The following publication Chen, J., Liu, Z., Wang, S., & Chen, X. (2018). Continuum approximation modeling of transit network design considering local route service and short-turn strategy. Transportation Research Part E: Logistics and Transportation Review, 119, 165-188 is available at https://doi.org/10.1016/j.tre.2018.10.001.

Continuum approximation modeling of transit network design considering local route service and short-turn strategy

ABSTRACT

This paper proposes a continuum approximation (CA) modeling framework to optimize the hybrid transit network designed with grids in the central district and hub-and-spoke structure in the periphery. Two CA models are formulated incorporating the local route service and the short-turn strategy respectively. The transit network configuration is optimized through minimizing the objective functions, which consider costs pertinent to passengers and operating agency. The decision variables include service boundary, spacings and headways of the regular service and the complementary services. Numerical experiments show that the performances of CA models with different complementary services are quite distinct under various demand scenarios.

Keywords: transit network design; continuum approximation; local route service; short-turn strategy

1. Introduction

In many urban cities, the rapid urbanization and accompanying increment in travel demand require a constant expansion of the transportation system. Public transport plays a pivotal role in the sustainable urban development, which is considered as one of the most efficient travel modes in the transportation system. Meanwhile, it has been widely recognized that infrastructure expansion itself is neither a sustainable nor effective solution for good serviceability. Systematic optimization is also needed to enhance the efficiency and reliability of the transit network in urban areas, which includes the strategic, tactical, and operational stages (e.g. Ceder, 2007; Yao, 2007; Li et al., 2011; Cipriani et al., 2012; Ibarra-Rojas et al., 2015; Zhang and Xu, 2017).

The transit network design (TND) is mostly related to the strategic stage and the tactical stage. The main difference between these two stages for TND is that at the strategic stage, TND determines the network layouts and associated service frequencies, while the tactical stage considers supplementary strategies such as short turn, deadheading and limited-stop service that can further improve the level of service of transit network. Currently, a number of previous studies have been conducted to model the TND problem at the strategic planning stage (Sheth et al., 2007; Lin and Chen, 2008; Saidi et al., 2014; Szeto and Jiang, 2014; Obeng et al., 2016; Chen et al., 2018; Huang et al., 2018). There are also numerous optimization models focusing on the design of strategies at the tactical planning stage (Ceder and Stern, 1981; Furth and Day, 1985; Wirasinghe and Seneviratne, 1986; Li et al., 2012; Liu et al., 2013; Bie et al., 2015; Chen et al., 2015; Verbas and Mahmassani, 2015; Yan et al, 2016). Yet, comparative studies pertinent to the optimization of TND at these two planning stages have received relatively less attention.

The TND problem can be modeled as a discrete optimization problem based on discrete data sets (e.g. candidate stops, demand centroids and expected travel time per link on the network). However, most of the proposed discrete models belong to the NP-hard class, which means an efficient algorithm for the exact solution of this problem may not exist. Hence, some studies either used exhaustive enumeration techniques to solve small instances of TND problems (Guan et al., 2003; Wan and Lo, 2003), or used heuristic and metaheuristic methods to tackle the real-size problems (Ceder, 2007; Szeto and Wu, 2011; Yu et al., 2012).

Considering the complexity of discrete models, a concise alternative for solving TND problem is the continuum approximation (CA) models that express the passenger demand as a density function over the whole network rather than an origin-destination (OD) matrix. At the expense of a set of simplifying assumptions, the number of decision variables in the TND problem can be largely reduced (e.g. only spacings between routes and stops, and vehicle headways are needed) and hence the CA models can be solved analytically, sometimes even in closed-form (Ouyang et al., 2014). Due to reduced requirements on both the quality and the quantity of inputs, CA models have been steadily gaining attention from the research community (Ouyang 2007; Daganzo, 2010; Estrada et al., 2011; Nourbakhsh and Ouyang, 2012; Ouyang et al., 2014; Petit et al., 2016; Saidi et al., 2016; Amirgholy et al., 2017).

1.1 Hybrid transit network

Typically, CA models are usually based on an idealized representation of the transit system which considers particular grid structures such as rectangular, hub-and-spoke, circular, etc. As one milestone work, Daganzo (2010) generalized previous studied and then proposed a hybrid transit network in a square city. The hybrid structure serves the entire city by means of a grid of transit routes that cross a central (e.g. downtown) district and branch throughout the peripheral service area as shown in Fig. 1.

Based on the hybrid transit network, Nourbakhsh and Ouyang (2012) utilized flexible routes instead of hub-and-spoke, to take passengers in low demand areas. Afterwards, Petit et al. (2016) compared the performance of hub-and-spoke and flexible-route structure based upon CA models. It is worth noting that the hybrid concept was also adapted to a circular city with ring-radial transit networks (e.g. Saidi et al., 2016; Chen and Nie, 2017, 2018).

