TRR
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

Research Article

Transportation Research Record
1–12

DOI: 10.1177/03611981211049143
journals.sagepub.com/home/trr
SAGE

Demand-Driven Optimization Method for Microtransit Services

Cyril Veve[1] and Nicolas Chiabaut[2]

Abstract

Many shared mobility solutions have been developed over recent decades. In the case of mobile technological innovations, new solutions that are more flexible to user demands have emerged. These dynamic solutions allow users to be served by optimizing different aspects such as the detour to pick up a passenger or the waiting time for users. Such methods make it possible to satisfy requests quickly and to match as closely as possible user expectations. However, these approaches usually use fleets composed of numerous small-capacity vehicles to serve each user. By contrast, microtransit aims to serve a more massive demand than conventional shared mobility methods. Our study falls within this context. It aims to identify recurrent patterns of mobility and to verify the possibility of implementing microtransit lines to serve them. In other words, the proposed method identifies spatial and temporal areas where the implementation of a flexible transport line would meet a potential mobility demand. The recurrence of trips in these specific areas provides a guarantee of the reliability of the designed lines.

Keywords

data and data science, artificial intelligence, and advanced computing applications

The sharing of trips in the context of urban mobility has been the subject of numerous studies in the literature over the past 10 years. The emergence of new mobile technologies and technical innovations (GPS, internet wireless connection for mobile phones, and others) have allowed the quick deployment of new shared mobility systems. Many systems have then appeared in urban areas: ridesharing, carsharing, carpooling, bike-sharing, and so forth. These systems aim to overcome problems with other forms of transport, such as the lack of flexibility in public transport or the harmful environmental effects linked to the use of individual vehicles. As shown in Alonso-Mora et al. (*1*) and Najmi et al. (*2*), these solutions have made it possible to offer demand-tailored services and thus reduce certain inconveniences such as waiting or travel times. However, these highly dynamic solutions are designed to work with large fleets of vehicles. The dispersal of the vehicles across a given territory allows users to be picked up and dropped off as quickly as possible and with minimal detours. The vehicles can generally carry from 2 to 10 passengers. Unlike classic public transport, these methods make it possible to respond to an instantaneous and highly targeted geographic demand for mobility to minimize the inconvenience for each user. However, these approaches are not suitable for the design of massive customized lines. Indeed, the techniques used to find the optimal way to serve users are usually related to combinatorial optimization problems: the vehicle routing problem, the dial-a-ride problem, and so forth. By nature, these problems are not solvable in polynomial time, but in exponential time depending on the size of the inputs. That is why it is not conceivable to use these methods for large-capacity vehicles serving many users.

To overcome this problem, the literature has in part turned to new approaches, such as microtransit or customized buses (*3*). Microtransit is defined as a multi-passenger transport solution (usually a shuttle or bus) from the private sector, which can be operated on fixed, flexible, or on-demand routes and times (*4*). The

[1]Univ. Lyon, Univ. Gustave Eiffel, ENTPE, LICIT, Lyon, France
[2]Citec Ingénieur Conseils SAS, Lyon, France

Corresponding Author:
Cyril Veve, cyril.veve@entpe.fr

objective of this system is to detect and pool a consistent number of similar trips in large-capacity vehicles while offering more flexibility than public transport. It comes down to proposing a supply adapted to user demand without necessarily adapting the line to each request individually. Our works fall within this framework.

Even if works such as Simini et al. (*5*) allow the detection of potential corridors of mobility and provide considerable benefits in the literature, they remain relatively far from our goal. Indeed, we do not aim to identify potential demand based on attributes such as population density or other sociologic aspects. This work instead focuses on developing a method to identify a recurrent demand for shared mobility and then on studying the feasibility of a mobility solution to meet this demand. Unlike traditional shared mobility systems, the state of the art concerning microtransit is sparse. Despite the numerous studies that have been carried out in the scientific (*6*, *7*) and generalist (*8*) literature to analyze specific cases, only a few works have focused on the theory of designing microtransit or customized bus lines. Alonso-González et al. (*9*) provide an analysis of the interactions between demand responsive transit (microtransit, customized bus) and fixed transit. Ma et al. (*10*) propose a method for improving the customized bus network in Bejing. Although interesting, the main disadvantage of this study is that the number of seats per vehicle is fixed at 30. Otherwise, the authors do not consider vehicles of different sizes depending on the line and the estimated number of users. They propose improving on this in future work. Although these studies constitute an interesting contribution to the literature, they do not provide more information about the automatic detection of potential recurrent mobility demand. To the best of our knowledge, only one very recent paper addresses the subject of demand recurrence in the responsive transit lines design (*11*). In their study, the authors use the most frequent trip requests to adapt shuttles routes. In this case, historical data are used to adapt the functioning of the lines in real time, but not to entirely design new customized lines. Thus, this study also differs from ours. Indeed, we provide a method for identifying robust patterns of micro-mobility thanks to the analysis of a large data set of historical user trips. This study could constitute the first step to ensuring microtransit system reliability and offer an additional guarantee before deploying a mobility service. As mentioned in Volinski (*12*), microtransit systems can use a static or flexible routing system and schedules. It is important to underline that we do not propose a flexible or optimized method to match perfectly user demand in a specific area. The goal is rather to prove that the proposed method makes it possible to detect and serve a substantial and recurring shared mobility demand.

The rest of the paper is organized as follows. The section, Estimation of the Demand presents the data set and introduces the methodology used to detect the mobility demand patterns. To do this, we analyze a data set containing 314,245 trips from June 1st to 30th, 2011, in New York. The section, Customized Supply Design is dedicated to designing customized transport lines to satisfy the demand. Thereafter we focus on the analysis of the results of the demand estimation and the planning of new lines. The last part is devoted to a discussion of the results.

Estimation of the Demand

The main objective of this section is to show how demand can be divided into spatio-temporal areas containing significant numbers of similar trips. The study focuses on the demand of shared mobility in the Midtown and Upper East Side of New York City (NYC). The objective is to present the method used to obtain clusters of similar and recurrent trips over time (meta-clusters). These clusters will be used to define the instances of the optimization problem presented in the section, Customized Supply Design. The methodology is based on three steps: i) definition of a similarity function to estimate the likeness between two trips; ii) implementation of a clustering method to create clusters of similar trips; and iii) development of a method to detect recurrent clusters over time. However, it is essential to underline that the demand is analyzed from a transportation point of view even if many other aspects could be taken into account: economic, social, or behaviorial. Our method determines an upper bound of the potential of shared mobility.

We use an open-source data set released by the New York City Taxi and Limousine Commission (https://www1.nyc.gov/site/tlc/index.page). Although these data are not fully representative of human mobility since they only correspond to taxi trips, such a data set still provides an attractive proxy for studying individuals' routes within a city. The study focuses on morning peak hours from 8:00 to 11:00 a.m. in June 2011. The area studied is a well known high-density area of mobility in New York City: Midtown and Upper East Side (*13*, *14*). For each trip i, the data set gives access to the following information: departure time t_i^{PU} and location $p_i^{PU} = (x_i^{PU}, y_i^{PU})$ of the pick-up of the passenger(s); arrival time t_i^{DO} and location $p_i^{DO} = (x_i^{DO}, y_i^{DO})$ of the drop-off.

First, it has been shown that the similarity function is used to quantify the likeness between two trips. The method used to detect groups of similar trips in different spatio-temporal areas is then presented. Finally, we investigate if commonalities exist between the clusters of successive studied days.

