Taxi Service Area Design: Formulation and Analysis

Abstract

Multiple taxi types with different service areas exist in practice. However, to the best of our knowledge, there is no methodology to determine the best service region of each type of restricted area taxi in a taxi market. It is also unclear what the determinant factors are for the design. This paper proposes a nonlinear mixed integer programming model to determine the service areas. The objective is to maximize social welfare. A greedy heuristic is developed to solve the model. Numerical examples are given to show the performance of the heuristic and the determinant factors for the design.

Keywords:

Taxi service area design, urban taxi services, multiple vehicle modes; customers' equity; social welfare maximization

1. Introduction

Taxis are an important part of the transportation system of any city. Taxis offer 24-hour, comfortable, and door-to-door transportation services. The taxi industry is usually regulated in two main ways: entry restriction and price control. In order to examine the economic consequences of different regulatory policies, previous studies have been conducted (e.g., Douglas, 1972; De Vany, 1975; Manski and Wright, 1976; Foerster and Gilbert, 1979; Shreiber, 1981; Beesley and Glaister, 1983; Schroeter, 1983; Frankena and Pautler, 1986; Hackner and Nyberg, 1995; Arnott, 1996; Cairns and Liston-Heyes, 1996; Flores-Guri, 2003; Yang et al., 2005a,b; Fernandez et al., 2006; Moore and Balaker, 2006; Loo et al., 2007; Yang and Yang, 2011; Yang et al., 2014; Wang et al., 2016). The general objective of these studies is to understand the manner in which demand and supply are equilibrated in the presence of such policies, hence providing useful insights for the taxi policy-making process of the government (Yang et al., 2002).

All of the above studies use an aggregate approach and do not take into account the spatial structure of the taxi market. To address this problem, Yang and Wong (1998) first made an attempt to depict taxi movements in a road network under a given origin-destination (O-D) pattern, in which a simultaneous system of equations was proposed to describe the movements of both occupied and vacant taxis. Their model was calibrated and validated by Wong et al. (1999). Further enhancements and extensions on their network model of urban taxi services were made in various ways to deal with demand elasticity (e.g., Wong et al., 2001; Yang et al., 2002), bi-lateral customer and taxi searching (e.g., Wong et al., 2005), multi-class and multi-mode taxi services (e.g., Wong et al., 2004, 2008), non-linear taxi pricing (e.g., Yang et al., 2010a), searching and meeting frictions (e.g., Yang et al., 2010b), dynamic taxi demand and information (e.g., Yang et al., 2005b; Long et al., 2017), and e-hailing (e.g., He and Shen, 2015; Qian and Ukkusuri, 2017a; He et al., 2018).

Until most recently, the advancement of data science and technology provided a useful tool to further study the properties of the taxi market. Generally, the data-driven approach was used to study taxi ridership (e.g., Qian and Ukkusuri, 2015), the customer search behavior of vacant taxi drivers (e.g., Wong et al., 2014a,b, 2015), taxi group ride (e.g., Qian et al., 2017), time-of-day pricing (e.g., Qian and Ukkusuri, 2017b), and the efficiency of e-hailing services (e.g., Zhan et al., 2016). The most of the above aggregate and network models can be reformulated into single-level optimization models with regulatory variables, such as fare and fleet size, as model parameters. In contrast, only a few studies proposed bilevel optimization models to determine optimal regulatory variables for road networks with taxis (e.g., Zhu et al., 2013; Zhang and Ukkusuri, 2016).

Although the properties and regulations of the taxi market have been extensively investigated, some critical regulatory problems still need to be further studied, one of which is the taxi service area design (TSAD) problem. The service area of a taxi is a region of the city in which taxis are allowed to run and provide services to customers. Conventionally, taxis are allowed to operate anywhere in a city. However, in some cities, taxis are classified into different types according to their confined service areas. Taxis with a service area restriction are only allowed to provide their service within the designated area. For instance, Hong Kong introduced two types of taxis known as the New Territory (NT) taxis and the Lantau taxis in 1978 in addition to the urban taxis. The urban taxis are allowed to operate anywhere in the city, while the NT taxis and the Lantau taxis are only allowed to provide their services to customers within their respective service areas (see Fig. 1). According to a report from the Hong Kong government, customer demand for public transport, especially the long-haul demand, had been increasing rapidly with the development of new towns and major infrastructure projects in the remote areas of the city (HKSAR, 2008). However, the fare level of the urban taxis made them less competitive with public transport to serve long-distance travelers. Consequently, urban taxis tended to cluster in the more profitable and populated areas, which led to a shortage of taxis in the newly developed rural towns and thus the customer waiting time for a taxi in those areas became long. Therefore, the NT taxis and the Lantau taxis were introduced with the expectation of improving taxi availability and service quality in the remote areas. The practice of TSAD can also be found in New York City. A Boro Taxi program was launched in 2013 to introduce a new type of taxi to the city known as the Boro Taxis. According to the New York City Taxi & Limousine Commission, the GPS data collected from the existing yellow taxis indicated that 95% of the taxi pick-ups occurred in Manhattan below 96th Street and at John F. Kennedy International and LaGuardia airports, which made residents outside the aforementioned areas more difficult to hail a taxi on streets (New York City Taxi & Limousine Commission, 2018). Therefore, the Boro Taxis were introduced and their service areas are confined within the under-served areas with the expectation of improving taxi availability and service quality. Apart from the realistic applications of TSAD, we believe that further discussions on TSAD are needed for the following reasons. First, the impacts and necessity of TSAD need to be verified. Second, there is no methodology to determine the best service area of each taxi type. Last but not least, the determinant factors to the design need to be investigated. Unfortunately, despite the substantial previous studies mentioned above, the attention that had been paid to the TSAD problem was extremely rare. Wong et al. (2008) only developed a taxi network model with the consideration of the service region of restricted area taxis. Liang et al. (2016) only introduced a methodology to determine the service area of a homogenous taxi system used as the last mile of travelers after taking trains.

To fill the research gaps, this paper proposes an optimization model for designing the service areas of restricted area taxis by extending the taxi service model proposed by Wong et al. (2008), which considers multiple taxi types serving different areas. Our proposed model captures congestion effects and two modes, namely taxis and non-taxis (i.e., private cars and other public transports). The taxis can be further classified into normal taxis and restricted area taxis according to their respective service areas. Normal taxis can operate anywhere in the network whereas restricted area taxis are only allowed to operate in the area determined by the regulator. The TSAD problem considers only one type of normal taxi but multiple types of restricted area taxis. It is formulated as an optimization model that is aimed at maximizing social welfare. The problem has two sub-problems. The first sub-problem is the combined network equilibrium problem (CNEP). The second sub-problem is the regulatory problem, which is to select a specific region from the entire network as the service area of each type of restricted area taxi. A greedy heuristic that encompasses a sub-algorithm for solving the CNEP is developed to solve the model. Using numerical examples, we examine the performance of the developed greedy heuristic, compare it with that of genetic algorithm (GA) and the enumeration method, and provide insights into TSAD. The results indicate that the greedy heuristic is more efficient to GA. The efficiency of the developed greedy heuristic is more significant as the network size is larger. The results also show that the best service area design may vary with the proportion of the fleet size of restricted area taxis, the total fleet size of all taxi types, total travel demand and distribution, the fare level of restricted area taxis, and design objectives. Finally, a case study is performed to show that the insights into TSAD obtained from the small network can be scalable to a larger network.

The main contributions of this paper include the following:

1. We introduce a new TSAD problem with multiple taxi types and determine the service areas of restricted area taxis.
2. We propose a methodology to formulate the problem.
3. We develop a novel greedy heuristic to solve the problem.
4. We provide insights into designing taxi service areas.

The outline of this paper is as follows. Section 2 develops the mathematical formulation of the proposed optimization model. Section 3 introduces the greedy heuristic. Section 4 presents the numerical examples. Finally, Section 5 concludes the paper and gives future research directions.

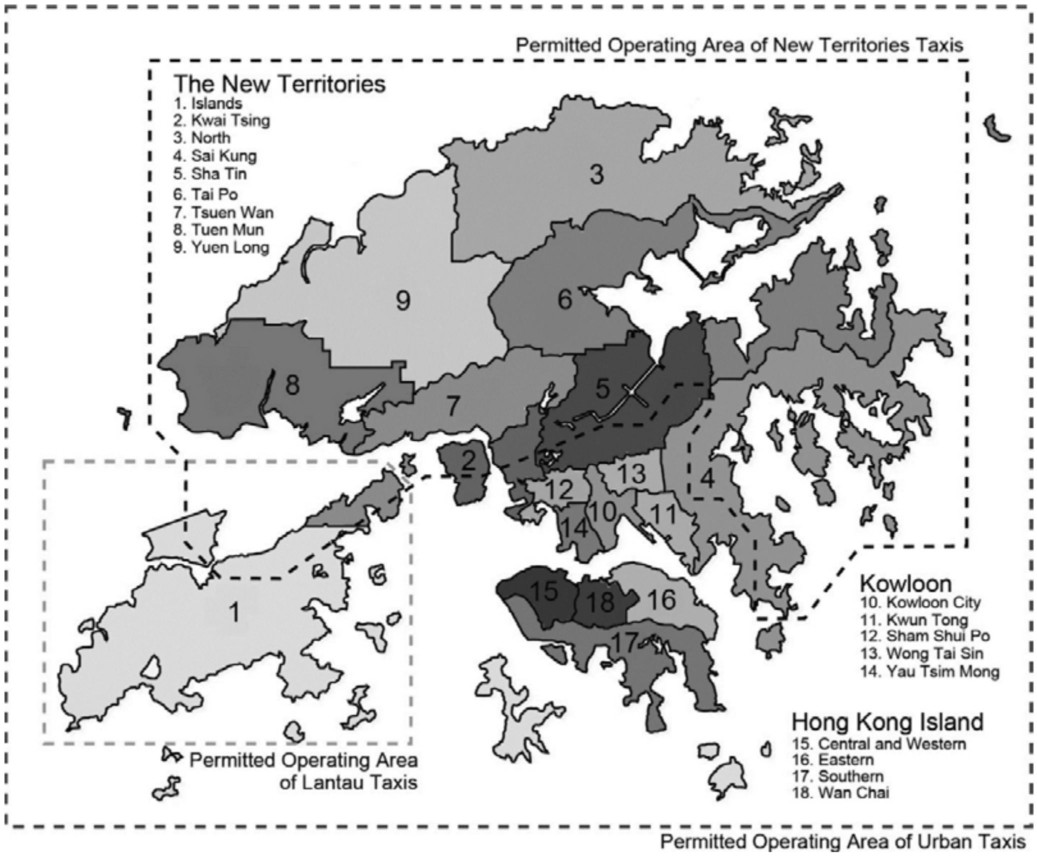

Fig. 1 District map of Hong Kong and permitted operating areas of taxis (Wong et al., 2015)

2. Mathematical formulation

Consider a network $G(V, A)$ where V is the set of nodes and A is the set of directed links. A node represents a zone of the study area and a directed link represents a one-way main road. There are three modes of vehicles: normal taxis, restricted area taxis, and other vehicles (i.e., non-taxi vehicles). The normal taxis are allowed to operate anywhere in the network, while the restricted area taxis are only allowed to operate within the specific region determined by the regulator. The TSAD problem is to select zones to form different regions, which is aimed at maximizing social welfare subject to specific design constraints and network equilibrium conditions. Each region is designated as the service area of a type of restricted area taxi.

2.1. Notations

The following notations are used in this paper:

Indices

i, j	Indices for the origin and destination nodes (zones) of travel;
e, u, l	Indices for nodes;
r	Index for paths;
q	Index for taxi types, which represents normal taxis when it equals 1, and represents restricted area taxis when it is greater than 1;
m	Index for vehicle classes.

Sets

I	Set of origin nodes (or zones);
J	Set of destination nodes (or zones);
Q	Set of taxi types;
M	Set of all vehicle classes: 1) non-taxis (n), 2) different types of occupied taxis (o,q), and 3) different types of vacant taxis (v,q);
R_{ij}^{m}	Set of paths for class *m* vehicles from zone i to zone j.

Parameters/Constants

b_0	Value of time;
b_1	Value of customer waiting time for taxis;
b^{n}	Mileage cost incurred by a traveler who takes non-taxi modes ($/km);
$b_1^{o,q}$	Mileage charge to a user in a type q taxi ($/km);
$b_2^{o,q}$	Congestion-based charge to a user in a type q taxi ($/h);
b_{d}^{q}	Mileage operating cost for a type q taxi ($/km);
b_{c}^{q}	Hourly operating cost for a type q taxi ($/h);
N^{q}	Fleet size of type q taxis;
K	A large constant;
$\delta_{ij,eu,r}^{m}$	Element of the link route incidence matrix, where $\delta_{ij,eu,r}^{m}=1$ if route r for class m vehicles between OD pair (i, j) passes through link (e,u), and 0 otherwise.

Decision variables

f_r^m	Traffic flow of class m vehicles on path $r \in R_{ij}^m$;
$\mathbf{f}$	$\left(f_r^m \right)_{r \in R_{ij}, m \in M, i \in I, j \in J}$;
T_{ij}^m	Auxiliary variable representing the travel demand for class m vehicles from zone i to zone j;
O_i^q	Auxiliary variable representing the total customer demand for type q taxis originated from zone i;
D_j^q	Auxiliary variable representing the total customer demand for type q taxis to zone j;
X_l^q	Binary decision variable, which equals 1 if zone l is included in the operating area of type q restricted area taxis, and 0 otherwise;
$\mathbf{X}$	$\left(X_l^q \right)_{l \in V, q \in Q-\{1\}}$;
y_{eu}^q	Non-negative continuous decision variable that indicates the amount of flow with respect to type q restricted area taxis from node e to its adjacent node u;
$\mathbf{y}$	$\left(y_{eu}^q \right)_{(e,u) \in A, q \in Q-\{1\}}$;
s_l^q	Binary decision variable, which equals 1 if node l is chosen as a sink with respect to type q restricted area taxis, and 0 otherwise;
$\mathbf{s}$	$\left(s_l^q \right)_{l \in V, q \in Q-\{1\}}$;
ϕ_{eu}^q	Binary decision variable, which equals 1 if both nodes e and u are in the service area of type q restricted area taxis, and 0 otherwise;
$\mathbf{\Phi}$	$\left(\phi_{eu}^q \right)_{(e,u) \in A, q \in Q-\{1\}}$.

Functions

$t_{eu}(\varsigma_{eu})$	Vehicle travel time on link (e,u) with a traffic volume of ς_{eu};
c_{eu}^m	Generalized travel cost for class m vehicles to travel on link (e,u) as perceived by the corresponding users;
C_r^m	Total generalized travel cost (disutility) for class m vehicles to travel on path r;

C_{ij}^{m} Minimum generalized travel cost for class m vehicles to travel from zone i to zone j;

W_{i}^{q} Customer waiting time for a type q taxi in zone i;

w_{i}^{q} Taxi waiting time of a type q taxi in zone i;

$P_{i/j}^{v,q}$ The probability of a vacant type q taxi that originates from zone j and meets a customer in zone i;

P_{ij}^{n} The probability of a traveler who chooses a non-taxi class to travel from zone i to zone j;

P_{ij}^{o} The probability of a customer who takes a taxi from zone i to zone j;

$P_{ij}^{o,q}$ The probability of a customer who takes a type q taxi from zone i to zone j;

L_{ij}^{o} Logsum of the disutility of travel perceived by users who take taxis from zone i to zone j;

L_{ij}^{D} Average consumer surplus perceived by users from zone i to zone j;

Y_{i}^{q} Expected profit of a driver of a type q taxi that meets customer in zone i;

$h_{ij,\mathrm{t}}^{m}$ Average travel time of class m vehicles from zone i to zone j;

$h_{ij,\mathrm{d}}^{m}$ Average travel distance of class m vehicles from zone i to zone j;

Z_{i}^{q} Number of vacant type q taxis that meet customers in zone i per hour.

