Decision Analytics Journal 2 (2022) 100013

Contents lists available at ScienceDirect

Decision Analytics Journal

journal homepage: www.elsevier.com/locate/dajour

A novel two-phase location analytics model for determining operating station locations of emerging air taxi services

Akhouri Amitanand Sinha [a], Suchithra Rajendran [a,b,*]

[a] Department of Industrial and Manufacturing Systems Engineering, University of Missouri Columbia, MO 65211, USA
[b] Department of Marketing, University of Missouri Columbia, MO 65211, USA

ARTICLE INFO

Keywords:
Location analytics
Urban air mobility
Electric vertical takeoff and landing
Air taxi
Infrastructure location decision

ABSTRACT

Air taxi is a new and emerging form of urban air mobility (UAM) for reducing traffic congestion in metropolitan cities by providing on-demand aviation services to millions of customers daily. These aviation services leverage the electric vertical takeoff and landing (eVTOL) technology, and hence, could operate even from building rooftops. This study aims to determine the location of infrastructure facilities for efficient eVTOL air taxi network operation in urban areas and its neighborhood. We utilize a unique two-phase procedure that integrates the multi-criteria warm start technique with an iterative k-means clustering algorithm. To evaluate the effectiveness of the proposed approach, we compare the model's performance with the results obtained in the previous studies for various settings, such as percentage of time savings, demand satisfaction, passenger willingness to fly rate, and on-road travel limits. The results show the proposed methods yield a better solution than previous studies in terms of several metrics such as the Davies Bouldin index, number of clusters, and multi-criteria attributes.

1. Introduction

Traffic congestion in major cities, such as Boston, Chicago, Washington DC, and New York City, has resulted in commuters spending over 130 h annually on road [1]. This causes an increase in accidents, in turn leading to stress, which negatively impacts the citizens. Besides, these gridlocks also have a significant impact on the economy due to a loss in productivity, as well on the environment because of an increase in pollution ([2,3]). In order to avoid these repercussions, individuals prefer to avail of public transportation, such as the subways and busses [4] and choose cycling for short to medium distances [5]. However, this causes an additional strain on passengers as they would have to adjust their work timings based on the transit system schedule [6] and affects their daily jobs due to physical strain on the body [7,8]. To address this issue and in an effort to provide a faster and more efficient method of commute, several logistic companies, such as Uber, Kitty Hawk, Airbus, Lilium, and Boeing, are currently in the process of manufacturing an innovative aviation service called air taxi. Although it is expected that during the initial stages of operation, commuters would have to pay a premium price to avail the services [9,10], a market study by the National Aeronautics and Space Administration (NASA) concluded that increase in operational efficiencies and technological advancement would make the cost per passenger mile for air taxis comparable to ground transportation [11]. The current study focuses on improving the network operations by identifying optimal infrastructure locations in an urban city.

Air taxi, urban air mobility (UAM) service, operates using the concept of electric vertical takeoff and landing (eVTOL), while minimizing energy and power requirements [12,13]. They can be used for a travel range of about 100 miles [9,14], and hence, could cater to the transportation needs of customers in metropolitan cities and their neighborhoods. In recent years, the design, development, and testing of air taxis are pursued by several companies across the globe in countries such as China, India, New Zealand, Singapore, and the USA [15–17]. For instance, Airbus has introduced the Airbus' A^3 Vahana program for testing the feasibility of self-piloting VTOL technologies [18]. Various parameters, such as cruise altitude, impact on the environment, speed of the vehicle, and completion time, are examined to evaluate the advantages in the design and use of air taxis [19–21]. Existing research suggested air taxi operation range of approximately 80 km in major metropolitan cities [22].

1.1. System overview and problem description

Fig. 1 depicts the operations involved in a typical air taxi network, based on the sequence of events given in [9,23,24], and [25]. In order to avail of the service, a customer would need to enter the pickup and destination details in the aviation app, like the current on-demand taxi

* Corresponding author.
E-mail addresses: ask54@mail.missouri.edu (A.A. Sinha), RajendranS@missouri.edu (S. Rajendran).

https://doi.org/10.1016/j.dajour.2021.100013
Received 23 June 2021; Received in revised form 25 October 2021; Accepted 6 November 2021
Available online 19 November 2021

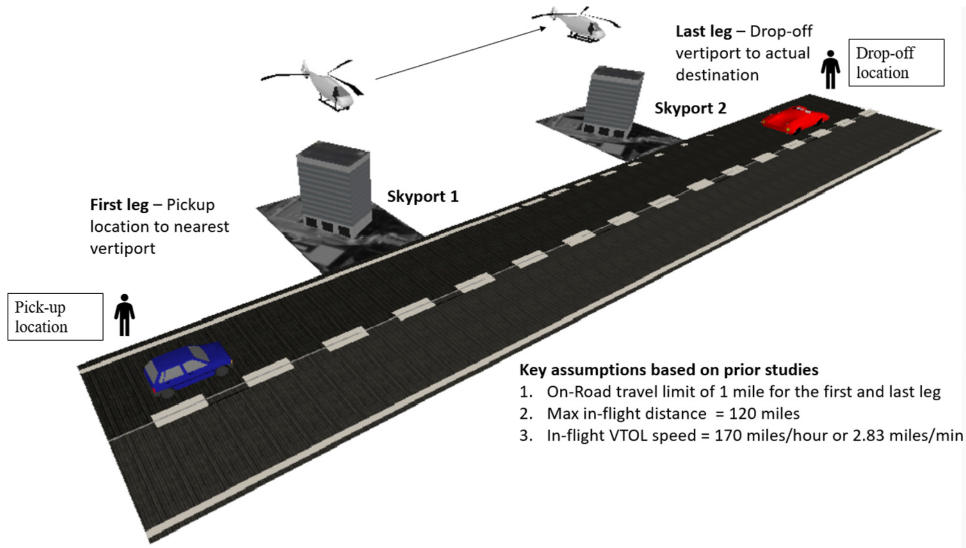

Fig. 1. Pictorial representation of air taxi operations.

Table 1
Terminologies used in this air taxi transportation study.

Term	Definition
Vertiport	A large station consisting of multiple takeoff/landing facilities along with charging and maintenance sites
Vertistop	A smaller station with only one helipad used entirely for customer pickup and dropoff
Tuple	List of elements that describe a potential site location consisting of the following components: rental cost per month, road facility, average salary per year, population coverage and total number of estimated trips made by air taxi per day per 1000 population in that region
Time-savings (TS)	Time saved by a passenger as a result of using air taxi service in comparison to a regular taxi ride
Passenger willingness to fly rate (PR)	Total percentage of eligible passengers who are willing to avail of the air taxi service. Some passengers might have reservations to leverage air taxi services due to safety concerns and affordability.
On-road travel limit (RL)	Overall portion of air taxi service travel made on road (in miles). It consists of two components: travel made from actual pickup point to the origin vertiport/vertistop (first leg) and travel made from destination vertiport/vertistop to actual dropoff location (last leg)
Demand satisfaction (DS)	Total percentage of customers whose on-road travel is less than a mile

reservation system. The app would then determine the cost of travel through all alternatives, such as air taxi (only if the user is eligible), regular taxi, and public transit. The traveler can choose to fly via air taxi based on factors like travel time, ride fare, and willingness to fly. If the passengers would like to avail of the air taxi service, depending on the distance from the pickup point to the nearest source skyport (first leg) as well as from the destination vertiport/vertistop to dropoff point (last leg), they can choose to either walk or travel by car trip that would be provided by the logistics company. Table 1 summarizes the terminologies used in this study.

Among the several decisions that have to be made for efficient network operations, air taxi site selection is extremely crucial as it has a significant impact on other decisions, such as demand estimation, fleet procurement and pricing strategy. Nonetheless, only very limited works in the literature have focused on developing location analytics models for air taxi services (e.g., [26,27]). Moreover, prior studies have predominantly examined a single objective function, such as minimizing total cost, maximizing revenue. Nevertheless, the application of multiple criteria models for location selection problems has been proven to perform better than single objective models [28]. In this paper, a multi-criteria warm start algorithm is adopted in Phase-1 to generate the potential list of air taxi vertiport/vertistop locations based on several conflicting socio-economic factors, such as rental cost, population density of an area, number of air taxi trips per day per 1000 customers, average salary in that neighborhood, and road facility. Phase-2 is associated with using a constraint clustering approach to provide recommendations on air taxi facility locations using the input from Phase-1 as the seed solution. We leverage the New York City (NYC) potential air taxi demand data existing in the literature and compare the results of the proposed approach with that of the traditional k-means algorithm and the technique discussed by Rajendran and Zack [27]. A hypothetical example illustrating the proposed approach is discussed in Appendix.

The remaining paper is organized as follows. The literature review on air taxi design and operations is presented in Section 2. The proposed multi-criteria warm start technique and the iterative k-means clustering algorithm are discussed in Section 3. The data description is given in Section 4. Results and comparison of the solution obtained using the proposed method with the existing models are presented in Section 5. Finally, conclusions and future work are summarized in Section 6.

2. Literature review

Several works have been conducted in recent years to examine the emerging air mobility design and network operations and are reviewed in this section.

2.1. Multiple criteria decisions making in air taxi design and operations

While the single objective problem aims to achieve the "best" solution with respect to one objective, multi-objective models result in "compromise" solution that achieves the best tradeoff between the conflicting criteria. For efficient air taxi design and operations, there are several conflicting criteria that have to be considered. From the design perspective, there are the three major air taxi vehicle types: vectored thrust, lift + cruise and wingless, each having its own strengths and limitations [19]. Vector thrust allows the air vehicle to modify its path based on propulsion direction (Hua et al., 2015). It provides a highly efficient cruising ability and speed to the aircraft. Harrier series is the most popular example of an aircraft using vector thrust technology (Zhou et al., 2020). The lift + cruise technology enables the manufacturers to add a dedicated lift engine alongside the cruise engine. This reduces excess drag and the amount of fuel used during the cruise. The thrust-to-weight ratio during the cruise is nearly 0.1 (Finger et al., 2019). It has multiple electric rotors, and the failure of one propeller will have no impact on the performance of others (Moore, 2020). A wingless multicopter design, as the name implies, has no wings present on the vehicle (Ozdemir et al., 2014). They rely on multiple propellers for thrust propulsion and thus provide better control and less vibration [29].