Modeling Similarity Between Individual Trips

Several attempts exist in the literature to express the similarity between two trips (*15–17*). But these studies aim mainly at estimating the likeness between trajectories using variants of Euclidean distance between points of interest or at measuring the number of points shared by two trajectories (*18*, *19*). This issue is quite different from our goal because the number of observations is much higher in these previous cases. The existing methods used to estimate the likeness between trips based on variants of the longest common subsequence problem (LCSS) need a critical number of observation points to define trips correctly. Here, only the origin and destination locations and times are considered. This choice has been motivated by two main factors. The first is the simplicity of our method and the low complexity of calculating the similarity index. The second is data accessibility. Many data sets exist for which we do not have access to trajectories to describe trips, but only origin–destination pairs. This method has been developed to estimate the similarity between trips easily and to work with many data sets, rather than just with data sets containing trajectories. Unfortunately, the research on the similarity between the origins and destinations of individual trips is very sparse. The main contribution to our work on the similarity between trips is the study of Ketabi et al. (*20*). The authors introduce a similarity measure calculated as the arithmetic mean of the distance between origin–destination locations to evaluate the proximity between two trips i and j. This index is the weighted geometric mean of Euclidean similarity (spatial) and their temporal similarity. In the following, l indicates a pick-up or a drop-off; p_i^l and t_i^l indicate respectively the spatial position and the timestamp of a point i. To be consistent with our modeling framework, we can formulate the index as follows:

$$Sim(i,j) = \left(\frac{1}{2} \sum_{l \in [PU, DO]} e^{\frac{w_1 \ln\left(\frac{1}{1 + d(p_i^l, p_j^l)}\right) + w_2 \ln\left(\frac{1}{1 + |t_i^l - t_j^l|}\right)}{w_1 + w_2}} \right)^{-1} \quad (1)$$

where d is the geodesic distance, and w_1 and w_2 are weighting factors. The main limitation of the approach described above to evaluate the similarity (such as the other attempts based on the Euclidean distance) is that it does not discriminate. It sufficiently enhances the difference between trips when applied to origin–destination locations and desired departure or arrival times only. Specifically, an index of two trips with similar origins or destinations, that is $d(p_i^l, p_j^l)$ close to zero (with d a distance function), but with significant difference in their desired departure or arrival times, that is high values of $|t_i^l - t_j^l|$, still remains low. As a consequence, the ability to cluster methods to capture similar trips deteriorates. To overcome this drawback, we define a function $S(i,j)$ to evaluate the similarity between two trips i and j based on the characteristics of their origins and destinations (*21*). From a physical point of view, the intuition is that two (or more) travelers may have and interest in sharing their trip if they start in the same neighborhood and at the same moment, and want to go to the same destination. The function S must encompass these different spatio-temporal attributes of the trips. The similarity is calculated according to the spatio-temporal commonalities between the trips. Let $S(i,j)$ be the similarity function between trips i and j.

$$\overline{S(i,j)} = \sum_{l \in [PU, DO]} \alpha_l e^{|f^l(i,j)|} \quad (2)$$

where $f^l(i,j)$ is the feasibility function, and α_l is a coefficient. Function f describes the potential of the service to operate the shared trips, that is the ability to pick up (or drop off) the two travelers before both of their desired departure times:

$$f^l(i,j) = |t_i^l - t_j^l| - \gamma d(p_i^l, p_j^l) \quad (3)$$

where γ is the average pace of connecting travelers who want to share a trip. This parameter is a general and synthetic formula to describe the operation of the service and the way in which this service gathers two demand requests into the same vehicle: defining a meeting point, successive pick-ups, and so forth. For example, if the first traveler must walk to the second traveler's pick-up point, then γ is equal to the inverse of the walking speed. If this distance is traveled by car, meaning that the service offers door-to-door service, then γ is equal to the inverse of the vehicle speed. Consequently, f is positive if the match can be realized before the two desired departure times t_i^l and t_j^l, whereas f is negative if travelers have to experience delay to make the match possible. Moreover, α_l is equal to $\frac{1}{2}$ if $f^l(i,j)>0$ and to $\frac{3}{2}$ otherwise because it is more disadvantageous to be delayed. In addition to this first index of similarity $\overline{S(i,j)}$, excessive distances/durations for rendezvous are penalized. Thus, penalties θ_x^l and θ_t^l are added when, respectively, the distances between origin (or destination) locations and departure (or arrival) times of trips i and j exceed, respectively, specific thresholds, δ_x^l and δ_t^l:

$$\theta_x^l = e^{d(p_i^l, p_j^l) - \delta_x^l} \quad \forall l \;/\; d(p_i^l, p_j^l) > \delta_x^l \quad (4)$$

$$\theta_t^l = e^{|t_i^l - t_j^l| . \frac{\delta_x^l}{\delta_t^l} - \delta_t^l} \quad \forall l \;/\; |t_i^l, t_j^l| > \delta_t^l \quad (5)$$

Otherwise, these penalties are null.

In this manner, $S(i,j) = \overline{S(i,j)} + \theta_x^l + \theta_t^l$ defines a sharp function that enhances the differences between trips and facilitates identification of similar travelers in the data set. Notice that S is minimal (and equal to 1) when the two trips are exactly identical.

Figure 1 highlights the evolution of S (in red) with regard to the difference between two hypothetical trips. A comparison with the similarity function *Sim* of Ketabi et al. (*20*) can be made. We complete this comparison by introducing a naive combination of the Euclidean distance and the absolute difference in time:

$$s(i,j) = 1 + \sum_{l \in [PU, DO]} (|t_i^l - t_j^l| + d(p_i^l, p_j^l)) \qquad (6)$$

The constant 1 is added to have the same minimal value as S and *Sim*. Notice that the different functions have been normalized to make them comparable. It is also important to remember that trips are similar when different functions are minimized. It turns out that *Sim* and s have a linear increase (quasi-linear for *Sim*) with the difference in departure and arrival times (curves of evolution with regard to the difference in origin–destination locations are strictly similar). By contrast, our function S is much more discriminant. Consequently, differences between trips are clearly enhanced. This means that it will be easier to cluster similar trips with good intra-cluster homogeneity and, simultaneously, significant inter-cluster dissimilarity.

Detection of Similar Trips for Individual Days

The method presented here allows us to detect spatio-temporal areas where there is a significant number of similar trips. The function of similarity introduced above only estimates the likeness between two trips but does not detect clusters of similar trips. To detect such groups, a clustering algorithm is used. The function of similarity allows us to compute the similarity matrix required by a clustering algorithm.