Using the above notations, the TSAD problem can be formulated into a single-level optimization model and has two sub-problems, namely the CNEP (see Section 2.2) and the regulatory problem (see Section 2.3).

2.2. The combined network equilibrium problem

The first sub-problem is the combined network equilibrium problem (CNEP) adapted from the study of Wong et al. (2008) and describes taxi movements in a network with multiple taxi types. A revision is made (Section 2.2.4) by using a different behavioral assumption on the customer search of vacant taxi drivers. The CNEP is expressed as a set of constraints as depicted below.

2.2.1. Occupied taxi movements and hierarchical mode choice

In the following, let the superscripts "n", "o", and "v" be, non-taxis, occupied taxis, and vacant taxis, respectively. We further denote occupied taxis and vacant taxis with respect to different taxi types as "o,q" and "v,q", respectively, in which q represents normal taxis when it equals 1, and represents restricted area taxis when it is greater than 1. We also denote $m \in M = \{(n),(o,q),(v,q)\}$ as the index of vehicle classes. Denote I and J as the sets of origins and destinations in the network $G(V,A)$, respectively. Let $\bar{D}_{ij}$ be the total travel demand from origin $i \in I$ to destination $j \in J$, which is fixed and known. With taxis and non-taxi vehicles running in the network, $\bar{D}_{ij}$ is expressed as

$$\bar{D}_{ij} = T_{ij}^{n} + T_{ij}^{o}, \forall\ i \in I,\ j \in J, \tag{1}$$

in which T_{ij}^{n} and T_{ij}^{o} are, respectively, the non-taxi demand and the total taxi demand from i to j and the latter one equals the sum of demand for each taxi type:

$$T_{ij}^{o} = \sum_{q \in Q} T_{ij}^{o,q},\ \forall i \in I,\ j \in J. \tag{2}$$

$T_{ij}^{o,q}$ is the number of occupied taxis of type q from i to j. For each taxi type q, we have the following trip end constraints:

$$\sum_{j \in J} T_{ij}^{o,q} = O_i^q,\ \forall i \in I,\ q \in Q \quad \text{and} \tag{3}$$

$$\sum_{i \in I} T_{ij}^{o,q} = D_j^q,\ \forall j \in J,\ q \in Q. \tag{4}$$

We further assume that travelers select their mode based on the hierarchical logit mode choice structure as shown in Fig. 2. They select between non-taxis and taxis in the upper-level and choose between normal and restricted area taxis in the lower-level.

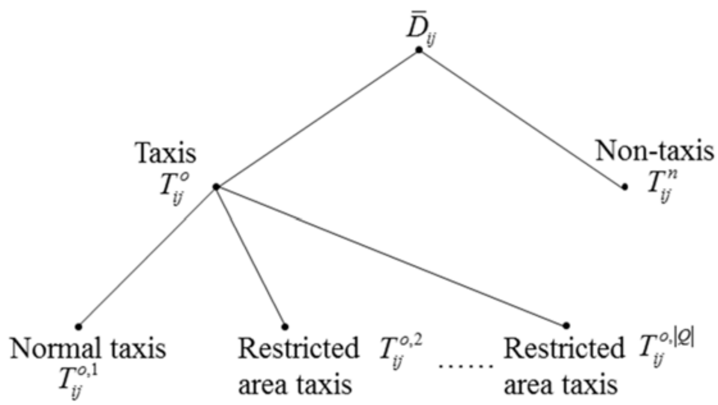

Fig. 2 A hierarchical mode choice structure for travelers

The probability of travelers taking non-taxis from i to j is given as

$$P_{ij}^{n} = \frac{\exp(-\beta_1 C_{ij}^{n})}{\exp(-\beta_1 C_{ij}^{n}) + \exp(-\beta_1 L_{ij}^{o})}, \ \forall i \in I, \ j \in J, \tag{5}$$

in which β_1 is the upper-level dispersion coefficient and C_{ij}^{n} is the minimum generalized travel cost perceived by travelers taking non-taxis from i to j. L_{ij}^{o} is the logsum of the disutility of travel perceived by customers who take taxis from i to j. It is expressed as

$$L_{ij}^{o} = -\frac{1}{\beta_2} \sum_{q \in Q} \exp(-\beta_2 C_{ij}^{o,q}), \ \forall i \in I, \ j \in J, \tag{6}$$

where β_2 is the lower-level dispersion coefficient and $C_{ij}^{o,q}$ is the minimum generalized travel cost perceived by the customers who take type q taxis from i to j. Clearly, the probability of customers taking taxis from i to j is expressed as

$$P_{ij}^{o} = 1 - P_{ij}^{n}, \forall i \in I, \ j \in J. \tag{7}$$

Given that a traveler has chosen to travel by taxi, the probability that he/she chooses a type q taxi from i to j is given as

$$P_{ij}^{o,q} = \frac{\exp(-\beta_2 C_{ij}^{o,q})}{\sum_{q' \in Q} \exp(-\beta_2 C_{ij}^{o,q'})}, \ \forall q \in Q, i \in I, \ j \in J. \tag{8}$$

For any given total travel demand from i to j, the number of trips taken by non-taxis and type q taxis can now be respectively expressed as

$$T_{ij}^{n} = \bar{D}_{ij} P_{ij}^{n}, \ \forall i \in I, j \in J \quad \text{and} \tag{9}$$

$$T_{ij}^{o,q} = \bar{D}_{ij} P_{ij}^{o} P_{ij}^{o,q},\ \forall q \in Q, i \in I, j \in J\ . \tag{10}$$

2.2.2. Cost structure

The generalized travel cost of a link is associated with its link travel time, in which the latter is determined by the hourly traffic flow of the link. Denote $(e,u) \in A$ as a link in the network, in which e and u form a pair of adjacent nodes (Note that $e = u$ when it represents an intra-zonal link). The hourly traffic flow of (e,u) is then given as

$$\varsigma_{eu} = \sum_{m \in M} \sum_{i \in I} \sum_{j \in J} \sum_{r \in R_{ij}^{m}} \delta_{ij,eu,r}^{m} f_{r}^{m},\ \forall (e,u) \in A, \tag{11}$$

where $\delta_{ij,eu,r}^{m}$ is an element of the link-route incidence matrix, which is equal to 1 if path r for class m vehicles between OD pair (i, j) traverses link (e,u) and 0 otherwise. f_{r}^{m} is the hourly traffic flow of vehicle class m on route $r \in R_{ij}^{m}$.

The travel time t_{eu} of link (e,u) is defined as

$$t_{eu}(\varsigma_{eu}) = t_{eu}^{0}(1 + 0.5(\varsigma_{eu}/S_{eu})^{2}),\ \forall (e,u) \in A, \tag{12}$$

in which S_{eu} represents the capacity of link (e,u) and t_{eu}^{0} is the free-flow travel time on that link.

The generalized travel cost of class m vehicles on link (e,u), which is denoted as c_{eu}^{m}, is defined by the weighted sum of the link travel time t_{eu} and link length d_{eu}. Then, we have c_{eu}^{m} expressed as

$$c_{eu}^{n} = b_0 t_{eu}(\varsigma_{eu}) + b^{n} d_{eu},\ \forall (e,u) \in A, \tag{13}$$

$$c_{eu}^{o,q} = b_0 t_{eu}(\varsigma_{eu}) + b_1^{o,q} d_{eu} + b_2^{o,q} t_{eu}(\varsigma_{eu}),\ \forall (e,u) \in A,\ q \in Q,\ \text{and} \tag{14}$$

$$c_{eu}^{v,q} = b_{\mathrm{c}}^{q} t_{eu}(\varsigma_{eu}) + b_{\mathrm{d}}^{q} d_{eu},\ \forall (e,u) \in A,\ q \in Q, \tag{15}$$

where b_0 is the value of time, b^{n} is the mileage cost incurred by a traveler who takes non-taxi classes. $b_1^{o,q}$ and $b_2^{o,q}$ are, respectively, the mileage and congestion-based fares charged to customers who take type q taxis. b_{d}^{q} and b_{c}^{q} represent, respectively, the mileage and hourly operating costs for a type q taxi.

With Eqs. (13)-(15), the generalized travel cost of class m vehicles on path $r \in R_{ij}^{m}$ can be defined as follows.

$$C_r^n = \sum_{(e,u)\in A} \delta_{ij,eu,r}^n c_{eu}^n,\ \forall r \in R_{ij}^n, i \in I,\ j \in J, \tag{16}$$

$$C_r^{o,q} = \sum_{(e,u)\in A} \delta_{ij,eu,r}^{o,q} c_{eu}^{o,q} + b_1 W_i^q,\ \forall r \in R_{ij}^{o,q},\ q \in Q,\ i \in I,\ j \in J, \text{ and} \tag{17}$$

$$C_r^{v,q} = \sum_{(e,u)\in A} \delta_{ij,eu,r}^{v,q} c_{eu}^{v,q},\ \forall r \in R_{ji}^{v,q},\ q \in Q,\ i \in I,\ j \in J, \tag{18}$$

where b_1 is the value of customer waiting time for taxis and W_i^q is the customer waiting time for a type q taxi in zone i. We can then define the minimum generalized travel cost of a class m vehicle from i to j as

$$C_{ij}^m = \min(C_r^m,\ \forall r \in R_{ij}^m, m \in M,\ i \in I,\ j \in J). \tag{19}$$

With the travel time on each link from Eq. (12) and the distance of each link, we can define the average travel time and distance of each path, respectively:

$$h_{ij,\mathrm{t}}^m = \frac{\sum_{r\in R_{ij}^m} \left(f_r^m \sum_{(e,u)\in A} \delta_{ij,eu,r}^m t_{eu}(\varsigma_{eu}) \right)}{\sum_{r\in R_{ij}^m} f_r^m},\ \forall m \in M,\ i \in I,\ j \in J \text{ and} \tag{20}$$

$$h_{ij,\mathrm{d}}^m = \frac{\sum_{r\in R_{ij}^m} \left(f_r^m \sum_{(e,u)\in A} \delta_{ij,eu,r}^m d_{eu} \right)}{\sum_{r\in R_{ij}^m} f_r^m},\ \forall m \in M,\ i \in I,\ j \in J, \tag{21}$$

where $h_{ij,\mathrm{t}}^m$ and $h_{ij,\mathrm{d}}^m$ are, respectively, the average travel time and distance of class m vehicles from i to j.

2.2.3. Customer and taxi waiting times

It is defined that the customer waiting time $W_i^q,\ \forall i \in I,\ q \in Q$, which is an endogenous variable of the CNEP, varies across zones and depends on both the density and the searching/waiting time of vacant type q taxis in zone i, i.e., $W_i^q = W_i^q(Z_i^q, w_i^q),\ \forall i \in I,\ q \in Q,$ where Z_i^q is the number of vacant type q taxis that meet passengers in zone i per hour, and $w_i^q,\ \forall i \in I, q \in Q$ is the searching/waiting time of vacant type q taxis in zone i. Note that at equilibrium, $Z_i^q = O_i^q,\ \forall i \in I, q \in Q$. The specified customer waiting time used in this paper was derived from Douglas (1972) and Yang et al. (2002) and is given as

$$W_i^q = \eta \frac{\Phi_i}{Z_i^q w_i^q} = \eta \frac{\Phi_i}{O_i^q w_i^q}, \ \forall q \in Q, \ i \in I, \tag{22}$$

where Φ_i is the area of zone i and η is a parameter that is assumed to be the same for all zones.

2.2.4. Profitability-based vacant taxi movement

When a customer ride is completed, a taxi becomes vacant and cruises either in the same zone or goes to another zone in search of customers. Unlike Wong et al. (2008), we assume that during the customer search, each taxi driver attempts to maximize the expected profitability before finishing the next customer ride and becoming vacant again (see Wong et al., 2003). Hence, the probability of a type q vacant taxi that originates from zone j and eventually meets a customer in zone i can be formulated as the following logit model:

$$P_{i/j}^{v,q} = \frac{\exp\{\theta^q (Y_i^{o,q} - (C_{ji}^{v,q} + b_0^q w_i^q))\}}{\sum_{i' \in I} \exp\{\theta^q (Y_{i'}^{o,q} - (C_{ji'}^{v,q} + b_0^q w_{i'}^q))\}}, \ \forall q \in Q, \ i \in I, \ j \in J, \tag{23}$$

in which θ^q is the dispersion coefficient of the logit model for vacant type q taxis and $(C_{ji}^{v,q} + b_0^q w_i^q)$ is the total searching and waiting cost for a type q vacant taxi from zone j before meeting a customer in zone i. $Y_i^{o,q}$ is the expected profit perceived by the driver of a type q taxi from the next customer ride that originates from zone i, which is given as

$$Y_i^{o,q} = \frac{\sum_{j \in J} T_{ij}^{o,q} ((b_1^{o,q} - b_d^q) h_{ij,\mathrm{d}}^{o,q} + (b_2^{o,q} - b_c^q) h_{ij,\mathrm{t}}^{o,q})}{O_i^q}, \ \forall q \in Q, \ i \in I. \tag{24}$$

In a stationary equilibrium state, the following conservation of flows of vacant taxis must also hold:

$$\sum_{i \in I} T_{ji}^{v,q} = D_j^q, \ \forall q \in Q, \ j \in J \quad \text{and} \tag{25}$$

$$\sum_{j \in J} T_{ji}^{v,q} = \sum_{j \in J} D_j^q \cdot P_{i/j}^{v,q} = O_i^q, \ \forall q \in Q, \ i \in I. \tag{26}$$

2.2.5. Taxi service time constraint

For each taxi type $q \in Q$, it is obvious that in one unit period (1 h), the sum of the total occupied and vacant times of all taxis is equal to the total taxi service time $N^q \times 1$. The total occupied time is given by $\sum_{i\in I}\sum_{j\in J} T_{ij}^{o,q} h_{ij,\mathrm{t}}^{o,q}$, while the total vacant time is the sum of the total searching and waiting times, which is given by $\sum_{j\in J}\sum_{i\in I} T_{ji}^{v,q}(h_{ji,\mathrm{t}}^{v,q}+w_i^q)$. Therefore, the following taxi service time constraint must be satisfied:

$$\sum_{i\in I}\sum_{j\in J} T_{ij}^{o,q} h_{ij,\mathrm{t}}^{o,q} + \sum_{j\in J}\sum_{i\in I} T_{ji}^{v,q}(h_{ji,\mathrm{t}}^{v,q}+w_i^q) = N^q,\ \forall q\in Q. \tag{27}$$

2.2.6. Equilibrium, flow conservation, and non-negativity constraints

In order to obtain the equilibrium path flow $f_r^m,\ \forall r\in R_{ij}^m, m\in M, i\in I, j\in J$ from the CNEP, the following constraints are needed:

$$f_r^m(C_r^m - C_{ij}^m) = 0,\ \forall r\in R_{ij}^m, m\in M, i\in I, j\in J, \tag{28}$$

$$\sum_{r\in R_{ij}^m} f_r^m = T_{ij}^m,\ \forall m\in M, i\in I, j\in J, \text{ and} \tag{29}$$

$$f_r^m \geq 0,\ \forall r\in R_{ij}^m, m\in M, i\in I, j\in J. \tag{30}$$

Eq. (28) is the user equilibrium constraint, which ensures that flow can be positive if the route used gives the lowest cost. Eq. (29) is the conservation of path flows, which indicates that the sum of path flows of each class m between each O-D pair is equal to the corresponding O-D flow. Eq. (30) is the non-negativity constraint for path flows.