On the operational side, Mane and Crossley (2012) utilized integer programming to synchronously solve aircraft design for improved performance and allocation to operators' problems. In contrast, de Jong (2007) explored the operational costs associated with urban air mobility using dynamic programming. Piwek and Wisniowski (2016) defined parameters such as passenger capacity, fuel consumption and flight level to be critical for small air transport (SAT) aircraft. Vascik and Hansman (2017) identified various constraints such as noise and social recognition, location of air taxi stations and air traffic services faced by urban air transportation in Los Angeles (LA). Clearly, each of the designs has its own advantages and limitations, and hence MCDM models could be used to determine the best-compromised design configuration.

The competitiveness of air taxi operations depends on several conflicting criteria, such as cost, travel time, safety, sustainability, and regulation policies [30–33]. Al Haddad et al. [34] studied the components influencing the consumer opinion for adopting UAM. By conducting a survey using exploratory factor analysis, the authors concluded that factors such as the amount of time savings, service reliability and cost are highly influential amongst the people. National Aeronautics and Space Administration (NASA) also explored the market potential for UAM for three different use cases: transfer of packages (last-mile delivery), autonomous public commuter system (air metro) and autonomous ridesharing (air taxi). They found a majority of the users were comfortable with the use cases, but the logistics organizations could face potential operational and technical challenges in the form of travel distance, overall demand, and scheduling [35]. Another research by Swadesir and Bil [36] investigated the competitiveness of urban air transportation with automobiles, bikes and other modes of public transportation in Melbourne, in which they concluded that customers were mainly concerned about the safety aspects and noise generated by the aircraft.

2.2. Air taxi location decision

Studies pertaining to facility location decisions for air taxi vehicles are still in the developmental stage and are not extensively available in the literature. Rajendran and Zack [27] evaluated the potential demand for air taxi services based on regular taxi data and utilized a clustering algorithm with multimodal transportation-based warm start technique to determine prospective locations for air taxi stations in NYC. They also evaluated the impact of various parameters, such as passenger willingness to avail of air taxi services and demand fulfillment rate. They concluded that the logistics companies would need to establish 21 air taxi stations in NYC to fulfill an annual demand of approximately five million users. Rath and Chow [26] formulated a mathematical model to identify the required number of air taxi hubs between the three major airports in NYC and its neighborhood to reduce the overall travel cost. They determined that the city needed more than nine stations to achieve at least 10% market penetration.

Lim and Hwang [37] investigated the use of a clustering algorithm to generate vertiport centers in Seoul and considered the three heavily utilized routes in the city in their study. Bonnefoy [38] explored the use of simulation for estimating annual demand, fleet size, and network configuration for air taxis. A similar investigation was conducted by Rothfeld et al. [39], in which they used a transportation simulation tool to analyze the network and infrastructure placement by incorporating no-fly zones, required flight paths, and height restrictions in the model. Swadesir and Bil [36] compared the travel time taken by commonly used transportation modes and air taxis in Melbourne and determined that UAM saved more than 24 min on average over driving. They inferred that each vertiport in the city should be placed at least 10 km (approx. 6 miles) apart due to the speed requirements of the vehicles.

Ale-Ahmad and Mahmassani (2021) developed a mixed integer programming model to investigate the capacitated location, allocation and routing problem for UAM services. The model addressed critical decisions related to the effective operation and suggested that the number of cancellations and aerial mileage can be decreased significantly with demand consolidation. Fadhil [40] investigated the minimum requirements for establishing ground infrastructure for air taxi operations in major cities. Based on an extensive literature review, they observed various factors such as population density, office rental price, median income and job density, annual transport cost, and points of interest on the demand side and existing helipads and noise on the supply side impacting the decision-making process. After determining these parameters, they developed an analytical hierarchy process (AHP) – Delphi tool by integrating the two techniques to generate suitable weights for their location analysis.

Facility location problems that are studied in similar emerging technologies could be adopted for air taxi network design as well. For example, this strategic decision has been widely analyzed for establishing charging stations for electric cars [41–45] and last-mile delivery systems [46,47]. Liu and Wang [41] investigated the use of a heuristic algorithm to optimally locate recharging terminals by minimizing travel cost, time, and delay due to charging. Similarly, Wang et al. [43] developed a network to model optimal charging strategies for electric vehicles and formulated the location decision problem based on available budget and charging capacity at each station. In comparison, Lee et al. [48] developed a discrete event model to study the utilization of a hypothetical air taxi network that would aid in setting up business decisions such as adjusting fares while maximizing overall profits.

2.3. Clustering algorithms

Aggregation of collected data points into numerous groups based on their properties is known as clustering (Yao et al., 2019). There are several modes of operation proposed in the literature, such as partitioning algorithm, density-based algorithm, hierarchical algorithm, etc. (Mouton et al., 2020). Other than partitioning algorithms, most approaches automatically determine the ideal number of cluster centroids in the system [49]. In the partitioning algorithm, the number of focal points is specified by the user. The current study focuses on using the k-means partitioning algorithm as the number of clusters can be easily varied to satisfy the restrictions in the sensitivity analysis.

2.4. Contributions to the literature

This research provides a multifaceted contribution to the existing literature. We are one of the first to propose location recommendations for air taxi vertiports and vertistops in metropolitan cities based on a two-stage approach. Phase-1 introduces a novel multiple criteria warm start algorithm that identifies the potential candidates for air taxi skyport locations based on several conflicting socio-economic factors, such as rental cost, population density of an area, number of air taxi trips per day per 1000 customers, average salary in that neighborhood, and road facility. These parameters are conflicting in nature because the logistic companies would favor minimizing rental cost, average employee salary and road facility while maximizing population coverage and number of trips per day. Phase-2 adopts a constraint clustering approach to determine potential sites using the input from phase-1.

We leverage the estimated air taxi NYC demand data provided by Rajendran and Zack [27]. Rajendran and Zack [27] estimated the demand based on the assumption that there is only one passenger per ride. However, when analyzing the estimated air taxi demand data, the actual number of passengers per trip varies from one to five. Hence, the total number of estimated air taxi ride records considered in our study increased from 3.6 million to 6.4 million. It is proven by prior research that the size of the dataset has a significant impact on the clustering quality [50]. Therefore, the increase in data records to 6.4 million is expected to result in better facility locations.

3. Methodology

The current study integrates a multi-criteria warm start technique with an iterative k-means constraint clustering algorithm. Several studies have been conducted on the aggregation of geographical coordinates using cluster models [37,51–54]. However, these techniques are not feasible for our current research because of a few unique constraints that pertain to air taxi operations. For example, Holden and Goel [9] and Rajendran [55] impose restrictions, such as a limit on first and the last mile on-road travel distance and threshold demand satisfaction rate. To incorporate these unique restrictions associated with eVTOL air taxi system, an iterative k-means cluster algorithm is adopted in this research as it allows the decision-makers to input a specific number of centroids. This method minimizes the distance between each data point and its associated center, hence satisfying the first constraint. Also, the number of facility locations can be modified based on the fulfillment of the second criteria.

3.1. Phase 1: Multi-criteria warm start technique

One of the limitations associated with the k-means clustering algorithm is that it randomly creates the initial solution, which significantly impacts the effectiveness of the final solution [56]. Therefore, it is necessary to generate a high-quality initial seed solution and provide it as input to the k-means clustering algorithm. A previous study by Rajendran and Zack [27] utilized a multimodal warm start technique, which is based on customers selecting various other modes of transportation. However, it does not take into account the intangible socio-economic factors. To address this issue, the current investigation proposes the use of a novel multi-criteria based warm start (MCWS) technique for preliminary seed generation. The reason for using the MCWS technique is that multiple criteria models have been effectively and widely applied for other strategic site selection problems, such as airport, gas and subway station (e.g., [57–59]). Based on previous literature, five important criteria were selected for this approach [60–62]: rental cost, population density of an area, number of air taxi trips per day per 1000 customers, average salary in that neighborhood, and road facility. A hypothetical example illustrating the proposed approach is discussed in Appendix.

Suppose if $x_1, x_2, \ldots, x_n$ are the set of potential vertiport/vertistop locations, each site location l is defined by a tuple $T = \left(r_l, f_l, s_l, p_l, t_l\right)$, where r_l is the rental cost per month, f_l is the road facility, s_l is the average salary per year, p_l is the population coverage and t_l is the total number of estimated trips made by air taxi per day per 1000 population in that region. Rental cost is defined as the median gross rent per month for each neighborhood in which the station is located. Population coverage represents the total population of the area, while trips per day per 1000 population for each district are determined using the customer pickup coordinates and the station location data. Road facility and employment costs are specified as the distance of the proposed air taxi station with a major road and average household income per year, respectively.

We use the weighted average multi-criteria approach to rank the set of potential site locations. If the criterion m has to be maximized (e.g., road facility, population coverage, total number of estimated trips made per day per 1000 population), then the ideal value (τ_m) is the maximum value noted for all sites. In such a scenario, the normalized value ($N_{m,l}$) is computed by dividing $V_{m,l}$ with τ_m, as shown in Eq. (1). Similarly, if a criterion m in the tuple is supposed to be minimized (e.g., rental cost per month, average salary per year) then the optimal value of that criteria m (τ_m) is the minimum value observed across all the stations. The normalized value of criteria m for site location l ($N_{m,l}$) is calculated by dividing τ_m with values of the measure ($V_{m,l}$), as shown by Eq. (2).