A variant of a well known clustering density-based method (*22*) is applied for each day to detect groups of similar trips. This method only requires two parameters: a threshold ε and a minimum number of points *MinPts*, which have to be in a radius ε so that the studied point is considered as an element of the cluster. The parameter ε is the maximal distance between trips, that is, the maximal value of S, allowed to consider them as similar and group them into the same cluster. However, this method must be slightly adapted to detect groups of different density. Thus, a successive DB-SCAN clustering is performed, that is *itdbscan* (see [*23*] for more information), using the similarity function S as the distance, while updating iteratively the values of the parameters. Starting with a large value of $MinPts = M$ and a drastic ε, makes it possible to identify large groups of travelers in the initial data set of trip T. In other words, we first detect large and high-density clusters. Then, the DB-SCAN method is applied

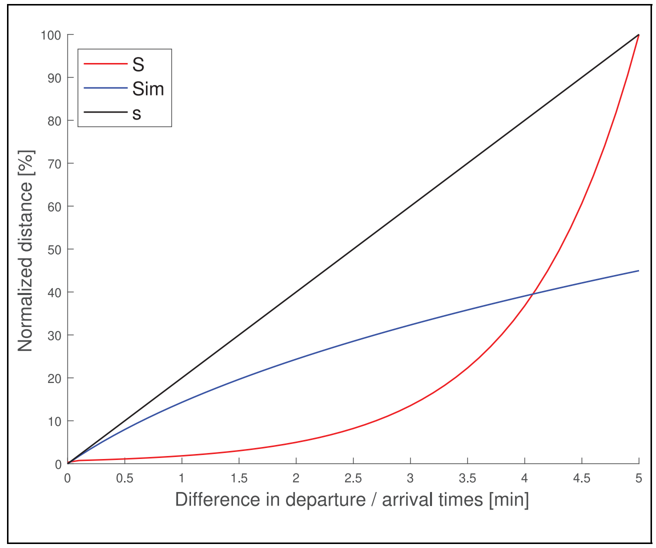

Figure 1. Comparison of different similarity indexes.

on the remaining non-clustered trips to detect groups of size $M - 1$. This process is repeated until $MinPts = 2$.

Clusters detected have different sizes, from two to 74 trips gathered into the same group. This highlights that the shared mobility demand may take many aspects requiring different forms of transportation services to be optimally satisfied. Figure 2 depicts four clusters with different sizes and characteristics. The average travel length $\overline{l_k}$ is directly the arithmetic average of the length of n_k trips within the cluster k, whereas the average travel time $\overline{\tau_k}$ is the arithmetic average of the duration of the n_k trips. Figure 2*e* shows the number of clustered trips and the total number of trips per day. The developed method detects almost 85% of similar trips per day on average in the studied zone.

Identification of Regular Demand Patterns for Multiple Days

Once this daily analysis is done, we investigate if commonalities in the clusters can be identified. Many approaches exist to derive the most representative partition from a group of partitions, such as meta-clustering or consensus learning (*24*). Here, we use the same clustering method to maintain consistency when scaling-up. In the following, to reduce the computational time, we focus the study on the 14 days of the data set for which the ratio of clustered trips is the highest: June 6th to 19th 2011. The objective is now to find out if similar trips (relatively close departure and arrival locations and times) are made several times during the studied period. These recurrent spatio-temporal areas are called meta-clusters. For that purpose, each cluster previously found is considered a new trip, formed by the centroid of its pick-up and the centroid of its drop-off. Centroids correspond to

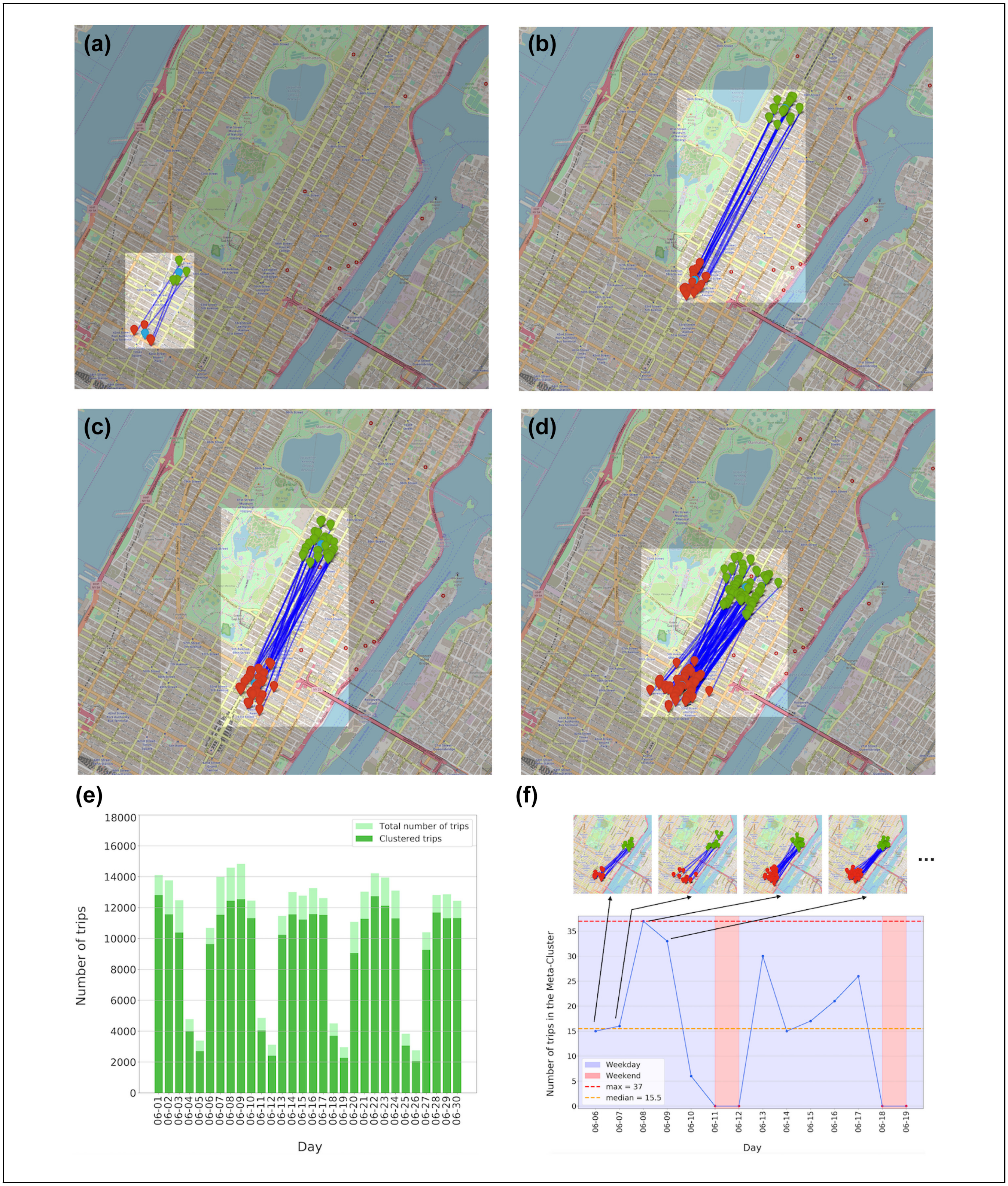

Figure 2. (*a–d*) Clusters with different characteristics, the pick-ups are depicted in green and drop-offs in red. n_k denotes the number of trips in the cluster *k*, $\overline{l_k}$ denotes the average length of trips in *k* and $\overline{\tau_k}$ denotes the average duration of trips in *k*; (*a*) $n_k = 4$ $\overline{l_k} = 0.93$ km $\overline{\tau_k} = 6.5$ min; (*b*) $n_k = 19$ $\overline{l_k} = 2.58$ km $\overline{\tau_k} = 13.3$ min; (*c*) $n_k = 30$ $\overline{l_k} = 1.95$ km $\overline{\tau_k} = 9.2$ min; (*d*) $n_k = 74$ $\overline{l_k} = 1.56$ km $\overline{\tau_k} = 9.8$ min; (e) ratio of clustered trips per day in Midtown and Upper East Side from 8:00 to 11:00 a.m.; (*f*) Example of demand graph for a randomly selected meta-cluster.