2.3. Design constraints

The second sub-problem is the regulatory problem in which the regulator selects regions in the entire network and each region is designated as the service area of a type of restricted area taxi. All regions must satisfy two specific criteria, namely the minimum-size and the contiguity constraints. The minimum-size constraint requires that the service area of each type of restricted area taxi must be non-empty, meaning that the service area must be comprised of at least one node (zone) from the study area. The contiguity constraint requires that every pair of nodes in the selected region is connected, which means that for every pair of nodes in the selected region, there exists at least one path in the region connecting them.

The minimum-size constraint is defined as

$$\sum_{l\in V} X_l^q \geq 1,\ \forall q\in Q-\{1\}, \tag{31}$$

where X_l^q is the binary decision variable which is equal to 1 if node l is included in the service area of type q restricted area taxis, and 0 otherwise.

To formulate the contiguity constraints, the formulation approach developed by Shirabe (2005) is used. The approach assumes that for the service area of each type of restricted taxi, there is only one arbitrarily chosen sink and every other node provides one unit of supply. For a service area to be contiguous, the supply sent from every source must ultimately arrive at the sink, without passing through the outside of the service area. The contiguity constraints are given as a set of linear constraints as follows.

$$\sum_{\{u|(e,u)\in A\}} y_{eu}^q - \sum_{\{u|(u,e)\in A\}} y_{ue}^q \geq X_e^q - K s_e^q,\ \forall e \in V,\ q \in Q - \{1\}, \tag{32}$$

$$\sum_{l\in V} s_l^q = 1,\ \forall q \in Q - \{1\}, \tag{33}$$

$$\sum_{\{u|(u,e)\in A\}} y_{ue}^q \leq (K-1) X_e^q,\ \forall e \in V,\ q \in Q - \{1\}, \tag{34}$$

$$X_e^q + X_u^q \geq 2\phi_{eu}^q,\ \forall e, u \in V,\ q \in Q - \{1\}, \tag{35}$$

$$K\phi_{eu}^q \geq \sum_{m\in\{(o,q),(v,q)\}} \sum_{i\in I} \sum_{j\in J} \sum_{r\in R_{ij}^m} \delta_{ij,eu,r}^m f_r^m,\ \forall (e,u) \in A, q \in Q - \{1\}, \tag{36}$$

$$y_{eu}^q \geq 0,\ \forall (e,u) \in A,\ q \in Q - \{1\}, \tag{37}$$

$$X_l^q = \{0,1\},\ \forall l \in V,\ q \in Q - \{1\}, \tag{38}$$

$$s_l^q = \{0,1\},\ \forall l \in V,\ q \in Q - \{1\},\ \text{ and} \tag{39}$$

$$\phi_{eu}^q = \{0,1\},\ \forall (e,u) \in A,\ q \in Q - \{1\}, \tag{40}$$

where y_{eu}^q is a non-negative continuous decision variable of type q restricted area taxis that indicates the amount of (imaginary) flow from e to u. s_l^q is a binary decision variable indicating whether node l is chosen as the sink of the service area of type q restricted area taxis ($s_l^q = 1$ if node l is a sink, $s_l^q = 0$ otherwise). ϕ_{eu}^q is a binary decision variable, which equals 1 if both nodes e and u are in the service area of type q restricted area taxis, and 0 otherwise. K is a large constant.

Constraint (32) represents the net outflow of each type of restricted area taxi from each node e. Constraint (33) requires that for each type of restricted area taxi, one and only one node is the sink. Constraint (34) avoids any (imaginary) flow entering into any node e outside the service area of type q restricted area taxis. Constraint (35) ensures that a link is included in any service area of restricted area taxis only if its tail and head nodes are selected into that area. Constraint (36) makes sure that the flow of each type of restricted area taxi on a link can be positive only if the link is inside the corresponding restricted area. Conditions (37)-(40) define the variable domains.

2.4. Social welfare

The social welfare (denoted as SW), which measures the effectiveness of TSAD, is a function of equilibrium path flows and the design variables of the regulatory problem. It is expressed as

$$SW = TCS + TPS, \tag{41}$$

where TCS denotes the total consumer surplus and TPS represents the total producer surplus, i.e., the profit of all taxi drivers per hour.

The total consumer surplus, TCS, can be calculated by taking a weighted sum of the average consumer surplus, where the "weight" is the total travel demand between each O-D pair. Therefore, TCS is formulated as

$$TCS = \sum_{i\in I}\sum_{j\in J} L_{ij}^{\mathrm{D}} \bar{D}_{ij}, \tag{42}$$

where L_{ij}^{D} is the average consumer surplus of travelers who travel from zone i to zone j by either taxi or non-taxi and is defined as

$$L_{ij}^{\mathrm{D}} = \frac{1}{\beta_1}\ln(\exp(-\beta_1 L_{ij}^{o}) + \exp(-\beta_1 C_{ij}^{n})),\ \forall i \in I,\ j \in J. \tag{43}$$

The total producer surplus TPS is the difference between the total taxi fares charged to all customers (denoted as TF) and the total taxi operating costs (denoted as TOC). The formula of TPS is expressed as

$$TPS = TF - TOC. \tag{44}$$

We assume that the individual fare of taking a type q taxi from zone i to zone j along path $r \in R_{ij}^{o,q}$ is the sum of the cost of total travel distance and the cost of total travel time:

$$F_{ij,r}^{o,q} = \sum_{(e,u)\in A} \delta_{eu,r}^{o,q}(b_1^{o,q} d_{eu} + b_2^{o,q} t_{eu}(\varsigma_{eu})),\ \forall r \in R_{ij}^{o,q}, i \in I,\ j \in J. \tag{45}$$

We can then obtain TF by summing up all individual collected fares:

$$TF = \sum_{q\in Q}\sum_{i\in I}\sum_{j\in J}\sum_{r\in R_{ij}^{o,q}} f_r^{o,q} F_{ij,r}^{o,q}. \tag{46}$$

Substituting Eqs. (45) into Eq. (46) and utilizing Eqs. (20), (21), and (29), we obtain

$$TF = \sum_{q\in Q}\sum_{i\in I}\sum_{j\in J} T_{ij}^{o,q}(b_1^{o,q} h_{ij,d}^{o,q} + b_2^{o,q} h_{ij,\mathrm{t}}^{o,q}). \tag{47}$$

The total operating costs (TOC) can be obtained by summing up the individual operating cost of each taxi, including occupied and vacant taxis. The operating cost for an individual type q taxi that travels from zone i to zone j is the sum of the mileage operating cost and the hourly operating cost along path $r \in R_{ij}^{o,q}$:

$$OC_{ij,r}^{o,q} = \sum_{(e,u)\in A} \delta_{eu,r}^{o,q}(b_{\mathrm{d}}^q d_{eu} + b_{\mathrm{c}}^q t_{eu}(\varsigma_{eu})),\ \forall r \in R_{ij}^{o,q}, i \in I,\ j \in J. \tag{48}$$

For vacant taxis, we need to add the waiting time cost into the operating cost. Hence, the operating cost for a vacant type q taxi that travels from zone j to zone i along path $r \in R_{ij}^{o,q}$ in search of customers can be expressed as

$$OC_{ji,r}^{v,q} = \sum_{(e,u)\in A} \delta_{eu,r}^{v,q}(b_{\mathrm{d}}^q d_{eu} + b_{\mathrm{c}}^q (t_{eu}(\varsigma_{eu}) + w_i^q)),\ \forall r \in R_{ji}^{v,q}, q\in Q,\ i \in I,\ j \in J. \tag{49}$$

Summing up $OC_{ij,r}^{o,q}$ and $OC_{ij,r}^{v,q}$ along all paths and utilizing Eqs. (20), (21), and (29), we have the total operating cost as

$$TOC = \sum_{q\in Q}\sum_{i\in I}\sum_{j\in J}\{T_{ij}^{o,q}(b_{\mathrm{d}}^q h_{ij,\mathrm{d}}^{o,q} + b_{\mathrm{c}}^q h_{ij,\mathrm{t}}^{o,q}) + T_{ji}^{v,q}(b_{\mathrm{d}}^q h_{ji,\mathrm{d}}^{v,q} + b_{\mathrm{c}}^q (h_{ji,\mathrm{t}}^{v,q} + w_i^q))\}. \tag{50}$$

With Eqs. (47) and (50), we rewrite Eq. (44) as

$$TPS = \sum_{q\in Q}\sum_{i\in I}\sum_{j\in J}\{T_{ij}^{o,q}((b_1^{o,q} - b_{\mathrm{d}}^q)h_{ij,\mathrm{d}}^{o,q} + (b_2^{o,q} - b_{\mathrm{c}}^q)h_{ij,\mathrm{t}}^{o,q}) - T_{ji}^{v,q}(b_{\mathrm{d}}^q h_{ji,\mathrm{d}}^{v,q} + b_{\mathrm{c}}^q (h_{ji,\mathrm{t}}^{v,q} + w_i^q)\}. \tag{51}$$

2.5. Mathematical program for the TSAD problem

Denote $\mathbf{\Delta} = (\mathbf{X}^T, \mathbf{y}^T, \mathbf{s}^T, \mathbf{\Phi}^T)^T$ and $\mathbf{\Gamma} = (O_i^q, D_j^q, T_{ij}^m),\ \forall m \in M, q \in Q, i \in I, j \in J$. The TSAD problem can be formulated as follows.

$$\max_{\mathbf{\Delta},\mathbf{f},\mathbf{\Gamma}} SW = TCS + TPS \tag{52}$$

subject to

Combined network equilibrium constraints: Eqs. (1)-(30);

Design constraints: Eqs. (31)-(40),

in which TCS and TPS are respectively defined by Eqs. (42)-(43) and (51).

3. A greedy heuristic for solving the TSAD problem

Generally, the mixed integer nonlinear formulation of the regulatory problem implies that the computational time required to solve the TSAD problem by exact methods such as enumeration or branch and bound methods increases exponentially with the network size. Hence, we propose a greedy heuristic to solve the TSAD problem, especially when the problem size is large. The proposed greedy heuristic determines the service areas of all types of restricted area taxis by sequentially adding each type into the network. For each type to be added, the initialization phase is first performed to generate a set of solutions, in which each solution represents a service area that consists of only one node in the network, and the corresponding social welfare is also calculated. Then, the expansion phase and the local search phase are performed on every solution in the solution set to get new solutions until the stopping criteria are satisfied. Lastly, the solution with the maximum social welfare as the service area of the selected type is outputted. The three phases are repeated until all types of restricted area taxis are added to the network. The calculation of social welfare for each given TSAD during the solution procedure requires solving the CNEP with the solution method proposed by Wong et al. (2008).

3.1. The initialization phase

In this heuristic, a solution for type q restricted area taxis is represented by a binary vector $\mathbf{X}^q = \left(X_l^q\right)_{l \in V}$. In the initialization phase, this heuristic generates a set of solutions, denoted as $\Omega = \{\mathbf{X}^q \mid X_i^q = 1, X_l^q = 0, \forall i \in V, \forall l \in V - \{i\}\}$. Each solution in Ω represents the service area design that consists of only one arbitrarily chosen node in the network. Fig. 3 demonstrates the generation of Ω using a simple nine-node network, in which k indicates the index of the solution. Afterward, the social welfare associated with each solution is calculated, denoted by $SW_k, k = 1, 2, ..., |\Omega|$, assuming that the previously determined service areas of other types of restricted area taxis are unaltered.

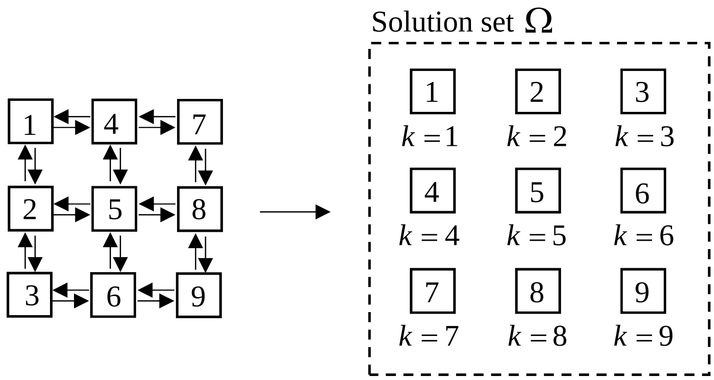

Fig. 3 Illustration of the initialization phase

3.2. The expansion phase

After the initialization phase, each solution in Ω is selected to perform the expansion phase. The expansion phase aims at improving the selected solution by expanding the corresponding service area. Each expansion includes all nodes adjacent to the currently selected service area to obtain a new one. If the social welfare increases after the expansion, this phase updates the solution to the new one and continues expanding the region. Otherwise, it keeps the current solution and selects the next solution in Ω. The expansion phase is repeated until all solutions in Ω have been selected. The main steps of the expansion phase are described in Table 1 and Fig. 4 illustrates how to improve a solution by the expansion with the first ($k=1$) and the fifth ($k=5$) solutions in Fig. 3 as examples. After the expansion phase, each solution in Ω will be updated to a new solution.

Table 1 Main steps of the expansion phase

The expansion phase	
Inputs: A solution set Ω and the social welfare associated with each solution $SW_k, k=1,2,\ldots,\lvert\Omega\rvert$	
1:	**for** $k=1,\ldots,\lvert\Omega\rvert$, **do**
2:	repeat
3:	Include all nodes adjacent to the sub-network represented by $\mathbf{X}_k^q$ to obtain a new solution $\mathbf{X}_k'^q$
4:	Calculate the social welfare SW_k' associated with $\mathbf{X}_k'^q$

if $SW'_k > SW_k$ **then**

Set $\mathbf{X}_k^q = \mathbf{X}_k'^q$

Set $SW_k = SW'_k$

else

Set $k = k+1$

end if

until all nodes in the network are included to $\mathbf{X}_k^q$

Set $k = k+1$

end for

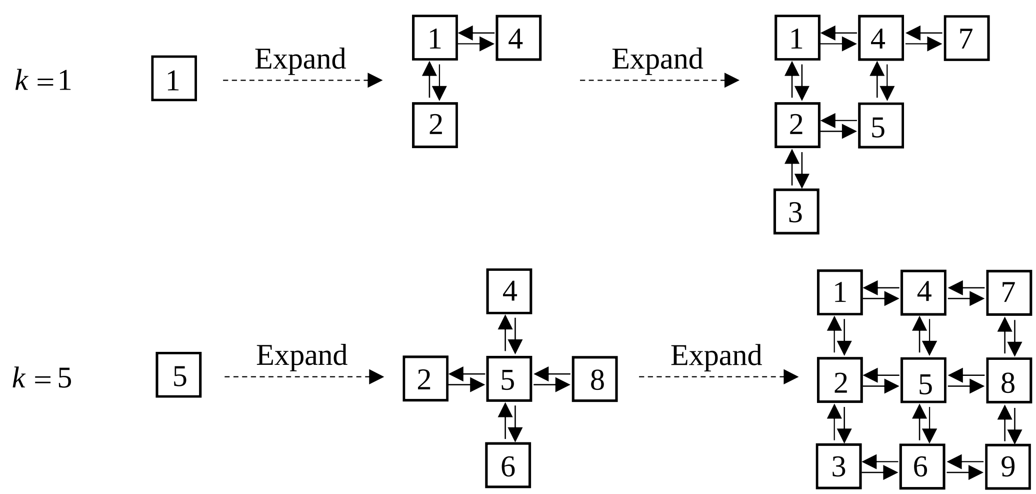

Fig. 4 Illustration of improving initial feasible solutions by expanding their service areas

3.3. The local search phase

The local search phase is performed on the updated solution set Ω after the expansion phase. It is proposed to fine-tune each solution $\mathbf{X}_k^q, k = 1, ..., |\Omega|$ in Ω and to finally determine the service area of the type of restricted area taxi under consideration. The main steps of the local search phase are displayed in Table 2. For each solution $\mathbf{X}_k^q, k = 1, ..., |\Omega|$ obtained from the expansion phase, we first put all nodes that are adjacent to $\mathbf{X}_k^q$ into a pool. Next, we form a new solution by picking a combination of nodes from the pool and changing the corresponding X_l^q in $\mathbf{X}_k^q$ to 1. Then, we try all possible combinations. Fig. 5 shows how to generate the service areas corresponding to all possible new solutions with a given $\mathbf{X}_k^q$. Afterward, we find the best new solution that improves the social welfare the most compared with $\mathbf{X}_k^q$. We replace $\mathbf{X}_k^q$ with the best new solution. After improving each $\mathbf{X}_k^q, k=1,\ldots,|\Omega|$, we output the one with the maximum social welfare as the final service area of the concerned type of restricted area taxi.