Once the normalized values for each criterion are calculated, it is then necessary to obtain the weight (W_m), which indicates the order of importance of each objective, and is usually acquired from the decision-maker. The overall fitness value of each center and the total score for all locations are given by Eqs. (3) and (4), respectively.

$$N_{m,l} = \frac{V_{m,l}}{\tau_m} \tag{1}$$

$$N_{m,l} = \frac{\tau_m}{V_{m,l}} \tag{2}$$

$$F_l = \sum_{m=1}^{M} W_m \times N_{m,l} \tag{3}$$

$$TS = \sum_{l=1}^{n} F_l \tag{4}$$

3.2. Phase-2: k-means clustering with MCWS technique

The output from the multiple criteria warm start technique (i.e., Phase-1) is provided as the input to Phase-2, as the seed solution. The goal of the algorithm involved in this phase is to minimize the average squared Euclidean distance between z data points and n cluster centers [63]. Let x_l be the center of cluster l, as shown in Eq. (5). According to Bock [63], the variance is given by Eq. (6). The objective is to minimize the sum of squares for all data points.

$$x_l = \frac{1}{|z|} \sum_{x \epsilon z} \vec{x} \tag{5}$$

$$Q = \sum_{\vec{x} \in z} |\vec{x} - x_l|^2 \rightarrow min.\ Q \tag{6}$$

4. Data description

The dataset used in this study is the estimated air taxi demand data used by Rajendran and Zack [27]. They leveraged the publicly available taxi records from the New York City Taxi and Limousine Commission database for a period of two years starting from January 2014 for demand estimation. Trip records from 2016 onwards are not considered in the present investigation because the government anonymized the pickup and dropoff locations due to privacy concerns, thus leading to the lack of accurate location specifics. Each data point consists of significant parameters, such as pickup and dropoff coordinates, date, time, total distance traveled in miles, and the total number of passengers on each trip. After pre-processing the data, Rajendran and Zack [27] estimated the potential demand for the air taxi service based on the constraints mentioned in Section 3, along with the assumption that a customer is eligible for eVTOL service only if they save at least 40% ride time when compared with the on-road travel [9]. As indicated earlier, the data estimated by Rajendran and Zack [27] was under the assumption that there is one customer in each ride, which is circumvented in this study.

Fig. 2 depicts the potential geo-mapping of customers on weekdays and weekends during the morning time period (9:00 AM–12:00 PM). It can be noted that both during weekdays and weekends, there are

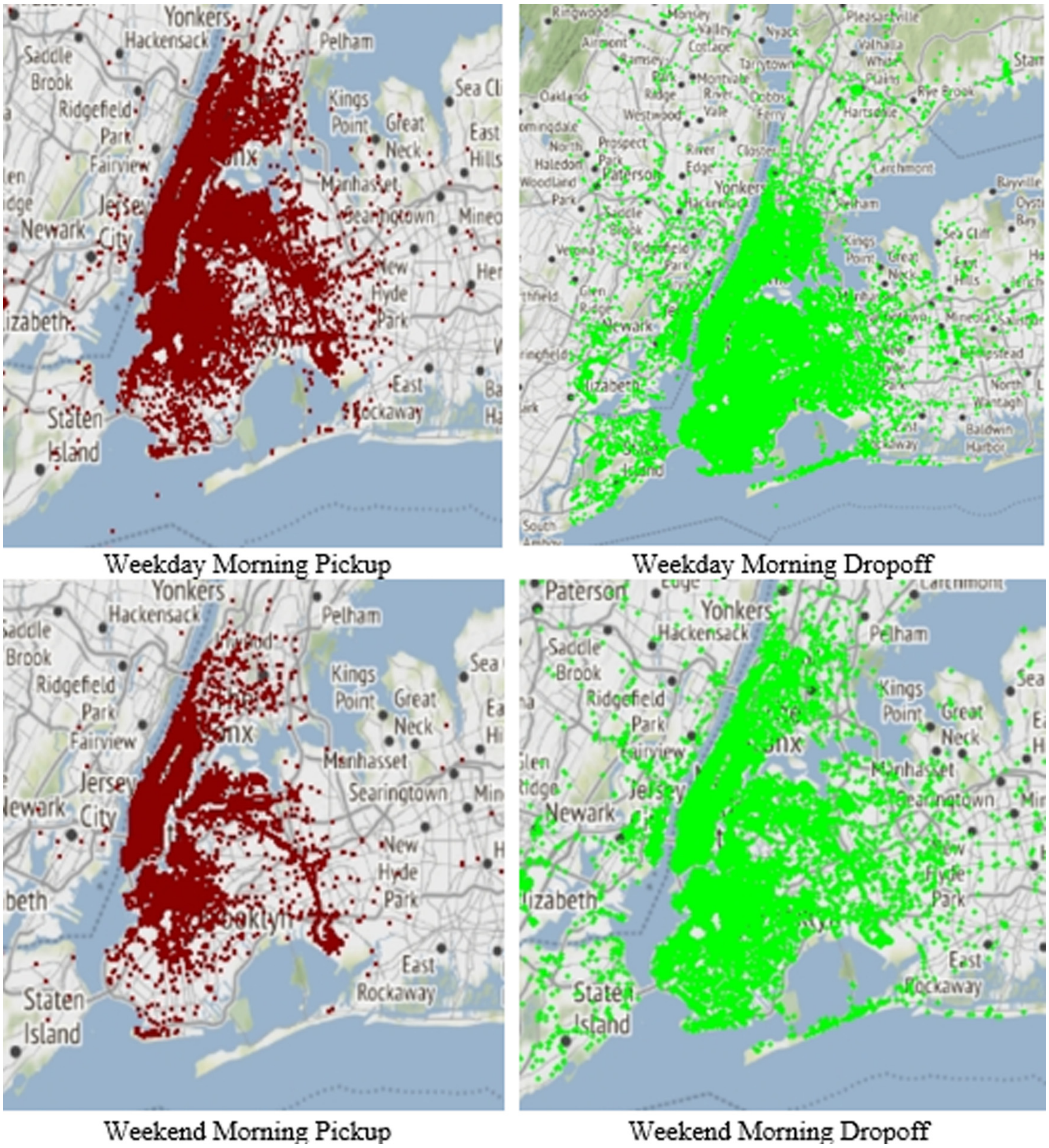

Fig. 2. Geospatial mapping of potential customers on weekdays and weekends.

more dropoffs compared to pickups. This is expected because many commuters travel to the city from suburban regions for work. We can also observe that the ride pickups near Queens during weekends are significantly lower when compared with weekdays.

5. Results and discussion

Based on the approach discussed in the previous section, we evaluate the locations for vertiports and vertistops for the baseline scenario, which is based on having at least 40% time savings compared with ground transportation. Furthermore, it comprises of satisfying at least 70% of the total demand with a one-mile on-road travel limitation for the first and last leg, respectively. Sensitivity analysis is then conducted by altering the percentage values of the four key parameters (time savings, on-road travel limit, passenger willingness to fly and demand satisfaction). The results of the current investigation are compared with the locations obtained by running the model developed by Rajendran and Zack [27] for our dataset.

5.1. Baseline results

The MCWS technique is first used to generate the list of potential seed solutions for the iterative clustering algorithm. There are 59 community districts distributed amongst the five major boroughs (Manhattan, Brooklyn, Queens, Bronx, and Staten Island) in NYC. As discussed earlier, the goal of the MCWS technique is to identify the best tradeoff solutions based on the five conflicting criteria: rental cost, population density of an area, number of trips per day per 1000 customers, average salary and road facility. Based on the procedure discussed in Section 3.1 and the recommended weights for each parameter provided in the literature [60–62], the total score is computed, and the best five sites with the maximum weighted values are used as initial input seed in the k-means algorithm.

The ideal number of stations generated by our model for the base case setting is 18 when compared to 20 stations suggested by the model proposed by Rajendran and Zack [27]. Six (#3, #9, #12, #15, #16 and #17) of those sites are proposed to be built in Manhattan. Sites #9 and #12 are approximately a mile apart. It is observed that facility

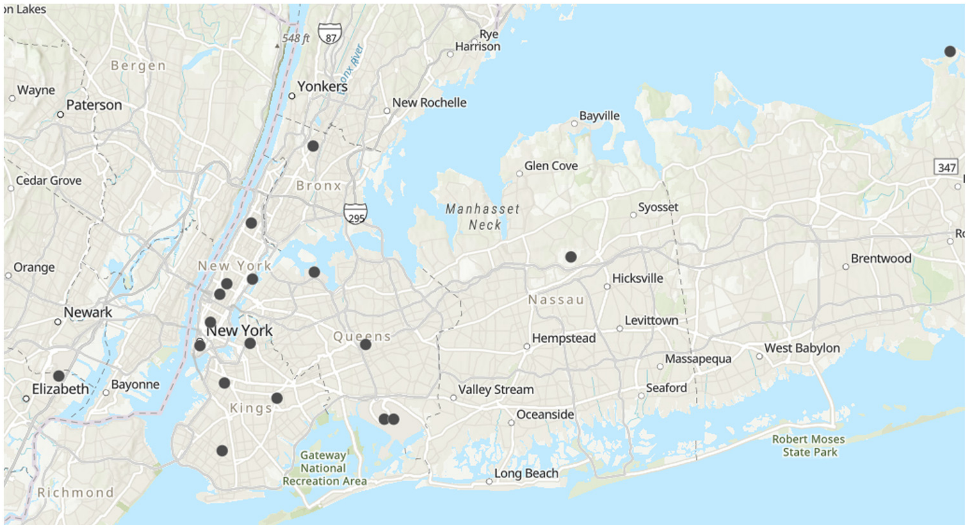

Fig. 3. Recommended location of the air taxi facility sites in NYC.

#12, near Times Square, serves over 12% of potential commuters while station #9 near South Central Park has approximately 10% customer demand. Other boroughs with a high volume of travelers are Queens and Brooklyn. A 2D view of the suggested potential locations is given in Fig. 3.

We observe that two air taxi stations (#6 and #11) are located in the John F. Kennedy (JFK) International Airport and one each (#4 and #14 respectively) in Newark Liberty International Airport and LaGuardia Airport (LAG). JFK and LAG cater to over 50% of the customer demand in total. Thus, it is recommended to build one large vertiport at each of these locations. Other locations that experience a high volume of traffic are sites #3 (near World Trade Center) and #17 (close to Empire State Building). It is seen that Columbia University is in close proximity to site #16. In addition, site #15 is located between Roosevelt Island and the Upper Eastside in Manhattan and has only about 2% of the potential customers. Nevertheless, it would be challenging to build infrastructure at site #15 since this facility has to be set up on the island to serve its customers. Despite the model recommending the construction of stations at sites #2 and #13, from a practical standpoint, it is not recommended since these facilities contribute to less than 0.5% of the demand. Our algorithm does not suggest major locations such as Yankee Stadium and Washington Square Arch, which is counter intuitive.

Table 2 shows the comparison of the locations reported by the current study and Rajendran and Zack [27]. It is to be noted that four of the 18 locations proposed in this study overlap with the prior literature. Further, we can see that eight facilities are within one-mile distance apart, whereas our results propose new infrastructure recommendations for the remaining number of sites.