the mean origin–destination locations and mean departure or arrival times of the clustered trips. This information can be useful to design the transportation supply because centroids can be the locations of common meeting points of the standby areas of shared vehicles. A second clustering is then performed, it returns clusters with similar characteristics (without taking into account the initial day when the trips were made). In other words, two clusters are in the same meta-cluster if their centroids have close departure and arrival locations and times. To this end, function S is extended to consider sets of trips. In other words, $|t_i^l - t_j^l|$ and $d(p_i^l, p_j^l)$ are respectively replaced by the mean distances, that is $\frac{1}{n_k}\sum_{i=1}^{n_k} |t_i^l - t_j^l|$ and $\frac{1}{n_k}\sum_{i=1}^{n_k} d(p_i^l, p_j^l)$ where n_k is the number of trips inside the cluster k. We then normalized these values because the acceptable delays are strongly related to the length of the trips. Consequently, the quality index of cluster k, $Q(k)$, is the function S applied to the set of clustered trips divided by the average length of the trips within cluster k. Notice that we aim at minimizing the quality index, that is, the best clusters present values of Q close to zero. Indeed, $Q(k)$ is low when i) the spatio-temporal distance between origins is low, ii) the spatio-temporal distance between destinations is low, and iii) the mean travel distance is large. A cluster k is selected if and only if $Q(k)$ is below a specific threshold Q_{max}. In the remainder of this study, we set Q_{max} at 3. It corresponds to a restrictive matching policy. Interestingly, we observe that more than 94% of the daily clusters are recurrent from one day to another.

The representation of a meta-cluster on a 2-D map is difficult to analyze because the time dimension is not accounted for. Consequently, we prefer to focus on the evolution of the daily cluster sizes and the localization of the related origin–destination locations. Each meta-cluster can be depicted as a graph of the demand. Figure 2*f* shows the graph of the demand for a randomly selected meta-cluster. This figure shows that in the same spatio-temporal area, similar trips can be seen every day, except on weekends. Each meta-cluster provides precise information about its location, its estimated departure and arrival times, and the total number of trips performed each day. It is important to note that different individuals perform these trips from one day to another. However, global human mobility is remarkably regular; this is a valuable insight to tune transportation services and favor shared mobility efficiently.

Customized Supply Design

As mentioned in the previous section, the spatio-temporal areas containing similar and regular trips (meta-clusters) are detected. The next objective of the study is to find a way to serve the pick-up and drop-off points in each of these meta-clusters while respecting a set of constraints: time windows, vehicle capacity, size of the fleet, and so forth. A minimal example of the developed method is depicted in Figure 3. Figure 3*a* shows a set of three meta-clusters; each of them contains several clusters of similar trips. A green marker and a red marker linked by a blue line depict a cluster containing several similar trips. The green and red markers designate respectively a cluster's pick-up and drop-off points. A meta-cluster is depicted by an aggregation of clusters in the same spatial area. On average, each cluster contains 6.22 similar trips. Moreover, a meta-cluster contains, on average 8.64 clusters. In other words, each meta-cluster contains on average 53 trips with very similar characteristics (see the section, Results). The method aims at designing a line of transport serving a set of meta-clusters with compatible characteristics (size, time windows, etc.). Depending on the meta-clusters chosen, the number and size of vehicles required will be different. To quantify the potential demand of a tour, we plot the total number of trips per day served by a tour going through these meta-clusters. Figure 3*c* depicts the total number of trips served per day for the set of meta-clusters depicted in Figure 3*a*.

Then, the selection of a set of meta-clusters to find a potential tour of vehicles is performed. We use the median number of trips per day in a meta-cluster as an indicator of its size. For each studied period, a minimal median value required is defined; we filter the meta-clusters with a median inferior to this value. This method allows us to obtain a reasonable number of meta-clusters to solve a relatively small instance of the optimization problem. However, it should be noted that the solution is optimal for each period on which we solve the problem, but an optimal result is not guaranteed for a set of periods. The meta-clusters found are used to define a dial-a-ride problem (DARP) instance. Given the large number of meta-clusters, it is impossible to directly model vehicle tours from the whole set of meta-clusters. That is why we focus this study only on high capacity vehicle tours. This method allows us to serve many passengers by solving smaller problem instances. The main advantage of this method is that the calculation time depends on the number of meta-clusters served rather than directly on the number of passengers. This is why it is crucial to find the largest possible meta-clusters. If we solve the problem for a set of large meta-clusters, the number of passengers effectively served will be significantly higher.

There are many variants for the DARP problem; (*25*, *26*) give a list of them based on different objective functions. In our case, we use a variant presented