Table 2 Main steps of the local search phase

The local search phase
Inputs: The updated solution set Ω and the social welfare associated with each solution $SW_k, k=1,2,\ldots,|\Omega|$ obtained from the expansion phase
for $k=1,\ldots,|\Omega|$, **do**
Determine Γ, which is the set of nodes adjacent to $\mathbf{X}_k^q$
repeat
Pick one combination of nodes from Γ and change the corresponding X_l^q in $\mathbf{X}_k^q$ to 1 to form a new solution $\mathbf{X}_k'^q$
Calculate the social welfare SW_k' associated with $\mathbf{X}_k'^q$
if $SW_k' > SW_k$ **then**
Set $\mathbf{X}_k''^q = \mathbf{X}_k'^q$
Set $SW_k'' = SW_k'$
end if
until all combinations of nodes from Γ are selected
if $SW_k'' > SW_k$ **then**
Set $\mathbf{X}_k^q = \mathbf{X}_k''^q$
Set $SW_k = SW_k''$
end if
Set $k=k+1$
end for
Output the solution with the maximum social welfare as the final service area of

the type of restricted area taxi under consideration

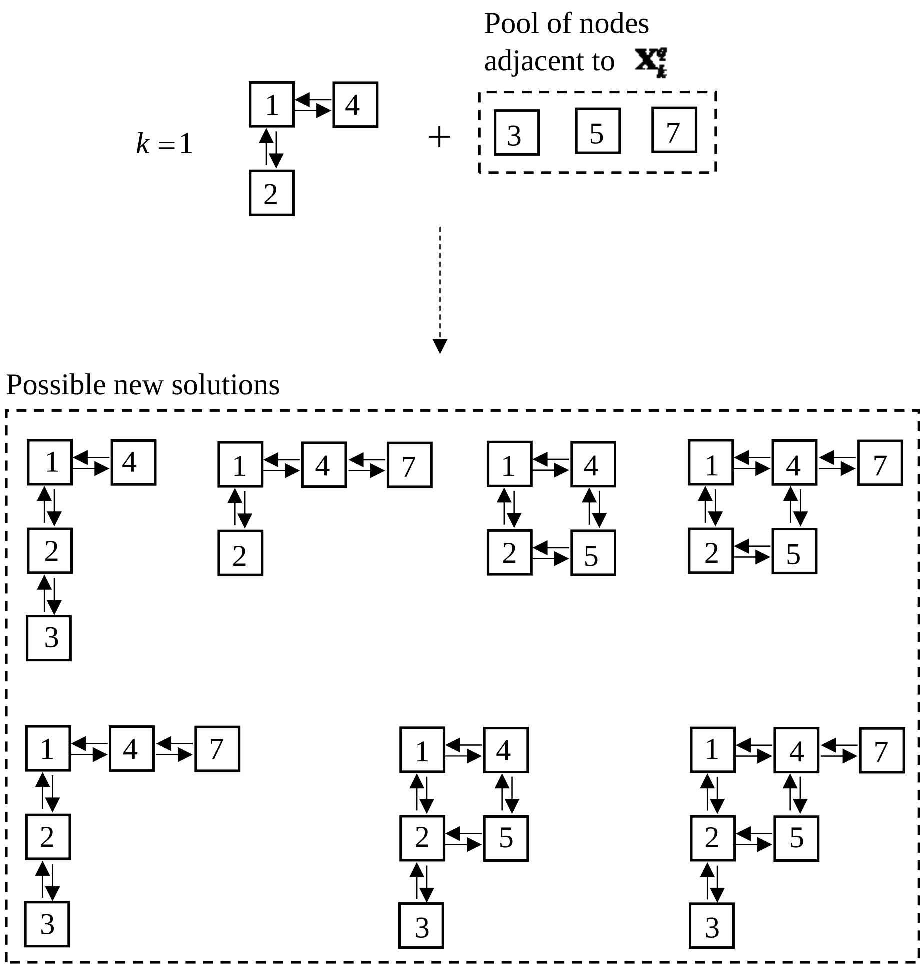

Fig. 5 Illustration of the service areas corresponding to all possible new solutions in the local search phase[1]

4. Numerical examples

In this section, three examples are given, in which the first one shows the performance of the greedy heuristic versus that of genetic algorithm (GA) and the enumeration method while the second example provides insights into TSAD using a small network. Finally, a case study is performed to show that the insights into TSAD from the small network can be scalable to a large network. The heuristic and GA were coded in MATLAB 2018a and were run on a Dell OptiPlex 7050 desktop with an Intel Core i7-7700 CPU@3.6 GHz and 64.0 GB RAM. For

[1] The TSAD {1,2,4} is obtained from the expansion phase assuming that the social welfare declines after the expansion from {1,2,4} to {1,2,3,4,5,7}

each given design, the equilibrium path flows and O-D flows of each class are obtained by solving the corresponding variational inequality (see Wong et al., 2008).

4.1. Performance of the developed greedy heuristic

The performance of the developed greedy heuristic is investigated in comparison to that of GA and the enumeration method using four network examples, in which there are three grid networks with their sizes as 4 × 4 (Fig. 6), 5 × 5 (Fig. 7), and 6 × 6 (Fig. 8) together with a real-world network based on the HKSAR, China (Fig. 9, the indices of zones are shown in circles). The number of taxi types in all the network examples is assumed to be 3, with 1 type of normal taxi and 2 types of restricted area taxis (denoted as type A and type B). In each network example, the fleet size of each type of taxi is given in Table 4. For the three grid network examples, the total travel demand originated from each node is assumed to be 1000 veh/h, with the destinations being evenly distributed among all nodes in the network. The travel impedance functions for all links are given as Eq. (12), where the free-flow travel times of all links are 0.06 h and the capacities of all links are 3000 veh/h. The lengths of all links are 3 km. The other input parameters in this sub-section are shown in Table 3.

The Hong Kong (HK) network was constructed based on the statutory planning zones defined by the Town Planning Board of the Hong Kong Government (see Fig. 9). This network consists of 125 zones and 440 links. The intra-zonal and inter-zonal demands (trips/h) were derived with reference to the AM peak-hour trip data recorded in the Travel Characteristic Survey (HKSAR, 2002) and were multiplied by the ratio of the population of Hong Kong between 2018 and 2002, which is around 1.11, to capture the most up-to-date travel demand pattern of the city. The free-flow speeds and capacities of all links are assumed to be 50 km/h and 10000 veh/h, respectively. As for the areas of zones $\Phi_i, \forall i \in V$, the values were obtained from the report of Land Supply in Hong Kong (Legislative Council, 1997) and the parameter $\eta = 0.01$ is assumed to be the same for all zones. The link lengths, total travel demand, and the areas of zones are available from http://web.hku.hk/~ceszeto/LiSzeto_TSAD_data.zip.

Table 3 Input parameters for the example

Input parameters	Values
Users' value of time	$b_0 = 100$ (\$/h)

Value of customer waiting time for taxis	$b_1 = 200\ (\$/\mathrm{h})$
Mileage cost to a user of a non-taxi mode	$b^n = 5\ (\$/\mathrm{km})$
Mileage charge to a taxi passenger	$(b_1^{o,q}, q \in Q) = (6, 4, 2)\ (\$/\mathrm{km})$
Congestion-based charge to a taxi passenger	$(b_2^{o,q}, q \in Q) = (80, 50, 40)\ (\$/\mathrm{h})$
Mileage operating cost	$(b_\mathrm{d}^q, q \in Q) = (1, 1, 1)\ (\$/\mathrm{km})$
Hourly operating cost	$(b_\mathrm{c}^q, q \in Q) = (60, 30, 20)\ (\$/\mathrm{h})$
Dispersion coefficient for the upper-level logit mode choice	$\beta_1 = 0.03\ (1/\$)$
Dispersion coefficient for the lower-level logit mode choice	$\beta_2 = 0.06\ (1/\$)$
Dispersion coefficient for vacant taxi search behavior	$(\theta^q, q \in Q) = (0.25, 0.25, 0.25)\ (1/\$)$
Parameter for the relationship of customer and taxi waiting times of zones (not applicable to the HK network example)	$(\eta\Phi_i, i \in I) = 2\ (\mathrm{veh \cdot h})$

Table 4 Taxi fleet sizes

Network	4×4	5×5	6×6	HK network
Normal taxis	1000	5500	9500	15000
Restricted area taxis				
Type A	400	1200	2300	3800
Type B	200	600	800	1200

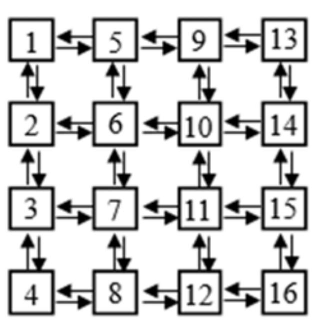

Fig. 6. A 4 × 4 grid network

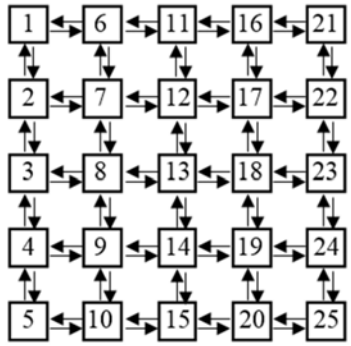

Fig. 7 A 5 × 5 grid network

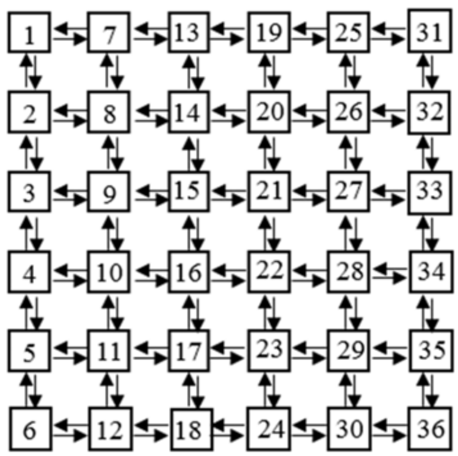

Fig. 8 A 6 × 6 grid network

The solution procedure of GA includes initialization, natural selection, crossover, and mutation. A solution was encoded as a binary vector $\mathbf{X}=\left(X_l^q\right)_{l\in V,q\in Q-\{1\}}$. New solutions (i.e., chromosomes) were generated in the initialization phase by randomly picking nodes in the network. Single-point crossover was implemented with the crossover rate as 0.8. The mutation rate was set at 0.05. For any infeasible solution encountered during the solution procedure, we discarded it and regenerated a new one in the same way as the initialization phase to complement the population. The fitness function is the same as the objective function. The population size of GA is equal to the corresponding number of nodes $|V|$ and the running time of GA was set to be equal to that required by the heuristic. For the three grid network examples, the enumeration method that tries all feasible solutions one by one to obtain the best solutions as benchmarks was also applied.

Table 5 shows the solutions to the grid network examples obtained by the heuristic, GA, and the enumeration method, together with the corresponding computational times. The results show that the solutions to the grid network examples produced by the heuristic and the enumeration are identical, while GA can also obtain the same results as the heuristic in the 4 × 4 and 5 × 5 network examples but fails to do so in the 6 × 6 network example. This is because on one hand, the probability that GA encountered infeasible solutions increases as the network size increases, which resulted in more time to discard the infeasible solutions and to regenerate new ones and thus a low convergence speed of the population. On the other hand, extra time is needed for GA to check the feasibility of the solutions with respect to the design constraints, while the solutions produced by the heuristic can always satisfy the design constraints. Note that due to the symmetric topologies and demand patterns, there are multiple optimal solutions to the three grid network examples. For instance, the optimal solution of types A and B restricted area taxis to Fig. 7 are respectively {7,8,9,12,13,14,17,18,19} and {8,12,13,18}, which means that the following solutions are also optimal: 1) {7,8,9,12,13,14,17,18,19} and {8,13,14,18}; 2) {7,8,9,12,13,14,17,18,19} and {12,13,14,18}; 3) {7,8,9,12,13,14,17,18,19} and {8,12,13,14}.

GA and the heuristic were also run to solve the HK network example with the same computational time and the summary of the results are displayed in Table 7, in which we can see that the result produced by the heuristic (the network illustrations of the results can be found in Fig. 10 and Fig. 11) is superior to that by GA with the same computational time. Moreover, Fig. 12 shows the convergence speed of the two algorithms for solving the HK network example, which shows that the heuristic can always find a better result than GA.

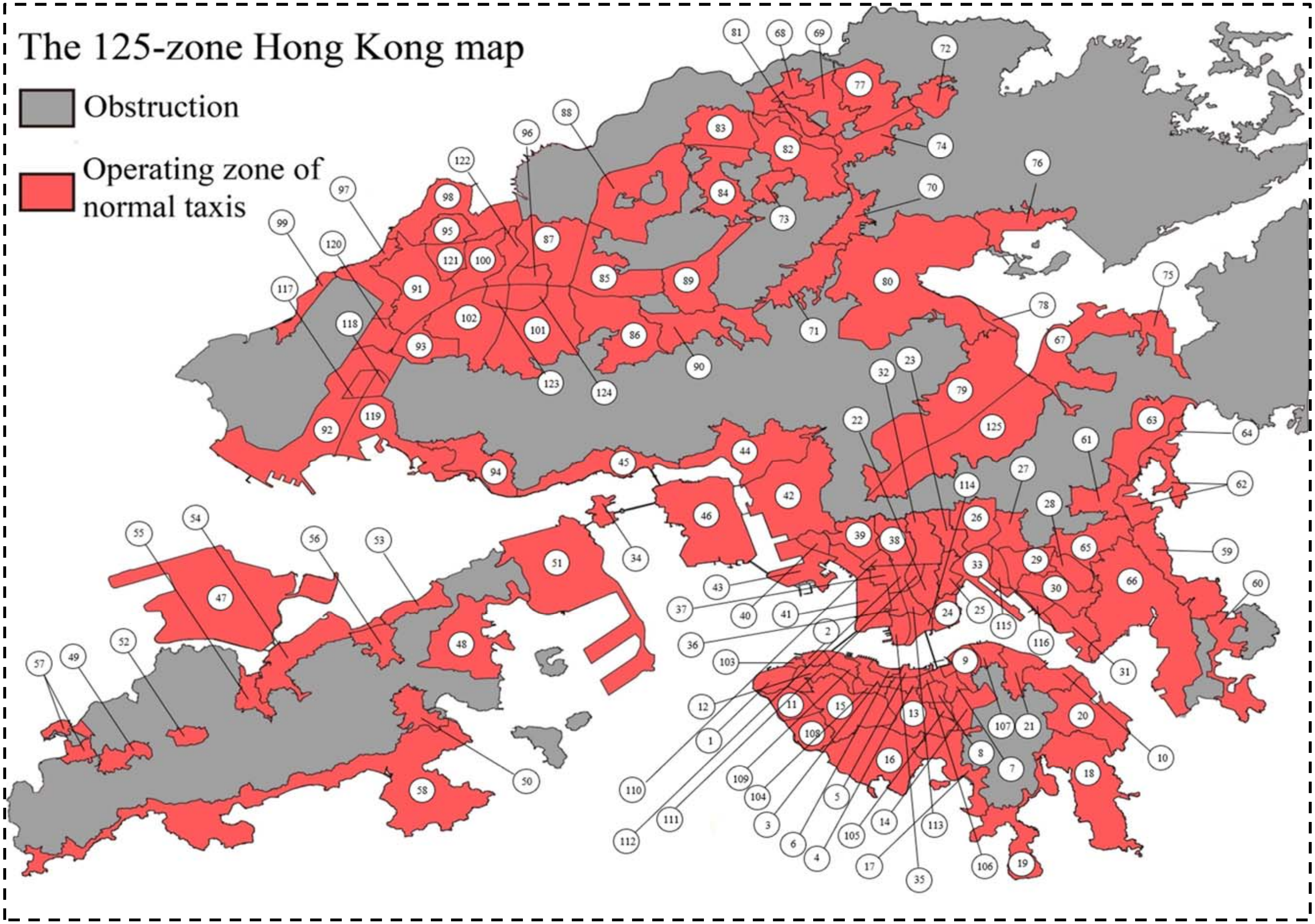

Fig. 9 The 125-zone Hong Kong network (Town Planning Board, 2018)