Table 2
Comparison of the locations obtained by current study and Rajendran and Zack [27].

Location	Current Paper	Rajendran and Zack [27]
Briarwood, Queens (site #1)	✔	✘
Long Island Sound (site #2)	✔	✘
Vesey Street, Lower Manhattan (site #3)	✔	•
Newark Liberty International Airport (site #4)	✔	✔
61st Street, Brooklyn (site #5)	✔	•
JFK International Airport (repeated twice) (site #6 and #11)	✔	✔
Douglass Street, Brooklyn (site #7)	✔	•
Woodland, Bronx (site #8)	✔	✘
South Central Park (site #9)	✔	✔
Grafton Street, Brooklyn (site #10)	✔	•
40th Street, near Times Square (site #12)	✔	•
Jericho Union District (site #13)	✔	✘
LaGuardia Airport (site #14)	✔	✔
Roosevelt Island (site #15)	✔	✘
West Harlem (site #16)	✔	•
5th avenue, Midtown Manhattan (site #17)	✔	•
Lorimer Street, Brooklyn (site #18)	✔	•

Note: ✔ means the location is present in the study. ✘ means location is not present in the study. • means the location reported by Rajendran and Zack [27] and is within a 1-mile radius of the site in the current study.

5.2. Sensitivity analysis

In this section, we examine the influence of the four key input parameters appertaining to the performance of the model used in the current study. Table 4 presents a comparison of the results generated in all the cases with those obtained by running the models proposed by Rajendran and Zack [27] and the traditional k-means algorithm for our dataset.

5.2.1. Time savings (TS)

In accordance with the assumptions made in Section 3, air taxi services will only be availed by a customer if there is a time saving of at least 40% compared to ground transportation [9,55]. The performance of the proposed approach is investigated by linearly varying the time savings (TS - 1 to TS - 4) percentage (Table 3). It is observed that the change in TS does not impact the number of facilities. Further, when compared to the number of sites obtained by the model proposed by Rajendran and Zack [27], we find that the proposed model performs better for all the TS settings with a percentage deviation of 11%, 14%, 15% and 10% for TS-1–TS-4, respectively.

In comparison with the locations proposed by the baseline setting, it is seen that site #16 shifts 1.5 miles south from the original location in TS-3 and 1.5 miles north in TS-4. Similarly, site #12 moves in the north direction by a mile in the new settings closer towards South Central Park. The new location fulfills 18% of total demand compared with the initial vertistop serving nearly 12% of potential customers. The location of the initial facility suggested for South Central Park (site

Table 3
Varying various parameters for sensitivity analysis.

Case	Time savings (TS) in %	Passenger willingness to fly rate (PR) in %	On-road travel limit (RL) in miles	Demand Satisfaction (DS) in %
TS-1	30	100	1	70
TS-2 (Base)	40	100	1	70
TS-3	50	100	1	70
TS-4	60	100	1	70
PR-1 (Base)	40	100	1	70
PR-2	40	90	1	70
PR-3	40	80	1	70
PR-4	40	70	1	70
RL-1	40	100	0.5	70
RL-2 (Base)	40	100	1	70
RL-3	40	100	1.5	70
DS-1	40	100	1	60
DS-2 (Base)	40	100	1	70
DS-3	40	100	1	80
DS-4	40	100	1	90

#9) moved across in a similar direction by approximately 2.5 miles. The recommended number of stations for various scenarios is shown in Fig. 4.

5.2.2. Passenger willingness to fly rate (PR)

Aviation safety has been proclaimed as a crucial factor by commuters in the literature [32,36]. Therefore, a certain proportion of users might be hesitant to utilize the air taxi service [64]. In the baseline scenario, it is assumed that 100% of the eligible riders would be willing to fly in the air taxis. In this analysis, the passenger willingness to fly rate is decreased from 100% to 70% in steps of 10%. The corresponding number of eVTOL stations achieved using the algorithm is shown in Table 3 (PR-1 to PR - 4). Across various cases, the number of facilities is found to be almost identical. However, compared to the number of sites obtained by the model proposed by Rajendran and Zack [27], we find that the developed approach performs better for all the PR settings with a percentage deviation of 10%, 15%, 10.53%, and 15.79% for PR-1–PR-4, respectively.

Unexpectedly, it is observed that the number of commuters traveling from Brooklyn is reduced by nearly one-third for the PR-2 setting while remaining the same for the other two scenarios. It is also suggested to build vertiports near Washington Square Arch and Midtown Manhattan (one mile from Empire State Building), which are not indicated in the baseline model, and provide service to over 7% and 10% customers in PR-2 and PR-4, respectively. Fig. 5 depicts the station locations for different PR settings.

5.2.3. On-road travel limit (RL)

As mentioned earlier, in the baseline setting, on-road travel distance was limited to one mile for the first and the last legs. In this section, this parameter is altered linearly from 0.5 miles to 1.5 miles, as shown in Table 3. As expected, the number of sites increased by approximately 70% for the first scenario (RL-1), relative to the baseline setting, and reduced by 50% for the last case (RL-3). Furthermore, a comparison with Rajendran and Zack [27] showcased a 4.92%, 14.29% and 10.00% deviation in the number of air taxi stations in the current study for settings RL-1–RL-3.

It is noticed that the majority of the demand (approximately 60%) is fulfilled by only ten facilities (18% of total sites) in RL-1. Other newly suggested potential centers are Staten Islands, Jamaica, and Madison Square Garden. Furthermore, in RL-3, site #12 is relocated by a mile north, where the number of users is doubled compared with the baseline case. Thus, it is recommended to build a large vertiport to cater to this extra demand, as showcased in Fig. 6.

5.2.4. Demand satisfaction (DS)

In the baseline case, we assumed that 70% of the total customers would be eligible to avail of the air taxi services [64]. In this analysis, the demand satisfaction rate is varied from 60% to 90%, incrementing by 10% for different settings, as shown in Table 3 (DS-1 to DS - 4). A linear increase in DS percentage generates an exponential growth in the number of locations. A similar trend was indicated by the model proposed by Rajendran and Zack [27] as well. It is seen that the number of locations outside the five major boroughs also increases, and it is recommended that the logistics company conducts a market study to determine the feasibility of establishing these stations. Fig. 7 depicts the recommended facility locations for all the scenarios investigated in this subsection.

The number of infrastructure sites rises by nearly 75% for DS-4 from the base case in the present investigation, comparable to the previously reported literature. Almost three-quarters of the total demand is fulfilled by 25% of sites in DS-3 and DS-4 settings. For DS-1, it is proposed to develop a vertistop in a facility close to site #12 (Times Square) as the number of customers is observed to double, which is similar to RL-3. The total number of eligible riders increased by 29% from DS-1 to DS-2. However, only 13% of the growth was observed when the commuter satisfaction level was increased from DS-2 to DS-3. A further increase to the subsequent scenario noted only an 11% rise in the overall number of rides.

5.3. Comparison of results of the proposed method with prior methods

We consider two metrics to evaluate the performance of the proposed approach. The first is the Davies Bouldin index (DBI). DBI determines the performance of clustering by calculating the ratio of the total spread of points within clusters and the distance between each cluster center. Therefore, the lower value of the metric is preferred. The second metric is the number of stations. The goal is to minimize the number of operating stations (thereby reducing the total operating cost) while maintaining a threshold demand fulfillment rate. Hence, the lower value of both metrics is preferred.

5.3.1. Evaluation using the DBI and number of clusters

Based on Table 4, it can be noticed that both the performance measures are observed to be lower under the present study than both the traditional k-means algorithm as well as the approach proposed by Rajendran and Zack [27]. This clearly indicates that the proposed method outperforms the existing air taxi facility location approaches discussed in the literature.

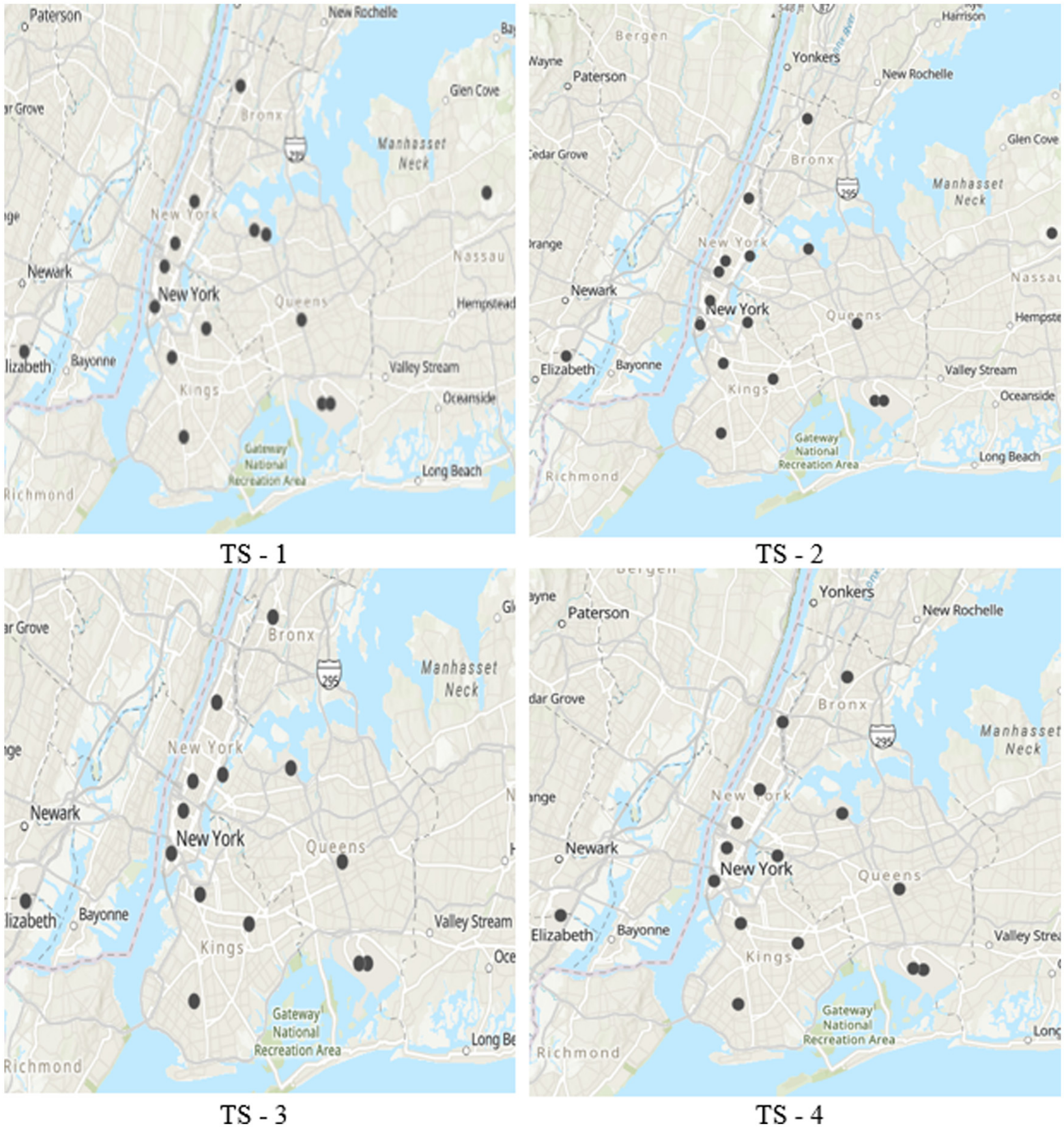

Fig. 4. Infrastructure locations for various time-saving scenarios.