Though the hybrid structure can yield suited heterogeneous network configurations, most of the current CA models treat transit demand density as homogeneous over the citywide space. Such uniform demand assumption is not practical and conceals the advantages of the hybrid structure for TND. Recently, several researchers began to consider a city’s spatial variations in travel demand. Ouyang et al. (2014) embedded the spatially-varying demand into the CA model, which assumes that demand density is a continuous function and changes smoothly over the space of a rectangular city. Saidi et al. (2016) focused on the planning of the ring-radial rail transit network, and proposed a more generalized model which is available to consider the realistic demand.

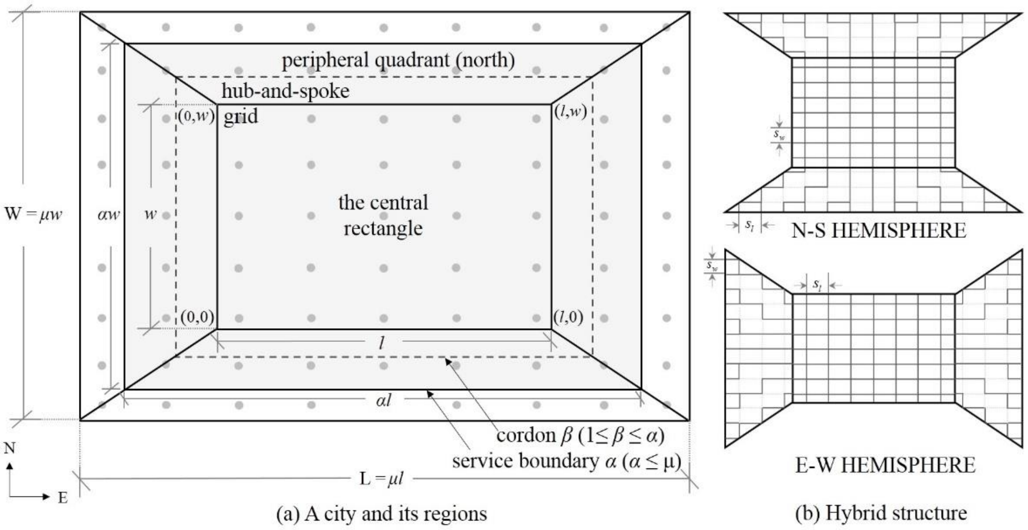


Fig. 1. Rectangular city and its hybrid transit network.

Source: the transit network structure and layout is adopted from Daganzo (2010).

1.2 Two complementary services

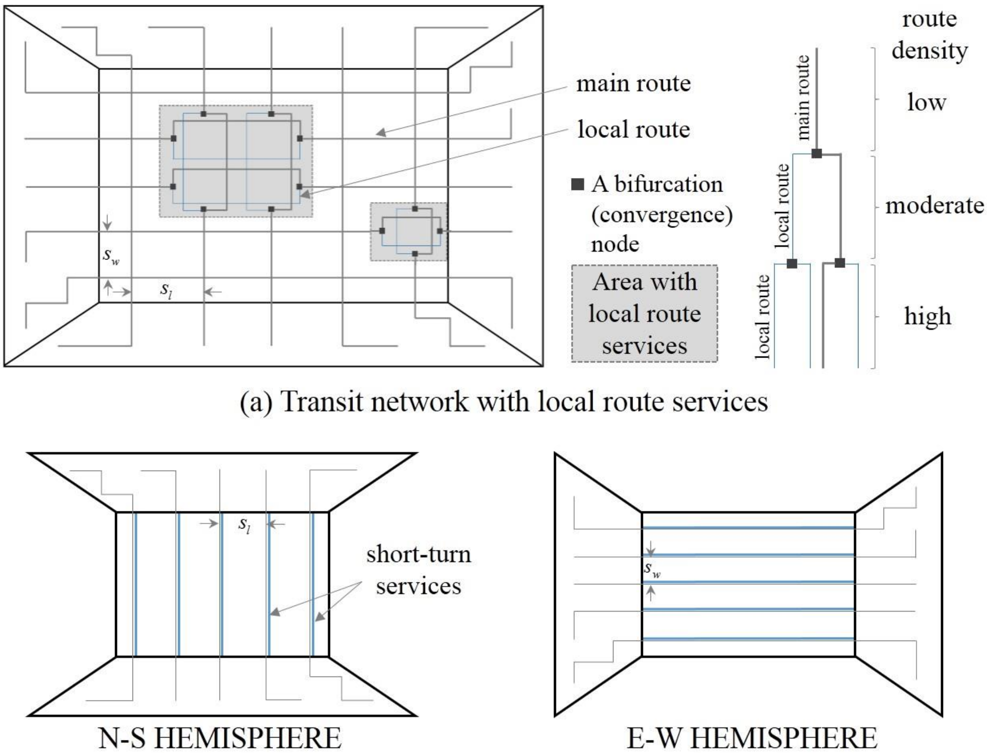
In practice, the efficiency of the transit network can be further enhanced by some complementary services at different planning stages. Ouyang et al. (2014) proposed a novel heterogeneous layout of routes distinct from Daganzo (2010)[1]. A grid of “main” transit routes which provide regular service are first formed to cater for the primary