Table 5 Summary of the results of the small network examples in Section 4.1

Network	4 × 4	5 × 5	6 × 6
Optimal solution ({nodes included})			
-Proposed greedy heuristic			
Type A	{2,5,6,7,10,11}	{7,8,9,12,13,14,17, 18,19}	{9,10,15,16,20,21, 22,23,27,28}
Type B	{6,7,10}	{8,12,13,18}	{9,15,16,21,22,27}
-GA			
Type A	{2,5,6,7,10,11}	{7,8,9,12,13,14,17, 18,19}	{8,9,10,11,13,14,1 5,16,20,21}
Type B	{6,7,10}	{8,12,13,18}	{8,9,10,14,15,21}
-The enumeration method			
Type A	{2,5,6,7,10,11}	{7,8,9,12,13,14,17, 18,19}	{9,10,15,16,20,21, 22,23,27,28}
Type B	{6,7,10}	{8,12,13,18}	{9,15,16,21,22,27}
Computational times (hrs)			
-Proposed greedy heuristic	0.26	1.41	4.04
-GA	0.26	1.41	4.04
-The enumeration method	0.81	56.13	398.67

Table 6 Summary of results of the Hong Kong network example

Network	Optimal solution Type A	Optimal solution Type B	Social welfare ($\times 10^8$\$)	Computational times (hrs)
Greedy heuristic	{3,4,5,6,7,9,12,22,23, 24,25,26,27,32,33,35, 36,37,38,39,41,106,1 09,110,111,112,113,1 14,115,116}	{22,23,24,25,32,36 ,37,38,39,110,111, 112,113,114,125}	-8.31	13.75
GA	{1,2,3,4,5,6,7,8,9,12, 22,23,24,25,32,35,36,	{1,2,3,4,5,6,7,8,9,1 2,13,35,103,104,10	-8.74	13.75

37,38,39,41,103,104,	6,109,114}
106,109,110,111,112,	
113,114}	

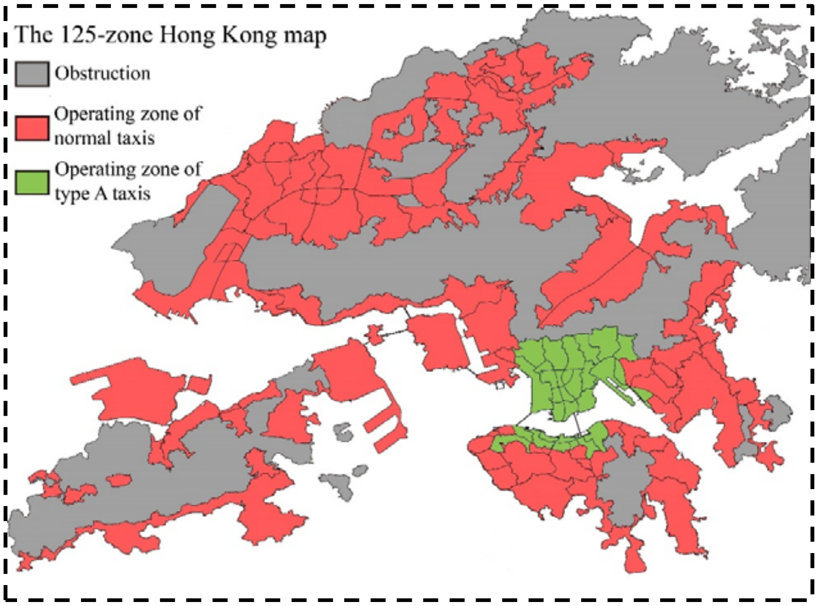

Fig. 10 Operating zones of type A taxis obtained by the heuristic

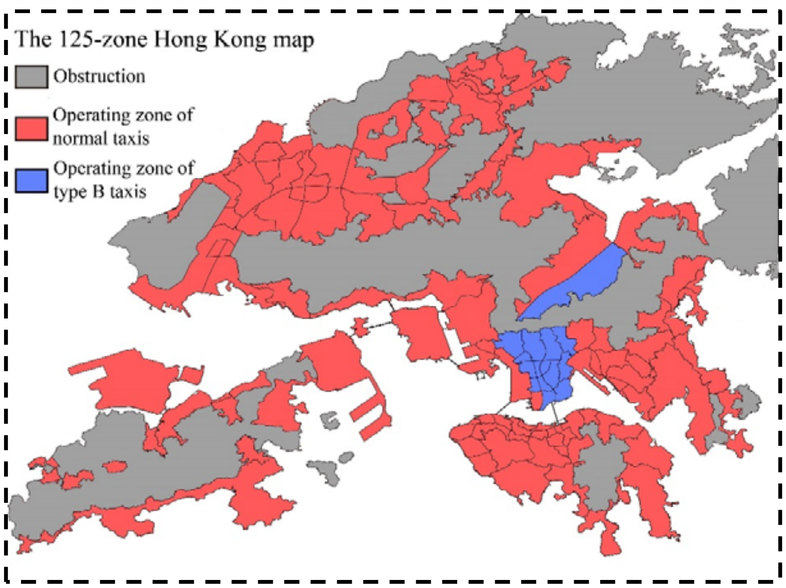

Fig. 11 Operating zones of type B taxis obtained by the heuristic

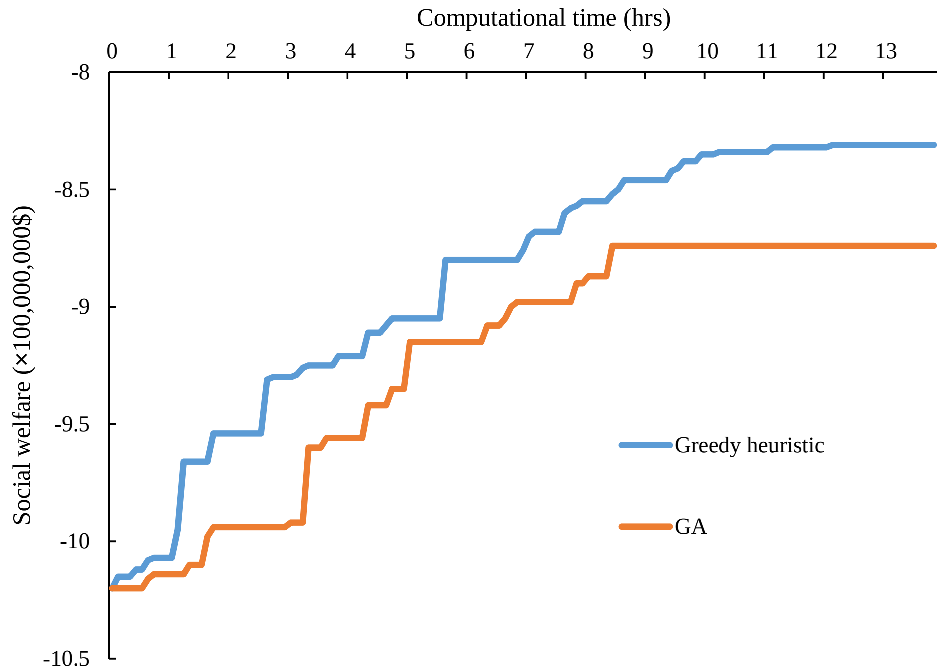

Fig. 12 Convergence graph of the HK network example

4.2. Insights into TSAD

This section provides the strategic analysis of TSAD. A 4-node network as shown in Fig. 13 is adopted for the analysis. The total travel demand $\bar{D}_{ij}$, $\forall i \in I, j \in J$, the link travel distance d_{eu}, $\forall (e,u) \in A$ are given in Table 7 and Table 8, respectively. The free-flow speeds of all links are assumed to be 50 km/h. Unlike Section 4.1, we only assume one type of restricted area taxi operating in the example network because this allows us to easily enumerate all possible TSADs and to ensure the optimality of the TSAD. Obviously, there are in total 15 possible TSADs in the example network, which are indexed from 1 to 15 as shown in Table 9 for a better presentation later. Note that case 15, which includes all four nodes in the example network, is a special case in which the restricted area taxis are allowed to operate anywhere in the network like the normal taxis.

For comparative analysis, we will also investigate the TSAD from a different perspective, namely the total absolute difference in customer waiting time (TADCWT). The TADCWT is the sum of the absolute difference in the average customer waiting time between each pair of zones in the network, where the average customer waiting time in a zone is obtained by dividing the sum of the customer waiting times for all taxi types by the total taxi demand in that zone. The TADCWT reflects the equity in customer waiting time. The lower the value is, the more equity the design is. Therefore, the best TSAD in view of equity in customer waiting time is expected to be the one with the lowest TADCWT.

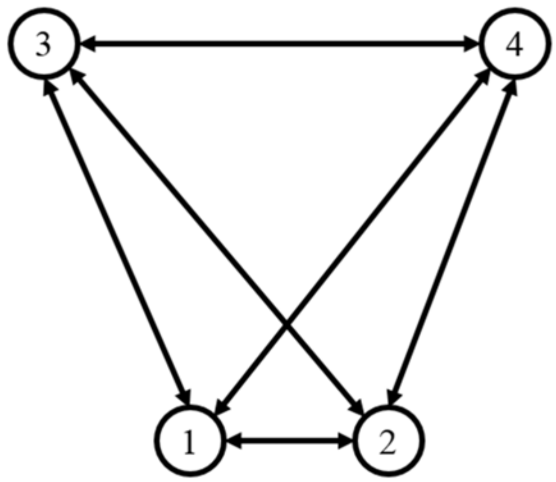

Fig. 13 The example network

Table 7 Total travel demand $\bar{D}_{ij}$ (persons/h)

	1	2	3	4	Total
1	1003	2287	573	512	4375
2	2195	1210	604	585	4594
3	601	613	1600	208	3022
4	532	598	197	1589	2916
Total	4331	4699	2974	2894	14907

Table 8 Link travel distances d_{eu} (km)

	1	2	3	4
1	2	3	7	9
2	3	2	9	7
3	7	9	2	7
4	9	7	7	2

Table 9 Indices for different TSADs

Index	TSAD	Index	TSAD
1	{1}	9	{2,4}
2	{2}	10	{3,4}
3	{3}	11	{1,2,3}
4	{4}	12	{2,3,4}
5	{1,2}	13	{1,2,4}
6	{1,3}	14	{1,3,4}
7	{1,4}	15	{1,2,3,4}
8	{2,3}		

4.2.1. Fleet size variation

The effects of taxi fleet size on TSAD are investigated in two ways. First, we look into how the ratio of two taxi types affects the best TSAD in terms of social welfare maximization and equity. To achieve this, we fixed the total fleet size at 3200 (veh) and considered three different combinations of fleet sizes of the two taxi types, which are represented as $I(3200,0)$, $II(2500,700)$, and $III(2310,890)$. I, II, and III are the indices of fleet size combinations. The first number in each bracket is the fleet size of normal taxis while the second one is the fleet size of restricted area taxis. Note that combination I is a special case in which there is no restricted area taxi in the network, meaning that the hierarchical mode choice model collapses into a simple binary logit mode choice model between non-taxis and normal taxis with the new dispersion coefficient assumed to be $\beta_1^* = 0.03\ (1/\$)$. All other unspecified input parameters take the same values as those in Table 3.

Fig. 14 plots the social welfares of different TSADs under different combinations. Clearly, the social welfare does not change with different designs under combination I since there is no restricted area taxi. However, we can observe that under combinations II and III, their best TSADs exist and are {1,2} and {1,2,3,4}, respectively, and the corresponding social welfares are both larger than the social welfare under combination I. This can be explained as follows. By replacing normal taxis with restricted area taxis that have a lower fare level, the consumer surplus rises and the producer surplus falls. However, the rise is greater than the fall so the resulting social welfare increases. It also implies that introducing restricted area taxis to the market can be better than the single-type case in terms of social welfare maximization. Meanwhile, the best TSAD under combination *III* indicates no service area restriction on restricted area taxis and the corresponding optimal social welfare is better than that under combination *II*. This tells us that on one hand, the best TSAD may differ as the proportion of restricted area taxis to the total fleet size varies. On the other hand, having a service area restriction on restricted area taxis is not a must since the corresponding social welfare may be lower than that in the case without such restriction. Nevertheless, the above conclusions are dependent on the demand elasticity for taxi traffic. If the demand for taxi traffic was inelastic (i.e., β_1 is a small value), introducing restricted area taxis would not attract sufficient passengers so that the increase in consumer surplus would not be greater than the decrease in producer surplus, which would, therefore, result in a decline in social welfare. Therefore, accurate calibration of parameters is necessary and further optimization can be conducted on both the fleet size proportion and the service areas of restricted area taxis simultaneously to draw a conclusion for the actual situation under consideration.