5.3.2. *Evaluation using the multi-criteria decision-making approach*

While Section 5.3.1 evaluates the effectiveness of the proposed approach theoretically, this section compares the model results from a practical standpoint. Based on the five criteria discussed in Section 3.1 (i.e., rental cost per month, road facility, average salary per year, population coverage, and the total number of estimated trips made by air taxi per day per 1000 population), this section analyzes the overall ranking of the location insights presented in our study and prior studies. Traditionally, the decision-maker is involved in providing the weights for each criterion. Since air taxi is still in the developmental stage, we used the weights suggested by Hawas et al. [60], Tzeng et al. [61,62]. The scaled weights for rental cost, population coverage, trips per day per 1000 population, road facility, and employment cost are 0.14, 0.39, 0.26, 0.06, and 0.15, respectively. These are used for the calculation of the weighted average for each location and the overall fitness values.

Tables 5 and 6 provide the values for each criterion of all the proposed sites under the current research as well as by Rajendran and Zack [27]. The data are obtained from multiple publicly available government sources (Department of city planning, NYU Furman Center, and United States census bureau). These data are normalized based on Eqs. (1) and (2), and then the weighted score for each station location is computed using Eq. (3). Based on Eq. (4), the total score per site of 0.45 is obtained for our suggested air taxi stations, which is higher than the score calculated for the location obtained using the model proposed by Rajendran and Zack [27], which is 0.43. This implies that the two-phase technique implemented in the present study provides a better solution than the existing literature.

5.4. *Implications and recommendations*

It is proven by prior research that the size of the dataset has a significant impact on the clustering quality [50]. Therefore, the increase in data records to 6.4 million in our study is expected to result in better facility locations. Even though the results obtained in the two studies that are compared in the paper are for NYC, several generic practical implications can be made to extend this for other cities as well.

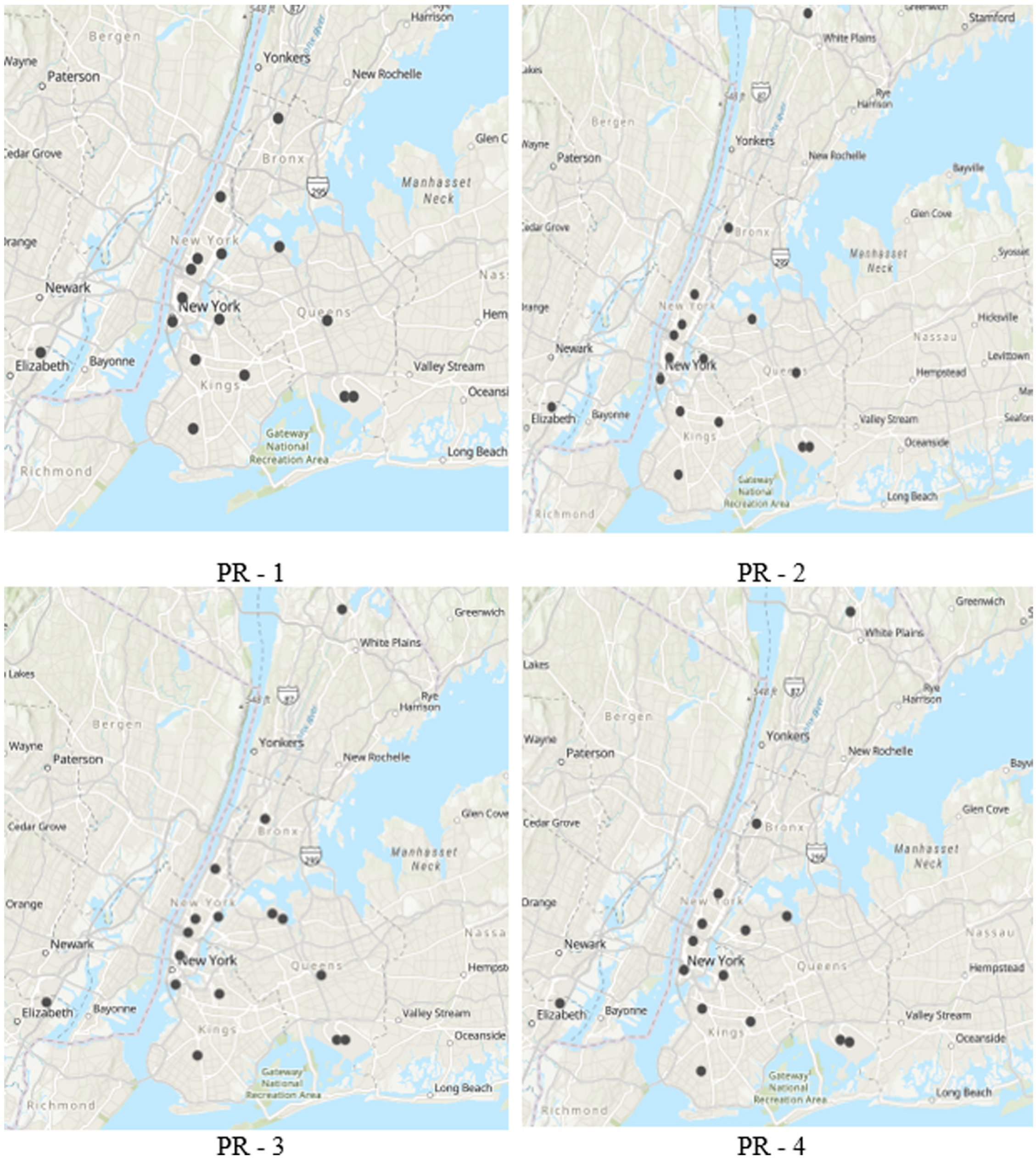

Fig. 5. Infrastructure locations of various passenger willingness to fly rate scenario.

Table 4
Comparison of performance metrics obtained from the current research with prior studies.

Scenario	Number of stations			Davies–Bouldin Index		
	Current study	% Deviation from [27]	% Deviation from traditional k-means	Current study	% Deviation from [27]	% Deviation from traditional k-means
TS-1	16	11.11	15.79	0.39	4.88	7.14
TS-2 (base)	18	14.29	18.18	0.36	12.20	18.18
TS-3	17	10.53	15.00	0.4	4.76	−2.56
TS-4	18	10.00	14.29	0.39	7.14	7.14
PR-1 (base)	18	10.00	18.18	0.36	20.00	21.74
PR-2	17	15.00	15.00	0.38	13.64	15.56
PR-3	17	10.53	10.53	0.42	2.33	2.33
PR-4	16	15.79	23.81	0.39	11.36	2.50
RL-1	58	4.92	6.45	0.22	61.40	69.01
RL-2 (base)	18	14.29	21.74	0.36	20.00	21.74
RL-3	9	10.00	35.71	0.57	−46.15	−42.50
DS-1	12	7.69	20.00	0.42	−2.44	2.33
DS-2 (base)	18	14.29	18.18	0.36	20.00	21.74
DS-3	30	9.09	0.00	0.4	34.43	35.48
DS-4	78	4.88	4.88	0.33	52.17	60.71