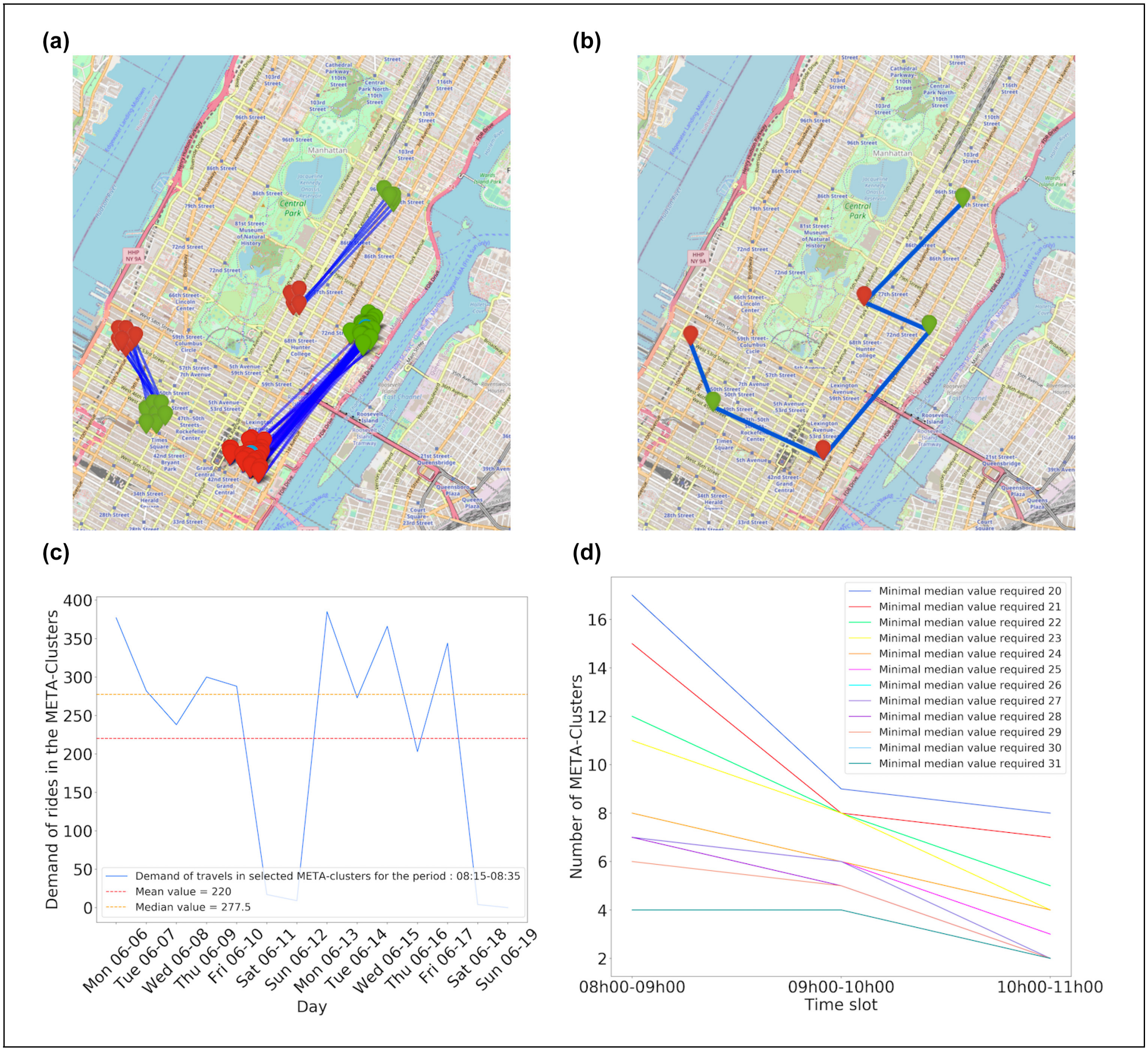

Figure 3. Green markers depict the pick-up and red markers the drop-offs: (*a*) three meta-clusters randomly chosen between 8:15 and 8:35 a.m.; (*b*) example of line designed, serving the centroids of pick-up and drop-off of each meta-cluster; (*c*) total number of trips per day served on the set of three meta-clusters; and (*d*) shows the number of meta-clusters for each time slot in function of the minimal median value of trips per day required.

in Cordeau (*27*). This model is depicted below in Equations 7 to 19. The model is based on a three-index formulation.

Let $G = (V, A)$ be a directed graph. The set of vertices V is partitioned as follows: the first and last elements are two copies of the depot, elements from index 1 to n are pick-up and elements from index $n+1$ to $2n$ are drop-off. P denotes the set of pick-ups and D the set of drop-offs. A request is a couple $(i, n+i)$, where $i \in P$ and $n+i \in D$. The load of each vertex is defined as q_i, with $q_0 = q_{2n+1} = 0$, $q_i \geqslant 0$ for i in $\{1, ..., n\}$ and $q_i = -q_{i-n}$ for i in $\{n+1, ..., 2n\}$. A service duration $d_i \geqslant 0$ with $d_0 = d_{2n+1} = 0$. K denotes the set of vehicles. The capacity of a vehicle $k \in K$ is Q_k, and T_k denotes the maximal duration of a route for a vehicle k. The arc set is defined as: $A = \{(i,j) \mid i = 0, j \in P$ or $i,j \in P \cup D, i \neq j$ and $i \neq n+j$, or $i \in D, j = 2n+1\}$ the cost of traversing an arc (i,j) with a vehicle k is c_{ij}^k, and the travel time between two nodes i and j is t_{ij}. L denotes the maximal ride time and the time window of a vertex i is $[e_i, l_i]$. x_{ij}^k is a binary variable equal to 1 if and only if (i,j) is traversed by a vehicle $k \in K$. Let u_i^k be the time at which a vehicle k

starts servicing a vertex i, w_i^k the load of vehicle k leaving vertex i, and r_i^k the ride time of user i.

(DARP)

$$Minimize \sum_{k \in K} \sum_{i \in V} \sum_{j \in V} c_{ij}^k x_{ij}^k \tag{7}$$

subject to

$$\sum_{k \in K} \sum_{j \in V} x_{ij}^k = 1 \qquad (i \in P), \tag{8}$$

$$\sum_{i \in V} x_{0i}^k = \sum_{i \in V} x_{i,2_{n+1}}^k = 1 \qquad (k \in K), \tag{9}$$

$$\sum_{j \in V} x_{ij}^k - \sum_{j \in V} x_{n+i,j}^k = 0 \qquad (i \in P, k \in K), \tag{10}$$

$$\sum_{j \in V} x_{ji}^k - \sum_{j \in V} x_{ij}^k = 0 \qquad (i \in P \cup D, k \in K), \tag{11}$$

$$u_j^k \geqslant (u_i^k + d_i + t_{ij}) x_{ij}^k \qquad (i,j \in V, k \in K), \tag{12}$$

$$w_j^k \geqslant (w_i^k + q_j) x_{ij}^k \qquad (i,j \in V, k \in K), \tag{13}$$

$$r_i^k \geqslant u_{n+i}^k - (u_i^k + d_i) \qquad (i \in P, k \in K), \tag{14}$$

$$u_{2n+1}^k - u_0^k \leqslant T_k \qquad (k \in K), \tag{15}$$

$$e_i \leqslant u_i^k \leqslant l_i \qquad (i \in V, k \in K), \tag{16}$$

$$t_{i,n+i} \leqslant r_i^k \leqslant L \qquad (i \in P, k \in K), \tag{17}$$

$$\max(0, q_i) \leqslant w_i^k \leqslant \max(Q_k, Q_k + q_i) \qquad (i \in V, k \in K), \tag{18}$$

$$x_{ij}^k = 0 \; or \; 1 \qquad (i,j \in V, k \in K), \tag{19}$$

This formulation of the constraints can be described as follows: constraints 8 and 10 ensure that each request is served once by the same vehicle. Constraints 9 and 11 aim at verifying that each vehicle starts and ends its trips at a specific point. This point is often a depot, but in our model it is a point defined beforehand. Constraints 12 to 14 define starts of service times, vehicle loads, and user ride times. Constraints 15 to 18 ensure that the constraints about the maximal duration of a route 15, a time window 16, the maximal ride time 17, and the maximal capacity of each vehicle 18, will be feasible. This model presents several interesting aspects: multiple vehicles and time windows for pick-up or drop-off. The main objective of this method is to minimize the total route length. However, several other constraints can be added, such as vehicle capacity, maximum route duration, or maximum ride time for users. Nevertheless, it is essential to note that the meta-clusters previously found are independent of the method chosen to serve them and vice versa. Indeed depending on the objective searched, one approach may be more interesting than another. For example, it could be interesting to use a method to minimize the total route length for a transportation network company. From a user point of view, it could be more interesting to use a technique allowing a reduction in the waiting time before being served. Several methods aim to satisfy an objective function depicted as a combination of constraints such as transportation time, ride time, excess of maximum ride time, waiting time, time windows violations, and so forth (*28*). A comparison with these sophisticated methods will be made in a future study.

Results

This section is devoted to the results of the proposed method for the case of NYC. First, the meta-clusters are presented and analyzed. Secondly, based on this demand decomposition, the optimization method is tested and evaluated.