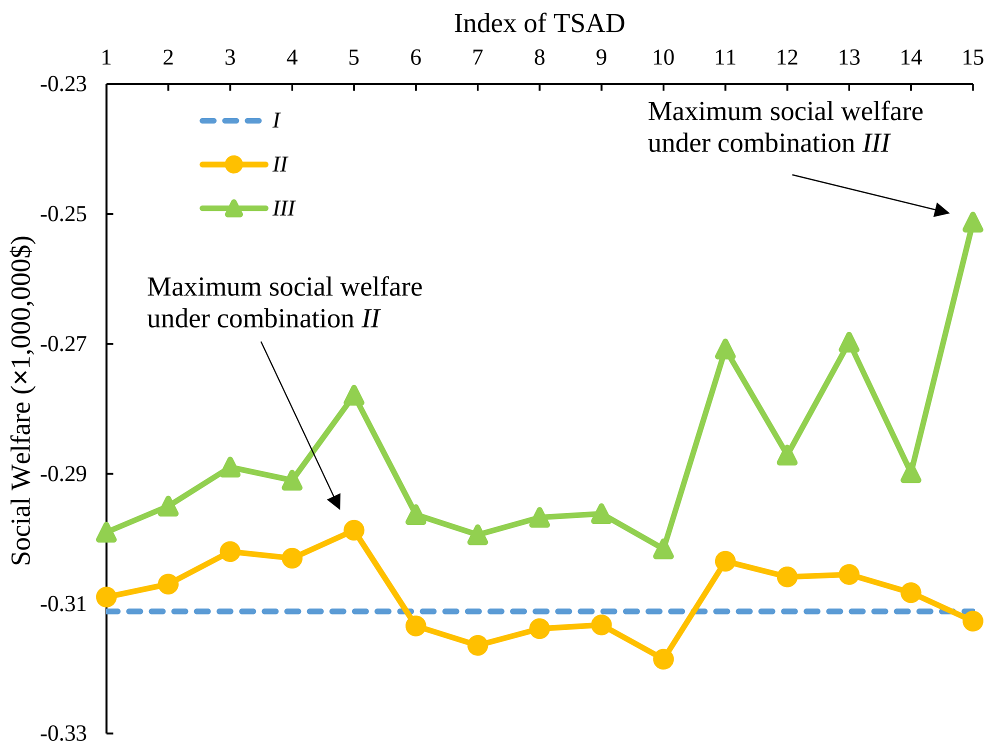

Fig. 14 Social welfares of different TSADs under different fleet size combinations

Fig. 15 shows that in terms of minimizing the TADCWT, the best TSADs under combinations *II* and *III* are the same, which is {3,4}, with the corresponding values of the TADCWT roughly as 0.078 (hr) and 0.07 (hr), respectively. Note that the two values are both lower than the TADCWT under combination I, which is around 0.22 (hr). To explain this, we list out the average customer waiting time in each zone under combination I and in case 6 under combination II for example (see Table 10), in which we see that under combination I, zones 1 and 2 have lower average customer waiting times than zones 3 and 4. This is because of the clustering of vacant taxis. Under combination I (with normal taxis only), zones 1 and 2 are higher demand areas for (and more profitable areas to) taxi drivers than zones 3 and 4 so that vacant taxis tend to cluster in zones 1 and 2, leading to longer customer waiting times in zones 3 and 4. Unlike combination *I*, in case 6 under combination II, restricted area taxis operate only within zones 3 and 4. The average customer waiting times in zones 1 and 2 are higher because of the decline in the fleet size of normal taxis compared to combination I. Meanwhile, the average customer waiting times are lower in zones 3 and 4 since the availability of restricted area taxis is better ensured within the two zones by TSAD. Therefore, the resulting TADCWT is lower than that in combination I.

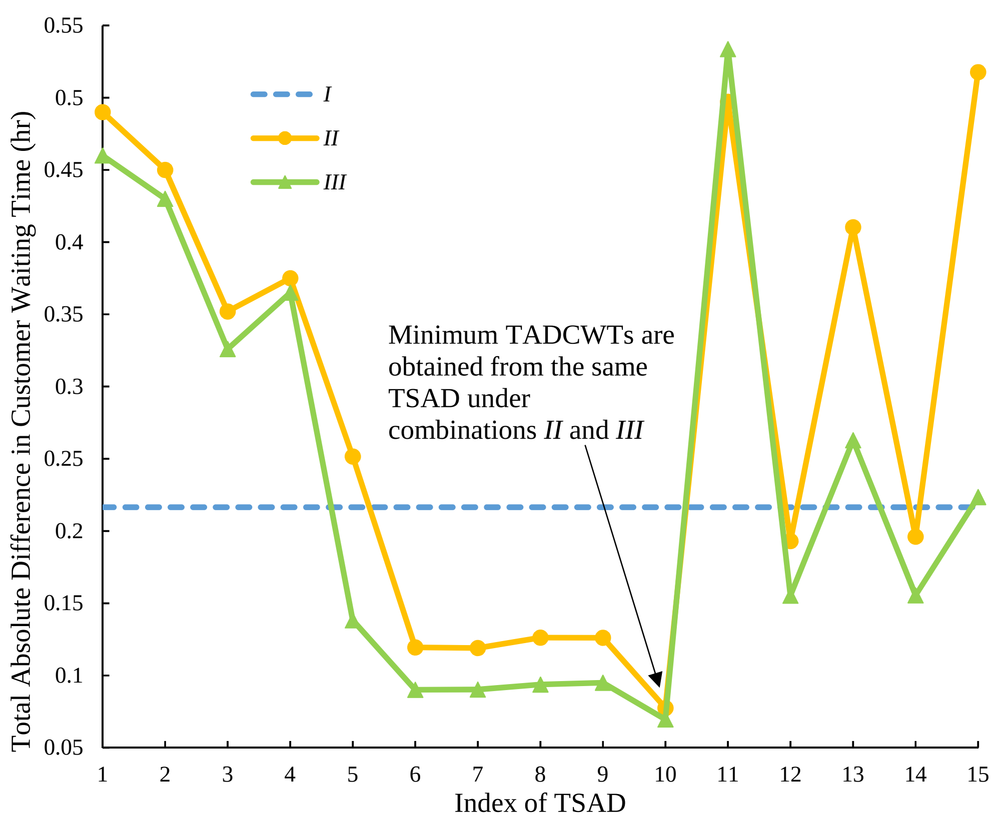

Fig. 15 The TADCWTs of different TSADs under different fleet size combinations

Comparing the results shown in Fig. 14 and Fig. 15, we see that under both combinations II and III, the design {3,4} is the worst case in terms of social welfare maximization but the best one in terms of equity in customer waiting time. This means that the best TSAD may also vary with different design objectives and tradeoffs exist between social welfare maximization and customers' equity. To avoid a large difference in customer waiting times among zones, equity or waiting time constraints should be incorporated into the proposed TSAD model.

Table 10 Average customer waiting time for taxis in each zone (combination I and case 6 under combination *II*)

Zone	1	2	3	4
Combination *I*	0.034	0.042	0.088	0.081
Combination *II*				
{3,4}	0.051	0.055	0.064	0.066

Next, we examine the impact of total taxi fleet size on the best TSAD with a fixed ratio of the fleet sizes of normal taxis to restricted area taxis. This ratio is the same as that of combination *III*, i.e., 2310:890. Three scenarios were designed with three different total fleet sizes: 3200 (veh), 2830 (veh), and 2500 (veh). All other input parameters remain unchanged. We calculated the corresponding single-type cases with only normal taxis in the market as benchmarks.

Fig. 16 plots the social welfare of each TSAD under different total taxi fleet sizes, in which the dash lines represent the single-type cases. The figure shows that the best TSADs under the aforementioned total taxi fleet sizes in terms of maximizing social welfare are, respectively, {1,2,3,4}, {1,2,3}, and {1,2}. Each best TSAD is better than the corresponding single-type case, which can also be explained by the fact that the consumer surplus rises with the introduction of restricted area taxis and the rise is greater than the fall in producer surplus. Besides, the area of the best TSAD shrinks as the total taxi fleet size decreases. This can be explained by the increase in customer waiting time with fewer taxis in service so that both the consumer and producer surpluses decrease.

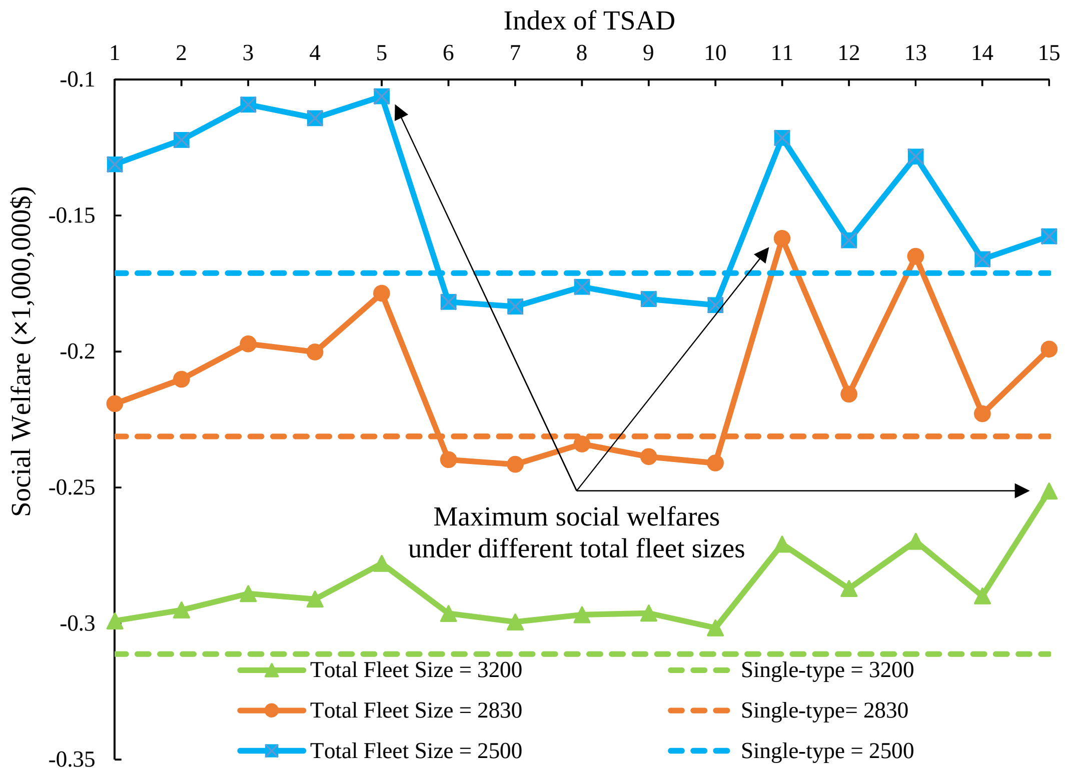

Fig. 16 Social welfares of different TSADs under different total taxi fleet sizes

Fig. 17 depicts the TADCWT of each TSAD under different total fleet sizes. It shows that the fleet size only affects the magnitude of the customer waiting time in each zone because the customer waiting time in each zone increases as the total fleet size decreases. However, the best TSADs under different total fleet sizes are the same, namely the design {3,4}. We can explain this as a consequence of the unchanged demand distribution despite the change in fleet size. The distribution of vacant taxis is only affected by the demand distribution. As the demand distribution remains the same, the distribution of vacant taxis remains the same. To conclude, the best TSAD in terms of equity may not change as total taxi fleet size varies.

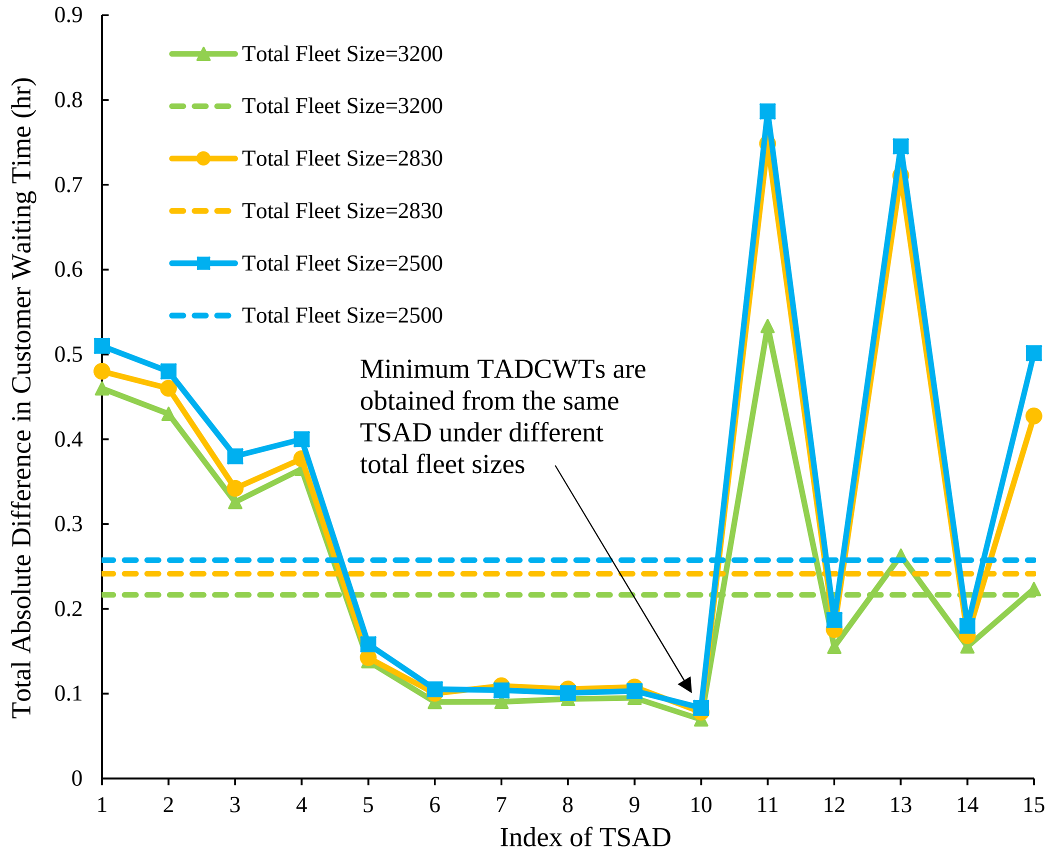

Fig. 17 The TADCWTs of different TSADs under different total fleet sizes

4.2.2. Total travel demand level and distribution

In this section, we analyze the impacts of demand level and distribution on the best TSAD. The parameter setting is based on that under combination $III(2310,890)$ in Section 4.2.1, while the modifications on total travel demand level and distribution are given separately.

To obtain a different distributing pattern of the total travel demand, we switched the positions of two pairs of elements in Table 7, namely $\bar{D}_{12}$ and $\bar{D}_{34}$, $\bar{D}_{21}$ and $\bar{D}_{43}$, so that travel demand between nodes 3 and 4 now becomes more intensive than that between any other pair of nodes. Then, we calculated the social welfare and TADCWT of each TSAD and obtained Fig. 18 and Fig. 19, in which the results of the single-type situations with only normal taxis in the market are also given.

Fig. 18 shows the social welfare of each TSAD under the modified total travel demand distribution. We can observe that the best TSAD shrinks from {1,2,3,4} (see the green line in Fig. 14) to {2,3,4} after the demand distribution changes. This can be explained by the longer travel distance between nodes 3 and 4 than that between nodes 1 and 2 that leads to longer occupied taxi hours and hence a lower taxi demand served. This further leads to a decrease in both consumer and producer surpluses and hence the social welfare falls.

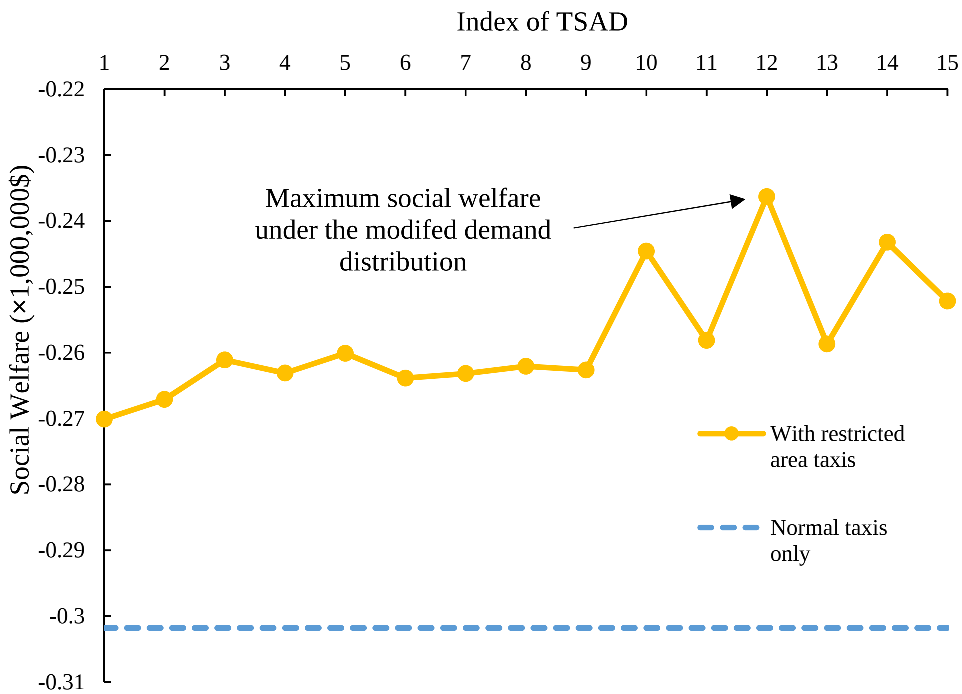

Fig. 18 Social welfares of different TSADs under the modified demand distribution

Fig. 19 plots the TADCWTs of different TSADs under the modified demand distribution. It is observed that the lowest TADCWT is obtained from a different TSAD ({1,4}) compared with the result in Fig. 15 ({3,4}). As the demand distribution changes, the clustering pattern of vacant taxis changes accordingly to satisfy the demand. As now the travel demand between nodes 3 and 4 is more intensive, more vacant taxis tend to cluster in the two nodes so that customers in nodes 1 and 2 suffer from a longer waiting time for taxis.