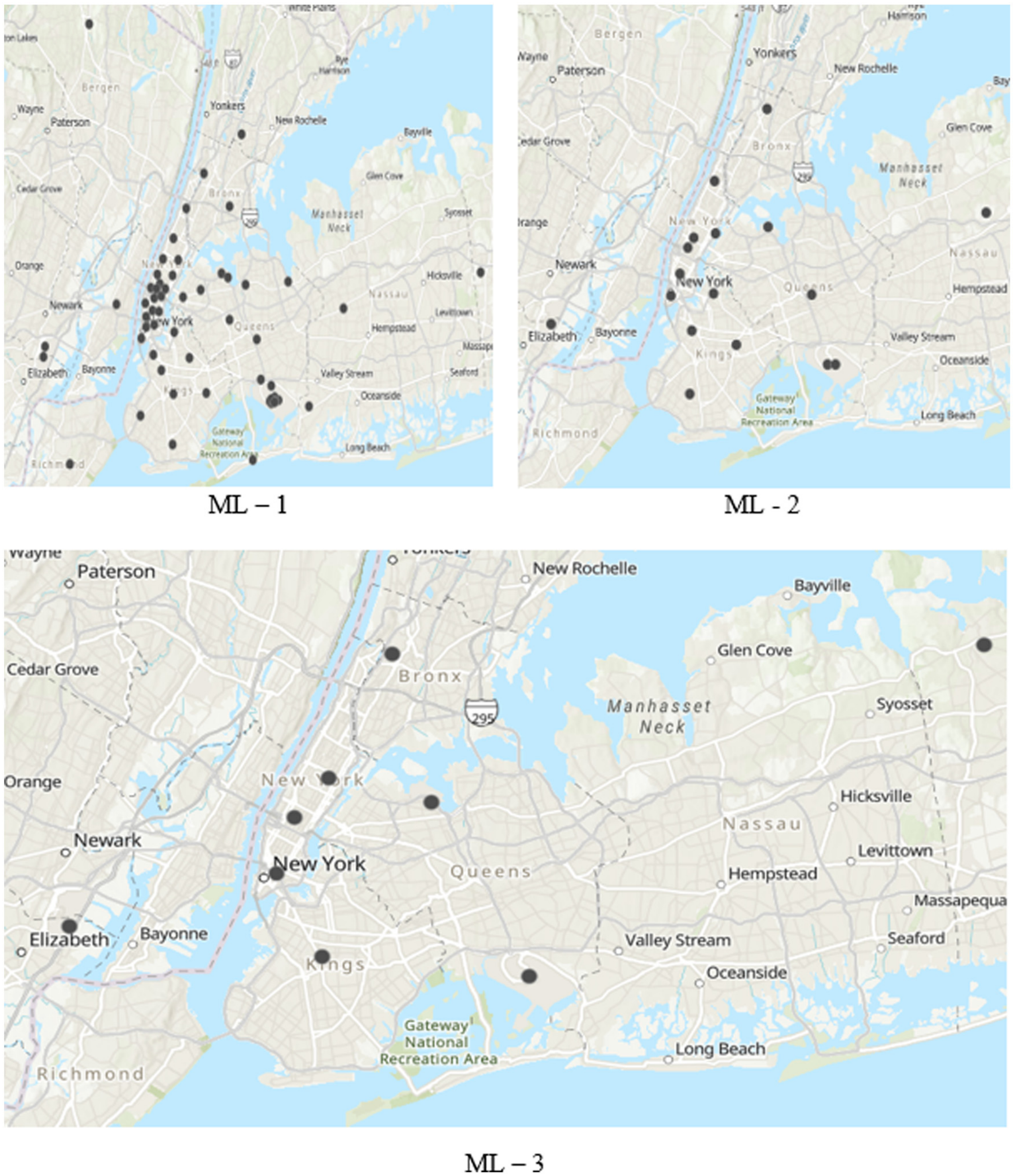

Fig. 6. Infrastructure locations for different on-road travel limit scenario.

First, urban areas having international airports are expected to have a high volume of customers using air transport and thus require huge vertistops to be built to satisfy the demand. Second, it is anticipated that popular tourist sites and downtown areas of a city would be the next highest in terms of proportions of commuters served and require a large number of air taxi vehicles at their respective stations. Third, varying percentages of time savings and passenger willingness to fly rate have no substantial effect on the total number of facilities. Lastly, linear change in on-road travel limit and customer demand satisfaction levels causes an increase in the number of customers assisted, and therefore, an exponential rise in the number of urban air facilities in the city.

Based on the results and observations, the following managerial recommendations are proposed:

1. It is essential to launch a large vertiport at JFK, LAG, and New York Liberty International Airports since they cater to a combined 60% of the total demand.
2. It is observed that the average number of dropoffs is greater than the average number of pickups in Brooklyn, Queens, Bronx, and Staten Island. Therefore, it is recommended to build smaller vertiports in these areas and the air taxi vehicles to be routed back to Manhattan for pickups.
3. Sites serving South Central Park and West 40th Street (0.2 miles from Times Square) are approximately a mile apart. Therefore, a common air taxi station can be developed catering to the demand from both locations.
4. Midtown and Lower Manhattan experience a high volume of commuters due to popular tourist attractions such as the Empire State Building, Madame Tussauds, World Trade Center, etc. Thus, we propose a high fraction of the fleet to serve the two areas.

Fig. 7. Infrastructure locations for demand satisfaction scenario.

Table 5
Values for different criteria for locations obtained in the current study.

Site #	Rental cost ($/month)	Population coverage	Trips per day per 1000 population	Road facility (miles)	Employment cost ($/year)
1	1800	165,000	0.0281	1	65,000
2	1420	12,000	0	2.1	43,000
3	2610	148,000	0.8102	0.009	147,641
4	1140	128,000	0.1417	0.001	47,000
5	1450	146,000	0.0223	0.2	46,229
6	1590	139,000	0.1634	0.001	75,300
7	2280	116,000	0.2886	0.7	137,000
8	1410	150,000	0.0101	0.6	60,000
9	2150	153,000	1.8109	0.2	104,000
10	900	111,500	0.0338	0.001	20,640
11	1590	139,000	2.1713	0.001	75,300
12	2150	153,000	2.3838	0.07	104,000
13	2044	14,000	0.000106	3.2	161,771
14	1530	170,000	1.4878	0.001	58,000
15	2380	215,000	0.3156	1.4	134,000
16	1280	136,000	0.0939	1.5	51,000
17	2610	149,000	1.0350	0.4	147,600
18	2980	152,000	0.1920	0.5	78,000

Table 6
Values of various criteria for locations obtained by model discussed in [27].

Site #	Rental cost ($/month)	Population coverage	Trips per day per 1000 population	Road facility (miles)	Employment cost ($/year)
1	2490	149,000	0.49	0.001	114,500
2	1777	15,000	0.01	0.28	142,928
3	1140	128,000	0.07	0.001	47,000
4	2610	148,000	0.36	0.001	147,641
5	1590	139,000	1.19	0.001	75,300
6	1300	120,000	0.02	0.25	49,000
7	1590	139,000	0.31	0.001	75,300
8	1670	164,000	0.06	0.5	67,670
9	1690	7900	0.00	1.3	163,795
10	2610	149,000	0.78	0.001	147,600
11	1150	152,000	0.02	0.08	71,200
12	1450	146,000	0.01	0.3	46,229
13	900	111,000	0.01	0.001	20,640
14	910	128,000	0.25	0.1	37,500
15	1300	219,000	0.03	0.08	57,500
16	2280	116,000	0.11	0.67	137,000
17	1838	15,000	0.00	0.1	98,065
18	2980	152,000	0.12	0.8	78,000
19	2610	149,000	0.90	0.001	147,600
20	1530	170,000	0.69	0.001	58,000

5. The two closest locations in the Brooklyn borough are 61st Street and Grafton, and thus, a logistics company might choose to have only one infrastructure site considering a low demand.
6. The most common locations observed across all the cases are JFK International Airport, LaGuardia Airport, Newark Liberty International Airport, Times Square, Central Park, Empire State Building, and 61st Street, Brooklyn. Therefore, it is recommended to begin services in at least these places, if the logistics company has operational restrictions.
7. The model found a potential to build 16 air taxi stations in NYC, with two infrastructures being in the suburban neighborhood near Long Island Sound and Jericho Union District.
8. When compared with the locations obtained using the model discussed by Rajendran and Zack [27], the present study recommends two unique skyports to be developed in Manhattan and one each in Queens and Bronx.
9. Only 3% of total customers utilize the stations in Brooklyn; therefore, four smaller vertistops are suggested to be built in the region.
10. It is recommended to build one small vertistop in Roosevelt Island subsequent to the rise of latent demand.

6. Conclusions

The present study proposes air taxi vertiport and vertistop location decisions in metropolitan cities using a two-phase approach. Based on the estimated air taxi demand in New York City (NYC), the potential sites are identified within the town and its neighborhood by coupling the multi-criteria warm start technique with an iterative k-means clustering algorithm. Our study reports 18 site locations, which is less than the number reported in the existing literature. It is also observed that approximately three-fifths of the demand is based on the three major airports located in the city.

Parameters, such as percentage of time savings and passenger willingness to fly rate, have a negligible impact on the number of sites, whereas on-road travel limit and percentage of customer demand satisfaction have an exponential effect. Therefore, it is essential for logistics companies to take the implications of these factors into account. Also, if the transportation company favors beginning with limited service, then they can consider establishing their operation from seven common locations (JFK International Airport, Laguardia Airport, Newark International Airport, Times Square, Central Park, Empire State Building and 61st Street, Brooklyn). The performance of the developed method is theoretically evaluated by computing the Davies–Bouldin Index (DBI) and the number of clusters. Both these measures obtained in the current research are lower than those obtained by running the models proposed by Rajendran and Zack [27] and the traditional k-means algorithms. To evaluate the solution from a practical standpoint, we utilize the multi-criteria decision-making technique to compute the total score of the sites suggested in the present study as well as in the previous investigation. It is noted that the total score per facility for this research is higher, indicating that the proposed model is more efficient than the traditional k-means and Rajendran and Zack [27].

The initial seed solution of the k-means clustering algorithm has a significant impact on the effectiveness of the final solution. Therefore, it is necessary to generate a high-quality initial seed solution and provide it as input to the k-means clustering algorithm. The multi-criteria based warm start technique used in our study for preliminary seed generation considers several conflicting socio-economic factors, such as rental cost, population density of an area, number of air taxi trips per day per 1000 customers, average salary in that neighborhood, and road facility. This proposed warm start approach results in infrastructure facility locations that yield better performance metrics compared to the existing models in the literature.

However, there are certain limitations in the present research. The current investigation does not consider tactical and operational level decisions for the effective functioning of the facility. Therefore, a study on policies involving scheduling, routing, number of vehicles at each location, etc., can be considered as a potential area of improvement. Moreover, it is assumed that a traveler is eligible to avail of the services if only they have at least 40%-time savings. However, a customer's mode choice could be more complicated. Future studies could involve examining other factors, such as cost and wait time, in order to determine overall demand. Also, a mathematical or simulation model could be developed to aid the logistic companies in deciding the set of sites to be developed in multiple phases to mitigate the impact of unknown variables such as community acceptance and safety records.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix. Hypothetical numerical example to illustrate the proposed approach

In this section, we use a hypothetical example illustrating the proposed approach.

Table A.1
Hypothetical example to illustrate MCWS approach.

