkusuri, S.V.; Yang, C. Optimal assignment and incentive design in the taxi group ride problem. *Transp. Res. Part B Methodol.* **2017**, *103*, 208–226. [\[CrossRef\]](#)
- Huang, X.; Yin, Y.; Lim, S.; Wang, G.; Hu, B.; Varadarajan, J.; Zheng, S.; Bulusu, A.; Zimmermann, R. Grab-Posisi: An Extensive Real-Life GPS Trajectory Dataset in Southeast Asia. In Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Prediction of Human Mobility, Chicago, IL, USA, 5 November 2019; Association for Computing Machinery: Chicago, IL, USA, 2019; pp. 1–10.
- Martins, L.D.C.; de la Torre, R.; Corlu, C.G.; Juan, A.A.; Masmoudi, M.A. Optimizing ride-sharing operations in smart sustainable cities: Challenges and the need for agile algorithms. *Comput. Ind. Eng.* **2021**, *153*, 107080. [\[CrossRef\]](#)
- Agatz, N.; Erera, A.; Savelsbergh, M.; Wang, X. Optimization for dynamic ride-sharing: A review. *Eur. J. Oper. Res.* **2012**, *223*, 295–303. [\[CrossRef\]](#)
- Mourad, A.; Puchinger, J.; Chu, C. A survey of models and algorithms for optimizing shared mobility. *Transp. Res. Part B Methodol.* **2019**, *123*, 323–346. [\[CrossRef\]](#)

17. Fielbaum, A. Optimizing a vehicle's route in an on-demand ridesharing system in which users might walk. *J. Intell. Transp. Syst.* **2021**, 1–20. [[CrossRef](#)]
18. Liu, X.; Titheridge, H.; Yan, X.; Wang, R.; Tan, W.; Chen, D.; Zhang, J. A passenger-to-driver matching model for commuter carpooling: Case study and sensitivity analysis. *Transp. Res. Part C Emerg. Technol.* **2020**, *117*, 102702. [[CrossRef](#)]
19. Czioska, P.; Mattfeld, D.C.; Sester, M. GIS-based identification and assessment of suitable meeting point locations for ride-sharing. *Transp. Res. Procedia* **2017**, *22*, 314–324. [[CrossRef](#)]
20. Stiglic, M.; Agatz, N.; Savelsbergh, M.; Gradisar, M. The benefits of meeting points in ride-sharing systems. *Transp. Res. Part B Methodol.* **2015**, *82*, 36–53. [[CrossRef](#)]
21. Fielbaum, A.; Bai, X.; Alonso-Mora, J. On-demand ridesharing with optimized pick-up and drop-off walking locations. *Transp. Res. Part C Methodol.* **2021**, *126*, 103061. [[CrossRef](#)]
22. Horváth, M.T.; Tettamanti, T.; Varga, I. Multiobjective dynamic routing with predefined stops for automated vehicles. *Int. J. Comput. Integr. Manuf.* **2019**, *32*, 396–405. [[CrossRef](#)]
23. Zheng, Y.; Li, W.; Qiu, F.; Wei, H. The benefits of introducing meeting points into flex-route transit services. *Transp. Res. Part C Emerg. Technol.* **2019**, *106*, 98–112. [[CrossRef](#)]
24. Aïvodji, U.M.; Gambs, S.; Huguët, M.J.; Killijian, M.O. Meeting points in ridesharing: A privacy-preserving approach. *Transp. Res. Part C Emerg. Technol.* **2016**, *72*, 239–253. [[CrossRef](#)]
25. Liu, B. Facility Location Problem. In *Theory and Practice of Uncertain Programming*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 157–165.
26. Mehar, S.; Senouci, S.M. An optimization location scheme for electric charging stations. In Proceedings of the 2013 International Conference on Smart Communications in Network Technologies (SaCoNeT), Paris, France, 17–19 June 2013; IEEE: Piscataway Township, NJ, USA, 2013.
27. Zanakis, S.H.; Evans, J.R. Heuristic “Optimization”: Why, When, and How to Use It. *Interfaces* **1981**, *11*, 84–91. [[CrossRef](#)]
28. Chen, S.; Shi, Y.; Chen, X.; Qi, F. Optimal location of electric vehicle charging stations using genetic algorithm. In Proceedings of the 2015 17th Asia-Pacific Network Operations and Management Symposium (APNOMS), Busan, Korea, 19–21 August 2015; IEEE: Piscataway Township, NJ, USA, 2015.
29. Stiglic, M.; Agatz, N.; Savelsbergh, M.; Gradisar, M. Enhancing urban mobility: Integrating ride-sharing and public transit. *Comput. Oper. Res.* **2018**, *90*, 12–21. [[CrossRef](#)]
30. Andréasson, I.J. Personal Rapid Transit as Feeder–Distributor to Rail. *Transp. Res. Rec. J. Transp. Res. Board* **2012**, *2275*, 88–93. [[CrossRef](#)]
31. Shaheen, S.; Chan, N. Mobility and the Sharing Economy: Potential to Facilitate the First- and Last-Mile Public Transit Connections. *Built Environ.* **2016**, *42*, 573–588. [[CrossRef](#)]
32. Scheltes, A.; de Almeida Correia, G.H. Exploring the use of automated vehicles as last mile connection of train trips through an agent-based simulation model: An application to Delft, Netherlands. *Int. J. Transp. Sci. Technol.* **2017**, *6*, 28–41. [[CrossRef](#)]
33. Zuo, T.; Wei, H.; Chen, N.; Zhang, C. First-and-last mile solution via bicycling to improving transit accessibility and advancing transportation equity. *Cities* **2020**, *99*, 102614. [[CrossRef](#)]
34. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. 2020. Available online: <https://www.R-project.org/> (accessed on 5 May 2021).
35. QGIS.org. QGIS Geographic Information System. QGIS Association. 2021. Available online: <http://www.qgis.org> (accessed on 5 May 2021).
36. Genetic algorithms and Machine Learning. *Mach. Learn.* **1988**, *3*, 95–99. [[CrossRef](#)]
37. Holland, J.H. Genetic Algorithms. *Sci. Am.* **1992**, *267*, 66–72. [[CrossRef](#)]
38. Ke, J.; Yang, H.; Zheng, Z. On ride-pooling and traffic congestion. *Transp. Res. Part B Methodol.* **2020**, *142*, 213–231. [[CrossRef](#)]
39. Syswerda, G. Simulated Crossover in Genetic Algorithms. In *Foundations of Genetic Algorithms*; Elsevier: Amsterdam, The Netherlands, 1993; pp. 239–255.
40. Shen, Y.; Zhang, H.; Zhao, J. Integrating shared autonomous vehicle in public transportation system: A supply-side simulation of the first-mile service in Singapore. *Transp. Res. Part A Policy Pract.* **2018**, *113*, 125–136. [[CrossRef](#)]