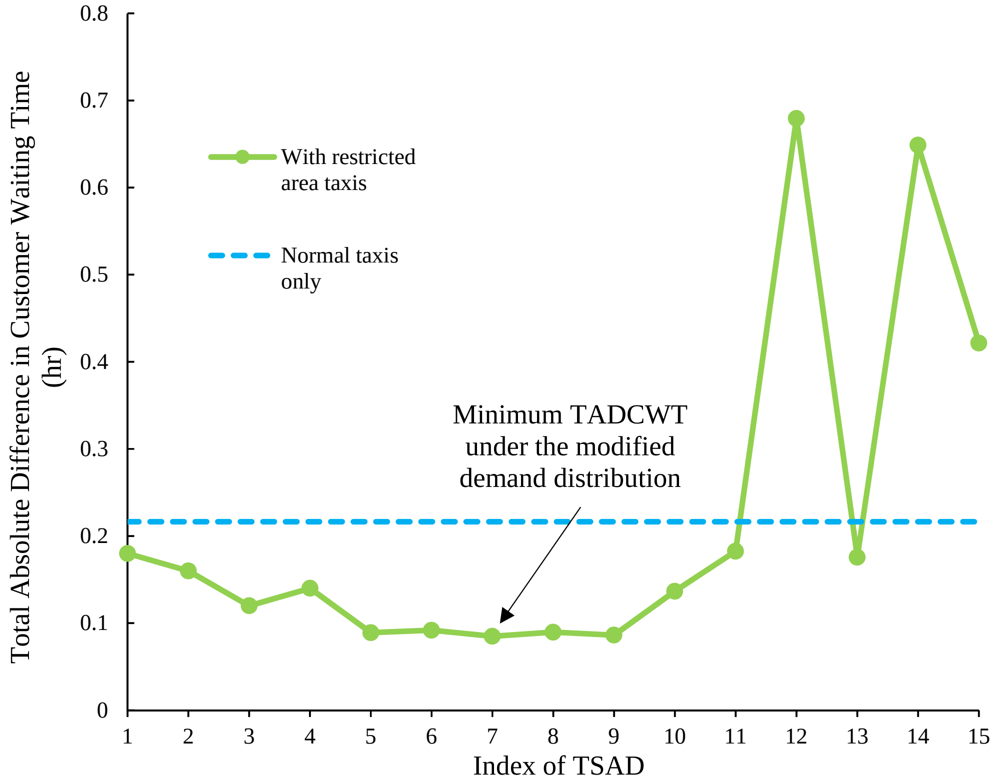

Fig. 19 The TADCWTs of different TSADs under the modified demand distribution

Then, we look into the impact of total travel demand level on the best TSAD. The modification of demand distribution is not considered here and the original total travel demand pattern as shown in Table 7 was used. Three cases were considered. In each case, all elements of Table 7 were multiplied by the same demand scaling factor. The factor took the values of 1, 1.15, and 1.3.

Fig. 20 plots the social welfare of each TSAD under different total travel demand levels, from which we learn that the best TSAD shrinks from {1,2,3,4} to {1,2,3} and eventually to {1,2} as the total travel demand level increases. This can be explained as follows. When the total travel demand level is relatively low, setting no service area restriction to both taxi types benefits all travelers with an alternative taxi choice. However, as the total travel demand level rises, the customer waiting times in all zones in {1,2,3,4} increases, leading to decreases in both the consumer and producer surpluses. Hence, a smaller service area of restricted area taxis, i.e., {1,2,3}, ensures that the customer waiting times within this service area will not be significantly affected. The same rule applies when this service area keeps shrinking from {1,2,3} to {1,2} with the further growth of the total travel demand. The result implies that the total travel demand level can also affect the best TSAD in terms of social welfare maximization.

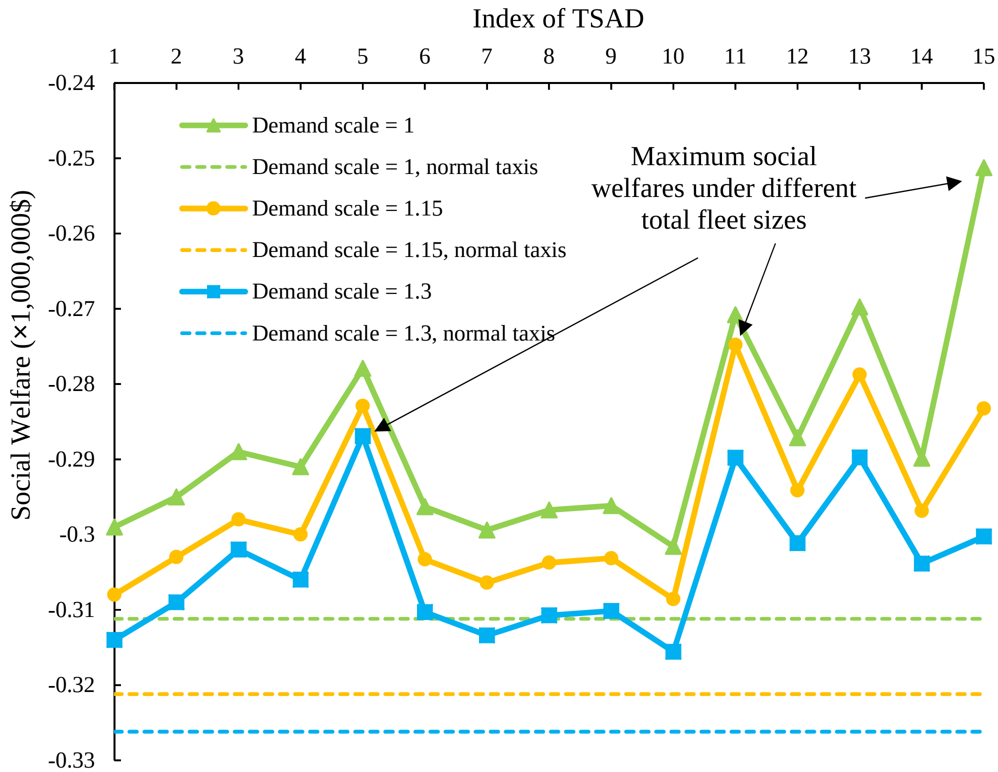

Fig. 20 Social welfares of different TSADs under different total travel demand levels

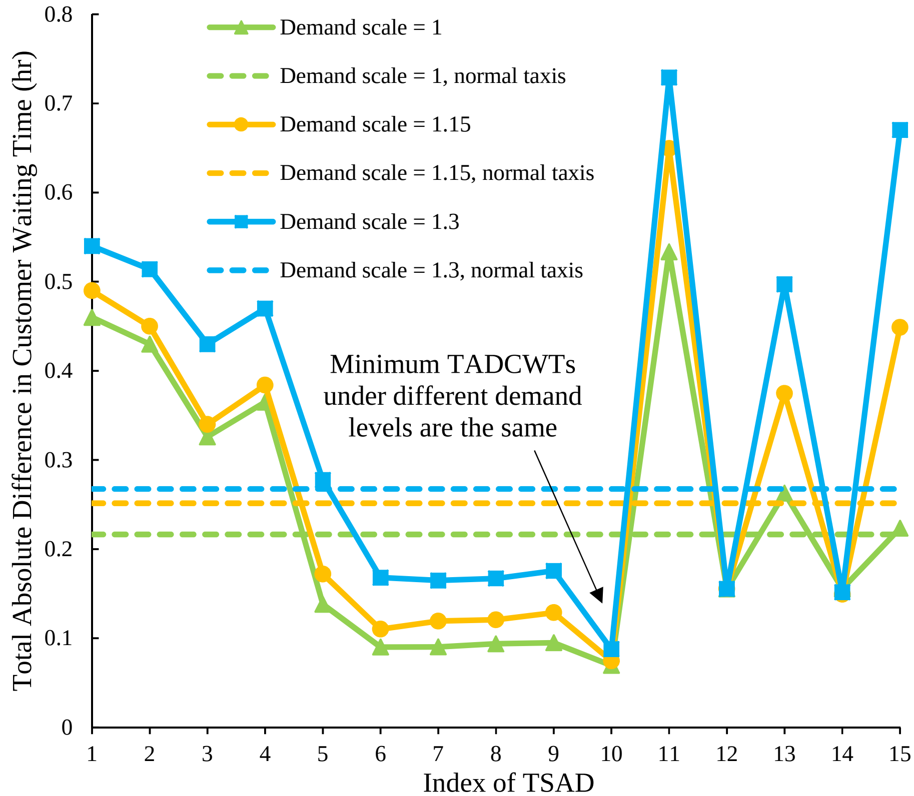

Fig. 21 The TADCWTs of different TSADs under different total travel demand levels

Fig. 21 depicts the TADCWT of each TSAD under different total travel demand levels and indicates that the lowest TADCWT under each demand scale is attained at the same TSAD, i.e., {3,4}. By increasing the total demand, the customer waiting time increases, but the distribution of vacant taxis remains the same as the demand distribution remains unchanged. Therefore, the total travel demand level does not affect the best TSAD in terms of equity.

4.2.3. Taxi fare level

In this section, we examine how taxi fare level affects the best TSAD. To see this, we simultaneously decreased the distance-based $(b_1^{o,2})$ and congestion-based $(b_2^{o,2})$ charges of the restricted area taxis to get two new fare levels other than those in Table 3, namely $(b_1^{o,2}, b_2^{o,2}) = (4, 50)$ and $(b_1^{o,2}, b_2^{o,2}) = (3, 40)$. All other parameters take the same values as those in Section 4.2.1, with the combination of the fleet sizes of the two taxi types the same as combination $III(2310, 890)$.

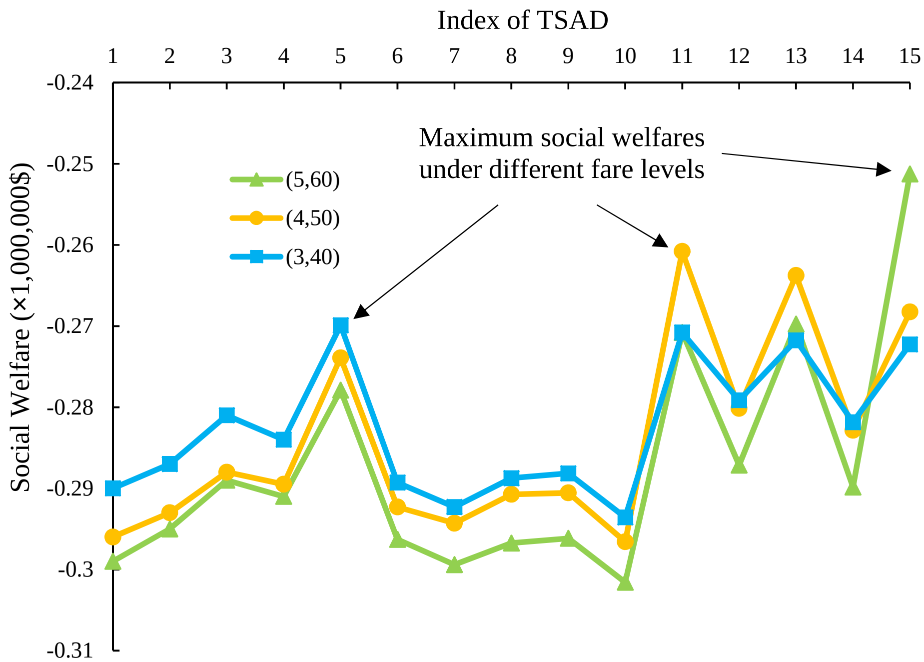

Fig. 22 Social welfares of different TSADs under different fare levels of restricted area taxis

Fig. 22 shows the social welfare of each TSAD under different fare levels of restricted area taxis. We observe that as the designed fare level decreases, the best TSAD changes from {1,2,3,4} to {1,2,3}, and {1,2} afterwards. This is because the demand for restricted area taxis increases as $b_1^{o,2}$ and $b_2^{o,2}$ decrease, which leads to the rise in the customer waiting time for restricted area taxis. Therefore, narrowing the service area of restricted area taxis is necessary to ensure that social welfare is maximized after the fare levels decrease.

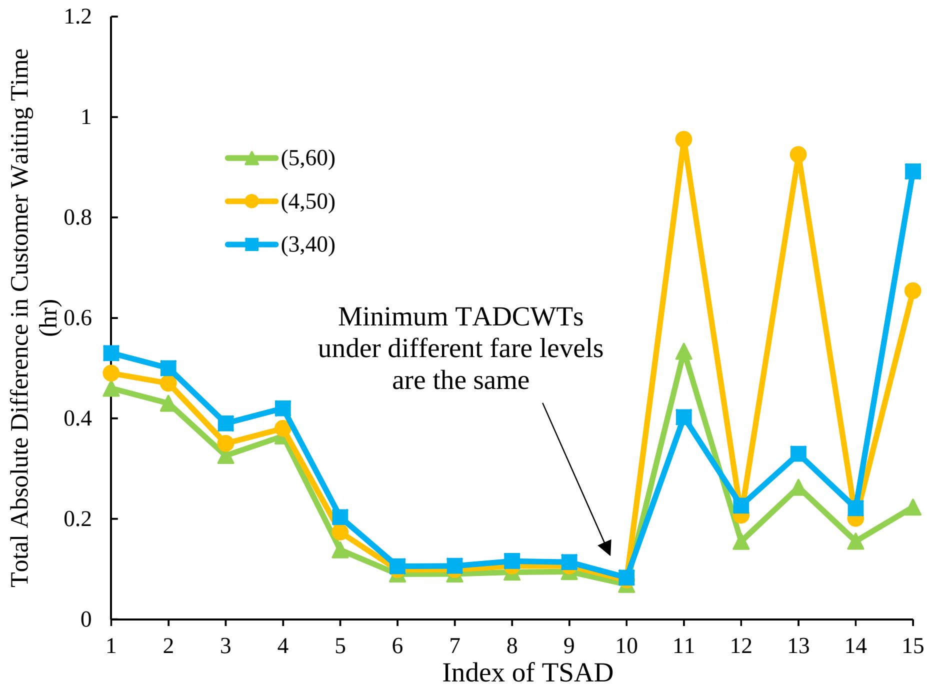

Fig. 23 The TADCWTs of different TSADs under different fare levels of restricted area taxis

Fig. 23 plots the TADCWT of each TSAD under different fare levels of restricted area taxis. We observe that the best TSAD in terms of TADCWT minimization or equity remains unchanged as the fare level decreases. The reason is that the reduction in fare only increases the total demand of restricted area taxis, but does not change the demand pattern for restricted taxis.