Facility location l	Rental cost r_l ($/month)	Population coverage p_l	Trips per day per 1000 population t_l	Road facility f_l (miles)	Employment cost s_l ($/year)
1	2610	148 000	0.8102	0.009	147 641
2	2150	153 000	2.3838	0.07	104 000
3	1590	139 000	0.1634	0.001	75 300
4	2980	152 000	0.1920	0.5	78 000
5	1410	146 000	0.0223	0.2	46 229

A.1. Numerical example for illustrating the proposed MCWS technique

Table A.1 presents hypothetical values of five different locations. The values of rental cost and average salary are normalized based on Eq. (2) are demonstrated through the set of Eqs. (A.1)–(A.5) and (A.6)–(A.10). Similarly, the values for population coverage, number of trips per day per 1000 customers and road facility are maximized and represented by Eqs. (A.11)–(A.15), (A.16)–(A.20) and (A.21)–(A.25), respectively. Finally, the fitness value for each center is computed using Eq. (3). It can be observed that location#2 has the highest fitness value showcasing that it has the greater priority when compared with other stations. Finally, the total score of all five centers is computed as shown by Eq. (A.31).

$$N_{r,1} = \frac{V_{r,1}}{\tau_r} = \frac{1410}{2610} = 0.54 \tag{A.1}$$

$$N_{r,2} = \frac{V_{r,2}}{\tau_r} = \frac{1410}{2150} = 0.65 \tag{A.2}$$

$$N_{r,3} = \frac{V_{r,3}}{\tau_r} = \frac{1410}{1590} = 0.88 \tag{A.3}$$

$$N_{r,4} = \frac{V_{r,4}}{\tau_r} = \frac{1410}{2980} = 0.47 \tag{A.4}$$

$$N_{r,5} = \frac{V_{r,5}}{\tau_r} = \frac{1410}{1410} = 1.00 \tag{A.5}$$

$$N_{s,1} = \frac{V_{s,1}}{\tau_s} = \frac{46\,229}{147\,641} = 0.31 \tag{A.6}$$

$$N_{s,2} = \frac{V_{s,2}}{\tau_s} = \frac{46\,229}{10\,400} = 0.44 \tag{A.7}$$

$$N_{s,3} = \frac{V_{s,3}}{\tau_s} = \frac{46\,229}{75\,300} = 0.61 \tag{A.8}$$

$$N_{s,4} = \frac{V_{s,4}}{\tau_s} = \frac{46\,229}{78\,000} = 0.59 \tag{A.9}$$

$$N_{s,5} = \frac{V_{5,1}}{\tau_s} = \frac{46\,229}{46\,229} = 1.00 \tag{A.10}$$

$$N_{p,1} = \frac{\tau_p}{V_{p,1}} = \frac{148\,000}{153\,000} = 0.96 \tag{A.11}$$

$$N_{p,2} = \frac{\tau_p}{V_{p,2}} = \frac{153\,000}{153\,000} = 1.00 \tag{A.12}$$

$$N_{p,3} = \frac{\tau_p}{V_{p,3}} = \frac{139\,000}{153\,000} = 0.90 \tag{A.13}$$

$$N_{p,4} = \frac{\tau_p}{V_{p,4}} = \frac{152\,000}{153\,000} = 0.99 \tag{A.14}$$

$$N_{p,5} = \frac{\tau_p}{V_{p,5}} = \frac{146\,000}{153\,000} = 0.95 \tag{A.15}$$

$$N_{t,1} = \frac{\tau_t}{V_{t,1}} = \frac{0.8\,102}{2.38} = 0.34 \tag{A.16}$$

$$N_{t,2} = \frac{\tau_t}{V_{t,2}} = \frac{2.38}{2.38} = 1.00 \tag{A.17}$$

$$N_{t,3} = \frac{\tau_t}{V_{t,3}} = \frac{0.1634}{2.38} = 0.07 \tag{A.18}$$

$$N_{t,4} = \frac{\tau_t}{V_{t,4}} = \frac{0.1920}{2.38} = 0.08 \tag{A.19}$$

$$N_{t,5} = \frac{\tau_t}{V_{t,5}} = \frac{0.0223}{2.38} = 0.009 \tag{A.20}$$

Table A.2
Hypothetical example to illustrate DBI calculation.

Air taxi trip record	Cluster #1	
	Latitude	Longitude
1	40.7525	−73.9802
2	40.7617	−73.9792
	Cluster #2	
	Latitude	Longitude
3	40.7636	−73.9813
4	40.7578	−73.9842
5	40.7457	−73.9876

$$N_{f,1} = \frac{\tau_f}{V_{f,1}} = \frac{0.009}{0.5} = 0.02 \tag{A.21}$$

$$N_{f,2} = \frac{\tau_f}{V_{f,2}} = \frac{0.07}{0.5} = 0.14 \tag{A.22}$$

$$N_{f,3} = \frac{\tau_f}{V_{f,3}} = \frac{0.001}{0.5} = 0.002 \tag{A.23}$$

$$N_{f,4} = \frac{\tau_f}{V_{f,4}} = \frac{0.5}{0.5} = 1.00 \tag{A.24}$$

$$N_{f,5} = \frac{\tau_f}{V_{f,5}} = \frac{0.2}{0.5} = 0.40 \tag{A.25}$$

$$F_1 = (0.14 \times 0.54) + (0.15 \times 0.31) + (0.39 \times 0.96) + (0.26 \times 0.34) + (0.06 \times 0.02) = 0.5861 \tag{A.26}$$

$$F_2 = (0.14 \times 0.65) + (0.15 \times 0.44) + (0.39 \times 1.00) + (0.26 \times 1.00) + (0.06 \times 0.14) = 0.8154 \tag{A.27}$$

$$F_3 = (0.14 \times 0.88) + (0.15 \times 0.61) + (0.39 \times 0.90) + (0.26 \times 0.07) + (0.06 \times 0.002) = 0.4731 \tag{A.28}$$

$$F_4 = (0.14 \times 0.47) + (0.15 \times 0.59) + (0.39 \times 0.99) + (0.26 \times 0.08) + (0.06 \times 1.00) = 0.6212 \tag{A.29}$$

$$F_5 = (0.14 \times 1.00) + (0.15 \times 1.00) + (0.39 \times 0.95) + (0.26 \times 0.009) + (0.06 \times 0.40) = 0.6868 \tag{A.30}$$

$$\text{TS} = 0.5861 + 0.8154 + 0.4731 + 0.6212 + 0.6868 = 3.1826 \tag{A.31}$$

A.2. Numerical example for calculating DBI

Let us assume that we have two clusters with two and three data records, respectively, as shown in Table A.2. Suppose the coordinates (40.7576, −73.9799) and (40.7470, −73.9836) are center of cluster#1 and cluster#2, respectively, that are generated by the proposed two-phase clustering approach. The coordinates of the points are presented in Table A.2.

We first compute the within cluster spread for both the clusters.

$$hav\left(\theta_{1,2}\right) = r \times d_{1,2}$$

$$= hav\,(40.7617 - 40.7525) + \cos(40.7617) \times \cos(40.7525) \times hav\,(-73.9792 - (-73.9802)) = 0.66\,\text{miles}$$

$$hav\left(\theta_{3,4}\right) = r \times d_{3,4} = hav\,(40.7578 - 40.7636) + \cos(40.7578) \times \cos(40.7636) \times hav\,(-73.9842 - (-73.9813)) = 0.51\,\text{miles}$$

$$hav\left(\theta_{4,5}\right) = r \times d_{4,5} = hav\,(40.7457 - 40.7578) + \cos(40.7457) \times \cos(40.7578) \times hav\,(-73.9876 - (-73.9842)) = 0.84\,\text{miles}$$

$$hav\left(\theta_{3,5}\right) = r \times d_{3,5} = hav\,(40.7457 - 40.7636) + \cos(40.7457) \times \cos(40.7636) \times hav\,(-73.9876 - (-73.9802)) = 0.42\,\text{miles}$$

Therefore,

$$ws_1 = \sum d_{1,2} = 0.66 \text{ miles}$$

$$ws_2 = \sum d_{3,4} + d_{4,5} + d_{3,5} = 0.51 + 0.84 + 0.42 = 1.77 \text{ miles}$$

Next, we calculate the between cluster segregation.

$$hav\left(\theta_{cluster\#1,cluster\#2}\right) = r \times d_{cluster\#1,cluster\#2} = hav\,(40.7470 - 40.7456) + \cos(40.7470) \times \cos(40.7456) \times hav\,(-73.9836 - (-73.9799)) = 1.16\,\text{miles}$$

Therefore, the ratio of within cluster spread with between cluster spread is given below.

$R_{cluster\#1,cluster\#2} = \frac{0.66+1.77}{1.16} = 2.09$

Considering we only have two clusters, the overall DBI for this example would be equal to 2.09. Whereas, if we increase the number of clusters, we would be computing the $n-1$ ratios for n clusters and select the maximum value as the DBI.

References

[1] INRIX, INRIX, INRIX Global Traffic Scorecard. INRIX – INRIX, 2019, inrix.com/scorecard/.

[2] F. Rajé, M. Tight, F.D. Pope, Traffic pollution: A search for solutions for a city like Nairobi, Cities 82 (2018) 100–107.

[3] Z. Kan, L. Tang, M.P. Kwan, C. Ren, D. Liu, Q. Li, Traffic congestion analysis at the turn level using Taxis' GPS trajectory data, Comput. Environ. Urban Syst. 74 (2019) 229–243.

[4] M.L. Anderson, Subways, strikes, and slowdowns: The impacts of public transit on traffic congestion, Amer. Econ. Rev. 104 (9) (2014) 2763–2796.

[5] H. Marquart, U. Schlink, M. Ueberham, The planned and the perceived city: A comparison of cyclists' and decision-makers' views on cycling quality, J. Transp. Geogr. 82 (2020) 102602.

[6] A. de Palma, M. Kilani, S. Proost, Discomfort in mass transit and its implication for scheduling and pricing, Transp. Res. B 71 (2015) 1–18.

[7] X. Delclòs-Alió, C. Miralles-Guasch, Suburban travelers pressed for time: Exploring the temporal implications of metropolitan commuting in Barcelona, J. Transp. Geogr. 65 (2017) 165–174.

[8] R. Wener, G.W. Evans, P. Boately, Commuting stress: Psychophysiological effects of a trip and spillover into the workplace, Transp. Res. Rec. 1924 (1) (2005) 112–117.

[9] Holden, Goel, Fast-Forwarding to a Future of on-Demand Urban Air Transportation, 2016, (27 October 2016).

[10] S. Shaheen, A. Cohen, E. Farrar, The potential societal barriers of urban air mobility (uam), 2018.

[11] C. Reiche, R. Goyal, A. Cohen, J. Serrao, S. Kimmel, C. Fernando, S. Shaheen, Urban Air Mobility Market Study, 2018.

[12] W. Johnson, C. Silva, E. Solis, Concept vehicles for VTOL air taxi operations, in: Proceedings of the AHS International Technical Meeting on Aeromechanics Design for Transformative Vertical Flight 2018, American Helicopter Society International, 2018.