Selection of the Spatio-Temporal Areas

First of all, it is interesting to analyze the characteristics of the meta-clusters found. As mentioned in section on the method for estimating the demand, in the studied area between 8:00 and 11:00 a.m., almost 85% of the trips can be considered similar. Moreover, more than 94% of trips are recurrent, that is, these trips can be observed almost every day. There were 2,136 spatio-temporal areas detected as zones where there was a recurrent potential demand of shared mobility. On average, each meta-cluster contained 53 trips. Once again, it is important to notice that different users surely perform these trips. We consider below that the user meeting points are defined as the centroids of pick-ups (respectively drop-offs) of a meta-cluster. It is thus interesting to know the spatial and temporal difference between the centroids and the points of pick-ups and drop-offs. Table 1 shows that the spatial distances are close to 200 m. The average temporal shifts are nearly 6 min which is entirely acceptable. It shows that the meta-clusters found are relatively close to the

Table 1. Average Spatial and Temporal Distances Between Pick-Up, Drop-Off and the Centroids of the Meta-Clusters

Variable	Result
Average spatial distance pick-up/centroid of pick-up	0.21 km
Average spatial distance drop-off/centroid of drop-off	0.21 km
Average shift between departure times/centroid of departure times	6.16 min
Average shift between arrival times/centroid of arrival times	6.27 min

initial clusters estimated from the real user rides. The average travel distance and time in the meta-clusters are respectively 1.71 km and 11.1 min. Although these data are not fully representative of human mobility since they only correspond to taxi trips, such a data set provides an attractive proxy for studying individual routes within a city.

First, it is necessary to select a reasonable number of meta-clusters that will be considered as an instance of the optimization problem. As noted above, the median number of trips per day in a meta-cluster is used as an indicator of its size. Figure 3*d* shows that for each one-hour period the number of meta-clusters depends on the chosen minimal median value. In other words, a meta-cluster is counted if and only if its median value of trips per day is greater or equal to the chosen value. It is important to note that the number of points effectively treated in the DARP will be for each period $2 * \textit{number of clusters}$, because a vehicle serves a pick-up and a drop-off for each meta-cluster. According to Figure 3*d*, the minimal median value 24 has been selected to find potential routes with many users with very short execution times.

Demand-Driven Route Optimization

This section aims to prove the possibility of designing microtransit lines to serve the meta-clusters previously found. We perform an analysis on 18 meta-clusters. However, all 2,136 meta-clusters could be used to design new lines in a case of real implementation. The meta-clusters can be selected according to their mean number of trips per day or to maximize the lines' efficiency. This study allows us to perform a robust and precise estimation of the historical daily demand before the design of the lines. By extension, it is easy to estimate the size of the vehicles needed on each designed line. Contrary to works such as Ma et al. (*10*), the size of the vehicles is not fixed to a particular value. In the specific case presented in this section, we use large-capacity vehicles containing 80 people, which is an adapted size to serve the selected meta-clusters.

According to the results shown above, the selection of the spatio-temporal areas, three periods of one hour for which the meta-clusters contain at least a median of 24 trips per day are selected (18 meta-clusters). Figure 4*a* shows the potential number of trips per day that can be served in the period from 8:00 to 11:00 a.m. This result illustrates the interests of the method: the 18 meta-clusters selected actually represent on average 614 trips per day. The median number of trips served per day for this set of meta-clusters is 756. Also, we note that the demand is extremely regular every day of the week (except weekends) for the two weeks of analysis. In the case of an effective implementation of optimized lines, it would be interesting for the service to be operated from Monday to Friday.

The parameters used for DARP are adjusted for each period of one hour. Each centroid of pick-up and drop-off of the meta-clusters is inserted in the sets P and D. The number of vehicles V for each period is depicted in Table 2. For each vehicle, its capacity $Q_k = 80$, which corresponds to the average capacity of a bus. For each arc (i,j), the cost c_{ij} is defined as the spatial distance between i and j. For each node i, we set the service duration $d_i = 2$ min. The load of each pick-up q_i is defined as the number of users to serve. The load of each drop-off is defined as $-q_i$.We consider that each point can theoretically be served until 20 min before or after the planned departure. However this value is not representative of the real difference between the desired service times and the effective times of service, but it provides an upper and lower limit that should not be exceeded. If the values e_i and l_i are not correctly chosen, the constraint 16 cannot be satisfied, then no solution will be found. The travel time between two nodes i and j is estimated according to the results presented in (*14*). We set the average speed for a vehicle to 9.65 km/h.

Figure 4, *b* to *d*, show the map of the designed routes for each period. Each color designates a specific transport line. Table 2 indicates the result of the DARP. For the three periods, routes allowing service serve to all the selected meta-clusters in less than 9 s are found. This result proves that the method is a good way to design lines serving many users (more than 600 trips per day on average). Besides, for each period, we calculate the average delays and time advances for each point served by the optimized tour. This value is estimated by the difference between the deesired times of departure and arrival (given by the centroids of the meta-clusters) and the hour of service given by the solving of DARP. These values show that the developed method has relatively little impact on the demand. Indeed, on average, the delay is 12 min and the advance is 10.6 min, which is acceptable since the number of users served is high.

There are no classical optimization methods that we know of to find round serving such a quantity of similar and recurrent trips in such a tight timeframe. The theoretical studies on DARP (*26*) show that the exact method used in this paper can solve instances up to 36 points. It would not be possible to solve instances with so many passengers without using an aggregation method in meta-clusters. The existing dynamic methods such as (*1*, *29*, *30*) obtain trips delay between 2 and 6 min; however, these services work with large fleets of vehicles with limited capacities (between 2 and 10). Moreover, these methods work only on networks with a limited number of nodes.

As mentioned in the introduction, the literature dedicated to microtransit is not as developed as shared