4.3. Case study

To demonstrate that the insights obtained from the small network are scalable to large networks, we also performed a case study of the Hong Kong network as shown in Section 4.1 by investigating the impact of fleet size combination on TSAD. We assumed a single restricted area taxi type (type A in Section 4.1) and fixed the total taxi fleet size as 18800 (veh). Like the 4-node network example, three fleet size combinations were designed as $I(16800, 2000)$, $II(15000, 3800)$, and $III(11000, 7800)$. All other unspecified parameters took the same as those in Section 4.1. The results obtained by the greedy heuristic are displayed in Table 11, from which we can tell that the service area of restricted area taxis obtained by the heuristic expands as the proportion of the fleet size of these taxis increases (e.g., the combination changes from combination *I* to *III*). This can be explained in the same manner as that in the small network example that the rise in consumer surplus is greater than the decrease in producer surplus by replacing normal taxis into restricted area taxis. Meanwhile, it is also interesting to observe that the output social welfare under combination *II*(15000,3800) (-9.02×10^8\$) is smaller than that in Table 6 (-8.31×10^8\$) in Section 4.1 in which there are three types of taxis. This implies that in this example, introducing type B taxis in addition to normal taxis and type A taxis can further contribute to the rise in social welfare.

Table 11 TSAD of Hong Kong network under different fleet size combinations

Fleet size combination	Solution	Social welfare ($\times10^8$\$)
I	{3,4,5,6,24,35,36,37,39,40,41,43,106,109,110,113,114}	-9.11
II	{3,4,5,6,7,9,12,22,23,24,25,26,27,32,33,35,36,37,38,39,41,106,109,110,111,112,113,114,115,116}	-9.02
III	{1,2,3,4,5,6,7,9,12,22,23,24,25,26,27,30,31,32,33,35,36,37,38,39,41,79,103,106,109,110,111,112,113,114,115,116,125}	-8.76

5. Conclusion

We have developed a mixed integer nonlinear optimization model to determine the taxi service area design with the objective to maximize social welfare. The model contains two sub-problems. The first one is a combined network equilibrium problem; the second one is a regulatory problem, which is to select a specific region from the entire network as the service area of each type of restricted area taxi. A greedy heuristic is proposed to solve the model. Numerical examples are given to examine the performance of the proposed heuristic and to provide insights into taxi service area design. The results show the following:

1. The proposed greedy heuristic can produce the same results as the enumeration method with a large decrease in computational time and can produce a superior result than the genetic algorithm within the same computational time in a Hong Kong network example. The efficiency of the developed greedy heuristic is more significant as the network size increases.

2. Under social welfare maximization, the best service area design may vary with the proportion of the fleet size of restricted area taxis to the total fleet size of taxis, the total fleet size of both normal and restricted area taxis, the total travel demand level and the demand distribution, and the fare level of restricted area taxis.
3. Under the minimization of the total absolute difference in customer waiting time between each pair of zones, the best service area design may vary with the proportion of fleet size of restricted area taxis to the total fleet size and demand distribution, but the best service area design may not vary with the total fleet size of both normal and restricted area taxis, the total travel demand level, and the fare level of restricted area taxis.
4. In terms of social welfare maximization, having a service area restriction for restricted area taxis is not a must.
5. Introducing restricted area taxis properly can reduce the customer waiting time inside the service area of restricted area taxis but increase the customer waiting time outside that area.
6. We may have contradictory conclusions on the best taxi service area design under different design objectives, e.g., social welfare maximization and the minimization of the total absolute difference in customer waiting time between each pair of zones in the network.
7. A tradeoff exists between social welfare maximization and customers' equity in terms of the total absolute difference in customer waiting time between each pair of zones.
8. Under social welfare maximization, the area of the best TSAD may shrink as the total taxi fleet size or the fare level of restricted area taxi decreases or as the total demand increases.

We believe that this paper opens up some new study directions. First, an extension can be made to the model by incorporating heterogeneous customers as proposed by Wong et al. (2008) to depict various demand elasticities among heterogeneous customers. Second, travel demand often varies over time of day. Introducing a time-dependent fleet size or service area scheme (e.g., introducing peak hour taxis in the central business districts or for specific commuting routes during peak hours) to our framework in responsive to the demand variation may help to improve customers' mobility and the service qualities of taxis. Third, traffic dynamics may be a key factor that affects service area design. One possible research direction is to extend the proposed framework to consider traffic dynamics. One approach can be based on network macroscopic fundamental diagrams (e.g., Yildirimoglu and Geroliminis, 2014;

Yildirimoglu et al., 2015; Ramezani and Nourinejad, 2018). Fourth, more behavioral considerations of taxi drivers can be considered in two possible ways. On one hand, taxi drivers may choose between different types of taxis to drive according to the possible income of each type. When a significant imbalance exists between incomes of the two types of taxi drivers, the supply of drivers to the two types may also be imbalanced. Consequently, it is necessary to consider the fairness among drivers in terms of the income by adding more constraints to the current model. On the other hand, the service time of drivers can also be an important issue. Our investigated system is purely centralized such that all taxis are mandated to serve the market all the time. This assumption is reasonable as we only consider one hour peak period. However, this assumption may not be applicable if the modeling period is extended, for example, to a whole day with a large demand variation over time of day. In this case, it is unreasonable and inefficient to have all drivers working all day long. Therefore, the assumption can be relaxed in future studies by allowing drivers having a choice of shift, such as morning, night, and peak hour shifts. Fifth, the equilibrium model proposed from Wong et al. (2008) was developed based on the concepts of “meta-zones” (in which every meta-zone consists of many links (i.e., streets between intersections) and many intersections) and “meta-links” (in which each meta-link contains many parallel and convoluted streets between meta-zones) and their model used the well-known BPR function to determine both intra- and inter-meta-zonal travel times. However, the BPR function has its limitations. For example, there is no evidence that the BPR function can accurately represent the relationship between traffic flow and travel time on meta-links. Moreover, the BPR function allows link flow to be greater than its capacity, which can also be unrealistic. Therefore, future studies can focus on the validation and calibration of BPR functions for meta-zones. Last but not least, one challenging research direction is to develop exact methods to get optimal solutions to our studied problem efficiently.

Acknowledgments

This study is supported by the Research Grants Council of the Hong Kong Special Administrative Region of China (17200618). The authors wish to thank the four anonymous reviewers for their useful comments and suggestions.

References

Arnott, R., 1996. Taxi travel should be subsidized. Journal of Urban Economics, 40 (3), 316-333.

Beesley, M.E., Glaister, S., 1983. Information for regulating: the case of taxis. The Economic Journal, 93 (371), 594-615.

Cairns, R.D., Liston-Heyes, C., 1996. Competition and regulation in the taxi industry. Journal of Public Economics, 59 (1), 1-15.

De Vany, A.S., 1975. Capacity utilization under alternative regulatory restraints: analysis of taxi markets. Journal of Political Economy, 83 (1), 83-94.

Douglas, G.W., 1972. Price regulation and optimal service standards: the taxicab industry. Journal of Transport Economics and Policy, 6 (2), 116-127.

Fernández L., J.E., de Cea Ch, J., Briones, J., 2006. A diagrammatic analysis of the market for cruising taxis. Transportation Research Part E-Logistics and Transportation Review, 42 (6), 498-526.

Flores-Guri, D., 2003. An economic analysis of regulated taxicab markets. Review of Industrial Organization, 23 (3-4), 255-266.

Foerster, J.F., Gilbert, G., 1979. Taxicab deregulation: economic consequences and regulatory choices. Transportation, 8 (4), 371-387.

Frankena, M.W., Pautler, P.A., 1986. Taxicab regulation: an economic analysis. Research in Law and Economics, 9, 129-165.

Hackner, J., Nyberg, S., 1995. Deregulating taxi services: a word of caution. Journal of Transport Economics and Policy, 29 (2), 195-207.

He, F., Shen, Z.J.M., 2015. Modeling taxi services with smartphone-based e-hailing applications. Transportation Research Part C-Emerging Technologies, 58, 93-106.

He, F., Wang, X., Lin, X., Tang, X., 2018. Pricing and penalty/compensation strategies of a taxi-hailing platform. Transportation Research Part C-Emerging Technologies, 86, 263-279.

HKSAR Government, 2002. Travel Characteristics Survey 2002 Final Report. Transport Department, HKSAR Government.

HKSAR Government, 2008. Report on Review of Taxi Operation. Transportation Advisory Committee, HKSAR Government.

Liang, X., Correia, G.H.D.A., van Arem B., 2016. Optimizing the service area and trip selection of an electric automated taxi system used for the last mile of train trips.

Transportation Research Part E-Logistics and Transportation Review, 93, 115-129.

Legislative Council, 1997. Land Supply in Hong Kong. Legislative Council Secretariat, Legislative Council, HKSAR.

Long, J.C., Szeto, W.Y., Du, J., Wong, R.C.P., 2017. A dynamic taxi traffic assignment model: a two-level continuum transportation system approach. Transportation Research Part B-Methodological, 100, 222-254.

Loo, B.P.Y., Leung, B.S.Y., Wong, S.C., Yang, H., 2007. Taxi license premiums in Hong Kong: Can their fluctuations be explained by taxi as a mode of public transport? International Journal of Sustainable Transportation, 1 (4), 249-266.

Manski, C.F., Wright, J.D., 1976. Nature of equilibrium in the market for taxi services. Transportation Research Record, 619, 296-306.

Moore, A.T., Balaker, T., 2006. Do economists reach a conclusion on taxi deregulation? Journal Watch, 3 (1), 109-132.

New York City Taxi & Limousine Commission, 2018. Background on the Boro Taxi Program. <http://www.nyc.gov/html/tlc/html/passenger/shl_passenger_background.shtml> (accessed on 22 October 2018).

Qian, X., Ukkusuri, S.V., 2015. Spatial variation of the urban taxi ridership using GPS data. Applied Geography, 59, 31-42.

Qian, X., Ukkusuri, S.V., 2017a. Taxi market equilibrium with third-party hailing service. Transportation Research Part B-Methodological, 100, 43-63.

Qian, X., Ukkusuri, S.V., 2017b. Time-of-day pricing in taxi markets. IEEE Transactions on Intelligent Transportation Systems, 18 (6), 1610-1622.

Qian, X., Zhang, W., Ukkusuri, S.V., Yang, C., 2017. Optimal assignment and incentive design in the taxi group ride problem. Transportation Research Part B, 103, 208-226.

Ramezani, M., Nourinejad, M., 2018. Dynamic modeling and control of taxi services in large-scale urban networks: a macroscopic approach. Transportation Research Part C-Emerging Technologies, 94, 203-219.

Schroeter, J.R., 1983. A model of taxi service under fare structure and fleet size regulation. Bell Journal of Economics, 14 (1), 81-96.

Shirabe, T., 2005. A model of contiguity for spatial unit allocation. Geographical Analysis, 37 (1), 2-16.

Shreiber, C., 1981. The economic reasons for price and entry regulation of taxicabs: a rejoinder. Journal of Transport Economics and Policy, 15 (1), 81-83.

Town Planning Board, 2018. Statutory Planning Portal 2. <https://www2.ozp.tpb.gov.hk/gos/#> (accessed on 22 October 2018)

Wang, X., He, F., Yang, H., Gao, H.O., 2016. Pricing strategies for a taxi-hailing platform. Transportation Research Part E - Logistics and Transportation Review, 93, 212-231.

Wong, K.I., Wong, S.C., Bell, M.G.H., Yang, H., 2005. Modeling the bilateral micro-searching behavior for urban taxi services using the absorbing Markov chain approach. Journal of Advanced Transportation, 39 (1), 81-104.

Wong, K.I., Wong, S.C., Wu, J.H., Yang, H., Lam, W.H.K., 2004. A combined distribution, hierarchical mode choice, and assignment network model with multiple user and mode classes. In: Lee, D.H. (Ed.), Urban and Regional Transportation Modeling. Edward Elgar Publishing Inc., Northampton, USA, 25-42.

Wong, K.I., Wong, S.C., Yang, H. 1999. Calibration and validation of a network equilibrium taxi model for Hong Kong. Proceedings of the 4th Conference of Hong Kong Society for Transportation Studies, The Hong Kong University of Science and Technology, 4 December, pp. 249-258.

Wong, K.I., Wong, S.C., Yang, H., 2001. Modeling urban taxi services in congested road networks with elastic demand. Transportation Research Part B-Methodological, 35 (9), 819-842.

Wong, K.I., Wong, S.C., Yang, H., Tong, C.O., 2003. The effect of perceived profitability on the level of taxi service in remote areas. Journal of Eastern Asia Society of Transportation Studies, 5, 79-94.

Wong, K.I., Wong, S.C., Yang, H., Wu, J.H., 2008. Modeling urban taxi services with multiple user classes and vehicle modes. Transportation Research Part B-Methodological, 42 (10), 985-1007.

Wong, R.C.P., Szeto, W.Y., Wong, S.C., Yang, H., 2014a. Modelling multi-period customer-searching behaviour of taxi drivers. Transportmetrica B: Transport Dynamics, 2 (1), 40-59.

Wong, R.C.P., Szeto, W.Y., Wong, S.C., 2014b. A cell-based logit-opportunity taxi customer-search model. Transportation Research Part C-Emerging Technologies, 48, 84-96.

Wong, R.C.P., Szeto, W.Y., Wong, S.C., 2015. A two-stage approach to modeling vacant taxi movements. Transportation Research Part C-Emerging Technologies, 59, 147-163.

Wong, S.C., Yang, H., 1998. Network model of urban taxi services: Improved algorithm. Transportation Research Record, 1623, 27-30.

Yang, H., Fung, C.S., Wong, K.I., Wong, S.C., 2010a. Nonlinear pricing of taxi services. Transportation Research Part A-Policy and Practice, 44 (5), 337-348.

Yang, H., Leung, C.W.V., Wong, S.C., Bell, M.G.H., 2010b. Equilibria of bilateral taxi-customer searching and meeting on networks. Transportation Research Part B-Methodological, 44 (5), 1067-1083.

Yang, H., Wong, S.C., 1998. A network model of urban taxi services. Transportation Research Part B-Methodological, 32 (4), 235-246.

Yang, H., Wong, S.C., Wong, K.I., 2002. Demand-supply equilibrium of taxi services in a network under competition and regulation. Transportation Research Part B-Methodological, 36 (9), 799-819.

Yang, H., Yang, T., 2011. Equilibrium properties of taxi markets with search frictions. Transportation Research Part B-Methodological, 45 (4), 696-713.

Yang, H., Ye, M., Tang, W.H., Wong, S.C., 2005a. Regulating taxi services in the presence of congestion externality. Transportation Research Part A-Policy and Practice, 39 (1), 17-40.

Yang, H., Ye, M., Tang, W.H.C., Wong, S.C., 2005b. A multiperiod dynamic model of taxi services with endogenous service intensity. Operations Research, 53 (3), 501-515.

Yang, T., Yang, H., Wong, S.C., 2014. Taxi services with search frictions and congestion externalities. Journal of Advanced Transportation, 48 (6), 575-587.

Yildirimoglu, M., Geroliminis, N., 2014. Approximating dynamic equilibrium conditions with macroscopic fundamental diagrams. Transportation Research Part B-Methodological, 70, 186-200.

Yildirimoglu, M., Ramezani, M., Geroliminis, N., 2015. Equilibrium analysis and route guidance in large-scale networks with MFD dynamics. Transportation Research Part C-Emerging Technologies, 59, 404-420.

Zhan, X., Qian, X., Ukkusuri, S.V., 2016. A graph-based approach to measuring the efficiency of an urban taxi service system. IEEE Transactions on Intelligent Transportation Systems, 17 (9), 2479-2489.

Zhang, W., Ukkusuri, S.V., 2016. Optimal fleet size and fare setting in emerging taxi markets with stochastic demand. Computer-Aided Civil and Infrastructure Engineering, 31, 647-660.

Zhu, J., Shuai, B., Huang, Z., Sun, C., 2013. The optimal taxi fleet size structure under various market regimes when charging taxis with link-based toll. Journal of Applied Mathematics, art. no. 535878.