[13] S. Rajendran, S. Srinivas, T. Grimshaw, Predicting demand for air taxi urban aviation services using machine learning algorithms, J. Air Transp. Manage. 92 (2021) 102043.

[14] B.J. Holmes, A vision and opportunity for transformation of on-demand air mobility, in: 16th AIAA Aviation Technology, Integration, and Operations Conference, AIAA AVIATION Forum, 2016.

[15] P. Butterworth-Hayes, India's DGAC takes first steps to urban air mobility with request for industry BVLOS partners - Urban Air Mobility News, 2019, Retrieved 26 June 2020, from https://www.urbanairmobilitynews.com/emerging-regulations/india-takes-first-steps-to-urban-air-mobility-with-request-for-industry-bvlos-partners/.

[16] Andrew J. Hawkins, Uber Narrows its Search for International City to Host 'Flying Taxis', The Verge, The Verge, 2018, (26 June 2020). www.theverge.com/2018/8/30/17795588/uber-elevate-flying-car-international-city-search-drone.

[17] G. Warwick, New Zealand welcomes flight tests of Kitty Hawk's eVTOL air taxi : full-scale prototypes of Cora air taxi in flight testing; transitional eVTOL combines rotors for vertical flight with wings for efficient forward flight, Aviat. Week Space Technol. (2018) Retrieved from https://trid.trb.org/view/1510576.

[18] Airbus, Vahana has come to an end. But a new chapter at Airbus has just begun, 2019, Retrieved 9 June 2020, from https://www.airbus.com/newsroom/stories/Vahana-has-come-to-an-end.html.

[19] A. Bacchini, E. Cestino, Electric VTOL configurations comparison, Aerospace 6 (3) (2019).

[20] J. Enconniere, J. Ortiz-Carretero, V. Pachidis, Mission performance analysis of a conceptual coaxial rotorcraft for air taxi applications, Aerosp. Sci. Technol. 69 (2017) 1–14.

[21] M. Shamiyeh, J. Bijewitz, M. Hornung, A review of recent personal air vehicle concepts. Bucharest, Romania, 2017.

[22] X. Sun, S. Wandelt, E. Stumpf, Competitiveness of on-demand air taxis regarding door-to-door travel time: A race through Europe, Transp. Res. E 119 (2018) 1–18.

[23] J. Holden, E. Allison, N. Goel, S. Swaintek, Session Presented at the Meeting of the Uber Keynote: Scaling UberAir, 2018.

[24] S. Hasan, Urban Air Mobility (UAM) Market Study, 2019.

[25] S. Rajendran, S. Srinivas, Air taxi service for urban mobility: A critical review of recent developments, future challenges, and opportunities, Transp. Res. E 143 (2020) 102090.

[26] S. Rath, J.Y. Chow, Air taxi skyport location problem for airport access, 2019, arXiv preprint arXiv:1904.01497.

[27] S. Rajendran, J. Zack, Insights on strategic air taxi network infrastructure locations using an iterative constrained clustering approach, Transp. Res. E 128 (2019) 470–505.

[28] M. Shao, Z. Han, J. Sun, C. Xiao, S. Zhang, Y. Zhao, A review of multi-criteria decision making applications for renewable energy site selection, Renew. Energy 157 (2020) 377–403.

[29] J. Lu, B. Li, H. Li, A. Al-Barakani, Expansion of city scale, traffic modes, traffic congestion, and air pollution, Cities 108 (2016) 102974.

[30] E. Pisoni, P. Christidis, P. Thunis, M. Trombetti, Evaluating the impact of Sustainable Urban Mobility Plans on urban background air quality, J. Environ. Manag. 231 (2019) 249–255.

[31] A. Straubinger, Policies addressing possible urban air mobility market distortions – a first discussion, Transp. Res. Procedia 41 (2019) 64–66.

[32] O. Cokorilo, Urban air mobility: Safety challenges, Transp. Res. Procedia 45 (2020) 21–29.

[33] S. Rajendran, E. Pagel, Recommendations for emerging air taxi network operations based on online review analysis of helicopter services, Heliyon 6 (12) (2020) e05581.

[34] C. Al Haddad, E. Chaniotakis, A. Straubinger, K. Plötner, C. Antoniou, Factors affecting the adoption and use of urban air mobility, Transp. Res. A 132 (2020) 696–712.

[35] NASA, Urban Air Mobility (UAM) Market Study, 2019, <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20190026762.pdf>.

[36] L. Swadesir, C. Bil, Urban air transportation journey time comparison for melbourne metropolitan area, in: 31st Congress of the International Council of the Aeronautical Sciences, ICAS 2018, International Council of the Aeronautical Sciences, 2018.

[37] E. Lim, H. Hwang, The selection of vertiport location for on-demand mobility and its application to Seoul Metro Area, Int. J. Aeronaut. Space Sci. 20 (1) (2019) 260–272.

[38] P.A. Bonnefoy, Simulating air taxi networks, in: Proceedings - Winter Simulation Conference, Vol. 2005, 2005, pp. 1586–1595.

[39] R.L. Rothfeld, M. Balac, K.O. Ploetner, C. Antoniou, Agent-based simulation of urban air mobility, in: 2018 Modeling and Simulation Technologies Conference, American Institute of Aeronautics and Astronautics Inc, AIAA, 2018.

[40] D.N. Fadhil, A GIS-Based Analysis for Selecting Ground Infrastructure Locations for Urban Air Mobility. Inlangen (Master's Thesis), Technical University of Munich, 2018.

[41] H. Liu, D.Z.W. Wang, Locating multiple types of charging facilities for battery electric vehicles, Transp. Res. B 103 (2017) 30–55.
[42] A. Sfyridis, P. Agnolucci, Annual average daily traffic estimation in England and Wales: An application of clustering and regression modelling, J. Transp. Geogr. 83 (2020) 102658.
[43] C. Wang, F. He, X. Lin, Z. Shen, M. Li, Designing locations and capacities for charging stations to support intercity travel of electric vehicles: An expanded network approach, Transp. Res. C 102 (2019) 210–232.
[44] F. Xie, C. Liu, S. Li, Z. Lin, Y. Huang, Long-term strategic planning of inter-city fast charging infrastructure for battery electric vehicles, Transp. Res. E 109 (2018) 261–276.
[45] A. Zhang, J. Kang, C. Kwon, Multi-day scenario analysis for battery electric vehicle feasibility assessment and charging infrastructure planning, Transp. Res. C 111 (2020) 439–457.
[46] D. Chauhan, A. Unnikrishnan, M. Figliozzi, Maximum coverage capacitated facility location problem with range constrained drones, Transp. Res. C 99 (2019) 1–18.
[47] J. Escribano Macias, P. Angeloudis, W. Ochieng, Optimal hub selection for rapid medical deliveries using unmanned aerial vehicles, Transp. Res. C 110 (2020) 56–80.
[48] D.W. Lee, E.J. Bass, S.D. Patek, J.A. Boyd, A traffic engineering model for air taxi services, Transp. Res. E 44 (6) (2008) 1139–1161.
[49] M. Salama, S. Srinivas, Joint optimization of customer location clustering and drone-based routing for last-mile deliveries, Transp. Res. Part C: Emerging Technol. 114 (2020) 620–642.
[50] M.Z. Rodriguez, C.H. Comin, D. Casanova, O.M. Bruno, D.R. Amancio, L.D.F. Costa, F.A. Rodrigues, Clustering algorithms: A comparative approach, PLoS One 14 (1) (2019) e0210236.
[51] H. Kim, Y. Ham, Participatory sensing-based geospatial localization of distant objects for disaster preparedness in urban built environments, Autom. Constr. 107 (2019).
[52] H. Millward, M.H. Hafezi, N.S. Daisy, Activity travel of population segments grouped by daily time-use: GPS tracking in Halifax, Canada, Travel Behav. Soc. 16 (2019) 161–170.
[53] M. Sodenkamp, J. Wenig, F. Thiesse, T. Staake, Who can drive electric? Segmentation of car drivers based on longitudinal GPS travel data, Energy Policy 130 (2019) 111–129.
[54] Y. Wang, Y. Deng, F. Ren, R. Zhu, P. Wang, T. Du, Q. Du, Analysing the spatial configuration of urban bus networks based on the geospatial network analysis method, Cities 96 (2020) 102406.
[55] S. Rajendran, Real-time dispatching of air taxis in metropolitan cities using a hybrid simulation goal programming algorithm, Expert Syst. Appl. 178 (2021) 115056.
[56] G. Usman, U. Ahmad, M. Ahmad, Improved k-means clustering algorithm by getting initial cenroids, World Appl. Sci. J. 27 (4) (2013) 543–551.
[57] A. Ballis, Airport site selection based on multicriteria analysis: the case study of the island of Samothraki, Oper. Res. 3 (3) (2003) 261–279.
[58] T. Semih, S. Seyhan, A multi-criteria factor evaluation model for gas station site selection, Evaluation 2 (1) (2011) 12–21.
[59] M. Ghorbanzadeh, M. Effati, M. Gilanifar, E.E. Ozguven, Subway station site selection using GIS-based multi-criteria decision-making: A case study in a developing country, Comput. Res. Prog. Appl. Sci. Eng. 6 (2020) 60–69.
[60] Y.E. Hawas, M.N. Hassan, A. Abulibdeh, A multi-criteria approach of assessing public transport accessibility at a strategic level, J. Transp. Geogr. 57 (2016) 19–34.
[61] G.H. Tzeng, M.H. Teng, J.J. Chen, S. Opricovic, Multicriteria selection for a restaurant location in Taipei, Int. J. Hosp. Manag. 21 (2) (2002) 171–187.
[62] G.H. Tzeng, C.W. Lin, S. Opricovic, Multi-criteria analysis of alternative-fuel buses for public transportation, Energy Policy 33 (11) (2005) 1373–1383.
[63] H. Bock, Origins and extensions of the k -means algorithm in cluster analysis, Electron. J. Hist. Probab. Stat. 4 (2) (2008) 1–18.
[64] S. Rajendran, J. Shulman, Study of emerging air taxi network operation using discrete-event systems simulation approach, J. Air Transp. Manage. (2020) http://dx.doi.org/10.1016/j.jairtraman.2020.101857.