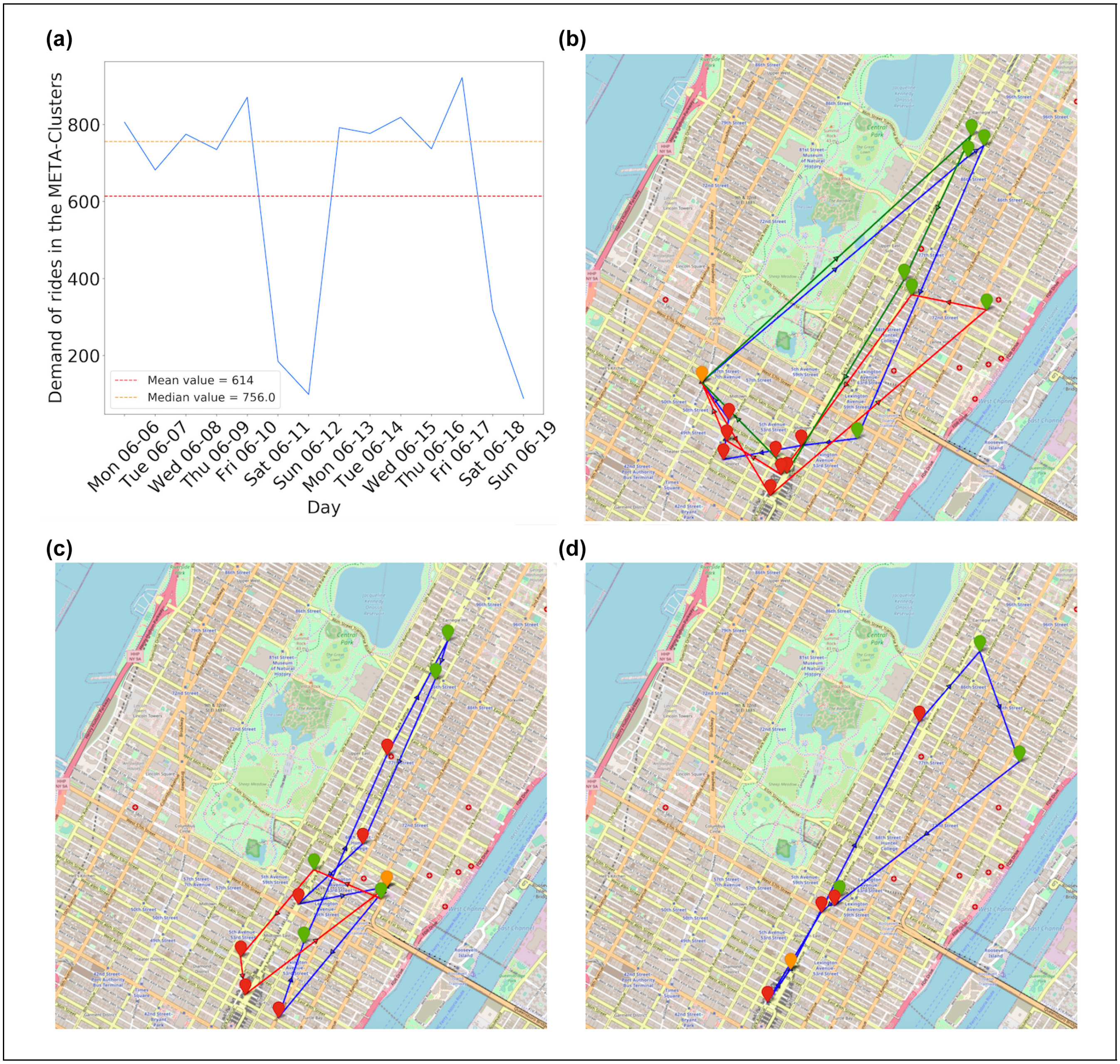

Figure 4. (*a*) The total number of users effectively served in the set of meta-clusters selected in the section on the selection of the spatio-temporal areas; and (*b–d*) the customized lines found for each period presented in Table 2.

Table 2. Result of the Search of Rounds for the Three Time Periods from 8:00 to 11:00 a.m.

	Period 1	Period 2	Period 3	Total
Time	8:00 to 9:00 a.m.	9:00 to 10:00 a.m.	10:00 to 11:00 a.m.	8:00 to 11:00 a.m.
Number of meta-clusters served	8	6	4	18
Number of vehicles required	3	2	1	6
Average delay (min)	16	11	9	12
Average advance (min)	9	12	11	10.6
Total travel distance	36.15 km	19.2 km	12.43 km	67.78 km
Computation time	6.85 s	1.37 s	0.05 s	8.27 s

Table 3. Comparison of Performances

	Meta-clustering method	RTRCB method
Studied time slot	8:00 to 11:00 a.m.	7:00 to 8:00 a.m.
Number of vehicles used	3	9
Average number of passengers served per bus	204	26
Average delivery delay (min)	12 min*	3 min**
Capacity of the vehicles	80	26
Average distance by vehicles	22.59 km	19.35 km
Total travel distance (km)	67.78	174.17
Execution time (s)	8.27	NA

Note: RTRCB = Real-Time Responsive Customized Bus.
*In Relation to the Schedules of the Centroids of the Meta-Clusters.
**For 94.44% of the Travels
NA = not available

mobility services. If several studies are devoted to analyzing specific cases of study, only a few of them have been focused on the theory of designing microtransit or customized bus lines. Among them, Han et al. (*11*) caught our attention because the authors address the subject of demand recurrence in the design of responsive transit lines. Even though the goals of our two respective studies are quite different, it is interesting to compare the two approaches. Unlike in our work, the authors do not provide a method to identify robust patterns of micromobility but use the most frequent trip requests to adapt the shuttle routes. In other words, they use the historical data to adjust the functioning of the lines in real time, but not to design entirely new customized lines. Table 3 summarizes different metrics to compare two methods. The first column shows the metrics obtained with our approach, and the second column refers to those presented in Han et al. (*11*). Their study focuses on a time slot from 7:00 to 8:00 a.m. and uses nine buses with a capacity of 26 people. Our experiment uses only three buses and focuses on a time slot from 8:00 to 11:00 a.m. The capacity of the vehicles is 80.

Table 3 shows that the average distances covered by the vehicles are relatively similar, close to 20 km. However, the total travel distance needed to serve the pick-up and drop-off is 39% lower in our study. This is a result of the difference in size between the two studied areas. We observe that the average delay is reduced by 9 min on average in their solution. However, the number of passengers served by one vehicle is eight times more important in our study. The execution time for this experiment is not clearly mentioned in their paper. For the method presented in our paper, the execution time for detecting the lines is about 8 s.

This comparison illustrates very well the differences between the two approaches. The method proposed by Han et al. (*11*) seeks to satisfy a demand by taking into account the specificities of each user trip. Conversely, our study aims to construct more massive lines to serve numerous users who are not captured by classic public transport services while conserving consistent average delay. Table 3 seems to show that the initial objectives of our paper have been achieved. Moreover, despite the constraints in vehicle speeds in the area studied (Midtown and Upper East Side), the indicators stay attractive in a context where the urban density of mobility is very high.

Conclusion

This article presents an optimization method based on a decomposition of the demand and a resolution of DARP on a reduced instance. This data-driven method allows us, for the first time, to identify clusters of similar and regular trips over time (meta-clusters). These meta-clusters are then considered as points to serve in an instance of DARP. The main interest of this method is to design tours of vehicles to serve many potential users. As shown in the section on results, the main advantage of this method is that the execution time of the optimization problem does not depend on the number of users served, but only on the number of meta-clusters served.

As part of this study, the analysis of pattern recurrence was carried out over two weeks. However, the proposed method makes it possible to detect regular patterns over much more extended periods. This can be particularly useful in the case of effective implementation of transport lines. It is also possible to integrate many other parameters into the objective function of DARP to design lines as close as possible to actual user demand.

The results obtained show us that rounds of large-capacity vehicles (80 people) can be identified. Making it possible to serve on average more than 600 trips a day with calculation times lower than 9 s. This method of designing microtransit lines allows the development of high capacity lines based on the demand of mobility. Moreover, it allows us to obtain lines with restricted spatio-temporal deviations from the demand described by the centroids of the meta-clusters. Setting up massive lines close to the initial demand of users provides a partial response to the last mile problem, which is one of the main obstacles to shifting users of private vehicles to shared modes of transport.

Several ways are studied to complete the current method. The first is to take real-time aspects into account in the method. This can be done in several ways, either with instant classification or by taking into account the results obtained to anticipate future demand. Another interesting aspect is the design of more or less dynamic lines according to the demand in a studied area. For example, depending on the number of users and the required responsiveness of the service, different solutions can be implemented: classic or dynamic bus lines, taxis, and so forth. Finally, taking into account the dynamic aspects of the network to choose routes according to the network's particular events: congestion, roadworks, and so forth, seems to be an excellent way to improve the current method. Finally, the method's scalability will be widely studied to maximize the number of data processed and thus the veracity of the results obtained.

Author Note

Nicolas Chiabaut is now affiliated to Univ. Lyon, Univ. Gustave Eiffel, ENTPE, LICIT, Lyon, France.

Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: Nicolas Chiabaut; data collection: Cyril Veve; analysis and interpretation of results: Nicolas Chiabaut, Cyril Veve; draft manuscript preparation: Nicolas Chiabaut, Cyril Veve. All authors reviewed the results and approved the final version of the manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Cyril Veve https://orcid.org/0000-0003-3336-1470
Nicolas Chiabaut https://orcid.org/0000-0003-0450-4890

References

1. Alonso-Mora, J., S. Samaranayake, A. Wallar, E. Frazzoli, and D. Rus. On-Demand High-Capacity Ride-Sharing via Dynamic Trip-Vehicle Assignment. *Proceedings of the National Academy of Sciences*, Vol. 114, No. 3, 2017, pp. 462–467.
2. Najmi, A., D. Rey, and T. H. Rashidi. Novel Dynamic Formulations for Real-Time Ride-Sharing Systems. *Transportation Research Part E: Logistics and Transportation Review*, Vol. 108, 2017, pp. 122–140.
3. Tsubouchi, K., H. Yamato, and K. Hiekata. Innovative On-Demand Bus System in Japan. *IET Intelligent Transport Systems*, Vol. 4, No. 4, 2010, pp. 270–279.
4. Shaheen, S., and A. Cohen. Shared Ride Services in North America: Definitions, Impacts, and the Future of Pooling. *Transport Reviews*, Vol. 39, No. 4, 2019, pp. 427–442.
5. Simini, F., M. C. González, A. Maritan, and A. -L. Barabási. A Universal Model for Mobility and Migration Patterns. *Nature*, Vol. 484, No. 7392, 2012, pp. 96–100.
6. Shaheen, S., A. Stocker, J. Lazarus, and A. Bhattacharyya. *RideKC: Bridj Pilot Evaluation: Impact, Operational, and Institutional Analysis*. Transportation Sustainability Research Center (TSRC), UC Berkeley, 2016.
7. Liu, T., and A. A. Ceder. Analysis of a New Public-Transport-Service Concept: Customized Bus in China. *Transport Policy*, Vol. 39, 2015, pp. 63–76.
8. METRO. Chariot Launches Crowd Sourced Shuttle in NYC. http://www.metro-magazine.com/shuttle/news/724657/chariot-launches-crowd-sourced-shuttle-in-nyc. 2017.
9. Alonso-González, M. J., T. Liu, O. Cats, N. Van Oort, and S. Hoogendoorn. The Potential of Demand-Responsive Transport as a Complement to Public Transport: An Assessment Framework and an Empirical Evaluation. *Transportation Research Record: Journal of the Transportation Research Board*, 2018. 2672: 879–889.
10. Ma, J., Y. Yang, W. Guan, F. Wang, T. Liu, W. Tu, and C. Song. Large-Scale Demand Driven Design of a Customized Bus Network: A Methodological Framework and Beijing Case Study. *Journal of Advanced Transportation*, Vol. 2017, 2017.
11. Han, S., H. Fu, J. Zhao, J. Lin, and W. Zeng. Modelling and Simulation of Hierarchical Scheduling of Real-Time Responsive Customised Bus. *IET Intelligent Transport Systems*, Vol. 14, No. 12, 2020, pp. 1615–1625.
12. Volinski, J. *Microtransit or General Public Demand–Response Transit Services: State of the Practice*. TCRP Synthesis of Transit Practice, No. Project J-7, Topic SB-30. Transportation Research Board, Washington, D.C., 2019.
13. Ferreira, N., J. Poco, H. T. Vo, J. Freire, and C. T. Silva. Visual Exploration of Big Spatio-Temporal Urban Data: A Study of New York City Taxi Trips. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 19, No. 12, 2013, pp. 2149–2158.
14. NYC Department of Transportation. New York City Mobility Report. https://www1.nyc.gov/html/dot/downloads/pdf/mobility-report-singlepage-2019.pdf. 2019.
15. Abraham, S., and P. S. Lal. Spatio-Temporal Similarity of Network-Constrained Moving Object Trajectories Using Sequence Alignment of Travel Locations. *Transportation Research Part C: Emerging Technologies*, Vol. 23, 2012, pp. 109–123.
16. Tiakas, E., A. N. Papadopoulos, A. Nanopoulos, Y. Manolopoulos, D. Stojanovic, and S. Djordjevic-Kajan. Trajectory Similarity Search in Spatial Networks. *Proc., 2006 10th International Database Engineering and Applications Symposium (IDEAS'06)*, IEEE, Delhi, India, 2006, pp. 185–192.
17. Toohey, K., and M. Duckham. Trajectory Similarity Measures. *Sigspatial Special*, Vol. 7, No. 1, 2015, pp. 43–50.
18. van Kreveld, M., and J. Luo. The Definition and Computation of Trajectory and Subtrajectory Similarity. *Proc., 15th annual ACM International Symposium on Advances in*

Geographic Information Systems, Seattle, Washington, ACM, 2007, p. 44.

19. Wang, H., H. Su, K. Zheng, S. Sadiq, and X. Zhou. An Effectiveness Study on Trajectory Similarity Measures. *Proc., 24th Australasian Database Conference-Volume 137, Australian Computer Society, Inc.*, Adelaide, Australia, 2013, pp. 13–22.
20. Ketabi, R., B. Alipour, and A. Helmy. Playing With Matches: Vehicular Mobility Through Analysis of Trip Similarity and Matching. *Proc., 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, Seattle Washington, 2018, pp. 544–547.
21. Chiabaut, N., and C. Veve. Identifying Twin Travelers Using Ridesourcing Trip Data. *Transport Findings*, 2019, p. 7.
22. Ester, M., H. -P. Kriegel, J. Sander, and X. Xu. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases With Noise. *KDD*, 1996, Vol. 96, pp. 226–231.
23. Veve, C., and N. Chiabaut, Estimation of the Potential Demand of Shared Mobility Services Through Clustering of Similar Individual Trips. Presented at 99th Annual meeting of the Transportation Research Board, Washington, D.C., 2020, No. 20-00168.
24. Lopez, C., L. Leclercq, P. Krishnakumari, N. Chiabaut, and H. Van Lint. Revealing the Day-To-Day Regularity of Urban Congestion Patterns With 3D Speed Maps. *Scientific Reports*, Vol. 7, No. 1, 2017, p. 14029.
25. Ho, S. C., W. Szeto, Y.-H. Kuo, J. M. Leung, M. Petering, and T. W. Tou. A Survey of Dial-A-Ride Problems: Literature Review and Recent Developments. *Transportation Research Part B: Methodological*, Vol. 111, 2018, pp. 395–421.
26. Cordeau, J. -F., and G. Laporte. The Dial-A-Ride Problem: Models and Algorithms. *Annals of Operations Research*, Vol. 153, No. 1, 2007, pp. 29–46.
27. Cordeau, J. -F. A Branch-and-Cut Algorithm for the Dial-A-Ride Problem. *Operations Research*, Vol. 54, No. 3, 2006, pp. 573–586.
28. Jorgensen, R. M., J. Larsen, and K. B. Bergvinsdottir. Solving the Dial-A-Ride Problem Using Genetic Algorithms. *Journal of the Operational Research Society*, Vol. 58, No. 10, 2007, pp. 1321–1331.
29. Santi, P., G. Resta, M. Szell, S. Sobolevsky, S. H. Strogatz, and C. Ratti. Quantifying the Benefits of Vehicle Pooling With Shareability Networks. *Proceedings of the National Academy of Sciences*, Vol. 111, No. 37, 2014, pp. 13290–13294.
30. van Engelen, M., O. Cats, H. Post, and K. Aardal. Enhancing Flexible Transport Services With Demand-Anticipatory Insertion Heuristics. *Transportation Research Part E: Logistics and Transportation Review*, Vol. 110, 2018, pp. 110–121.