[1] Readers are referred to Daganzo (2010) for details on the hybrid structure of transit network, and to Ouyang et al. (2014) for details on the local route service.

travel demand in the entire service area, and then one or more grids of closely-spaced "local" routes are generated in order to provide additional service for relatively high demand district (see Fig. 2a). This complementary service first proposed by Ouyang et al. (2014) is termed as the *local route service* at the strategic stage. Yet, the peripheral demand is always less than that in the city center. Compared with a grid of routes that cover the city, the hybrid structure in Fig. 2 can generate lower generalized cost in regard to the relatively low peripheral demand. Hence, we extend the pioneering work of Ouyang et al. (2014) by developing a new CA model to consider a hub-and-spoke configuration in the city's peripheral region.

Furthermore, all of the previous CA models only focus on the strategic planning stage. In practice, transit network can also benefit from several strategies (e.g. the short-turn strategy) at the tactical planning stage as well. However, none of the previous studies on CA models have investigated the transit network with these tactically complementary services.

This study aims to remedy the above gaps in the extant literature where *two complementary transit services* (local route service and short-turn strategy, see Fig. 2) are considered at the planning level. Based on the CA modeling framework, we build two models for these two complementary services.



(a) Transit network with local route services

(b) Transit lines with short-turn services

Fig. 2. General scheme of transit network with two complementary services.

Model 1 addresses the *local route service* at the strategic stage. As illustrated in Fig. 2a, it serves some city areas with higher travel demand through configuring denser routes and stops. Note that local routes are assumed to be only provided in the central district because demand in the periphery is relatively low and the value of local routes there is dwarfed.

Model 2 addresses the *short-turn strategy* at the tactical stage, which attracts much attention in the realm of transit system (e.g. Furth and Day, 1985; Furth, 1987; Cortés et al., 2011). Differed from the regular service spanning the entire service area, the short-turn service only serves a part of main routes in the central district to cover areas with higher demand, as shown in Fig. 2b.

1.3 Objectives and contributions

The main contributions of this study are twofold: (a) to build a new CA modeling framework and (b) to incorporate two complementary transit services into the hybrid structure of transit network (see Fig. 1). Based on the CA, the two services are respectively handled by two models. The focus of Model 1 lies in the determination of route layouts at the strategic stage. Short-turn strategy in Model 2 is one of the most widely adopted tactical strategies. These two models seek to optimize the transit network configuration plan, including service boundary, stop spacings, route spacings, and headways of the regular service and the complementary services.

These two complementary services are both beneficial to tackling a city's spatial and temporal variations in travel demand. In this study, a comparative analysis is further conducted to quantify the merits and benefits of two complementary services under the heterogeneous demand. We could gain insights on when and which transit service is appropriate to be considered in the transit network design under a given demand scenario.

The remainder of this paper is organized as follows. Section 2 below presents the two CA models and associated solution procedures. Numerical examples are presented in Section 3. Finally, conclusions are provided in the last section.

2. Methodology

This section presents the models and the solution methods. Variables/parameters used in the model and their notation are summarized in Appendix B.

2.1 Basic assumptions

2.1.1 City form and the supply of transit

The city form is an idealized rectangular shape with one central district and four peripheral quadrants (see Fig. 1). The central district has a grid structure with length l in the E-W travel directions (along the x axis), and width w in the N-S directions (y axis). The length and width of the entire city is set as L and W respectively, and we assume that $\mu = L/l = W/w$. Furthermore, we define a variable α to determine the service boundary to the most extent where the transit network can cover ($1 \leq \alpha \leq \mu$)[2]. Then, the transit service area is expressed as a rectangle with length αl and width αw. Four quadrants are trapezoids and differentiated as North (N), South (S), East (E) and West (W). The hourglass-shaped region (see the top right corner of Fig. 1) including the north peripheral quadrant, the central rectangle and the south peripheral quadrant is called as "N-S hemisphere". The "E-W hemisphere" is similarly defined.

[2] It is worth noticing that the variable α in this study is $\alpha \geq 1$, which differs from the model in Daganzo (2010) where $\alpha \leq 1$.

Assume that the entire city has a very dense rectangular mesh of two-way streets, which parallel with the city's boundaries. In the central district, the base route spacing between N-S routes is s_l and between E-W routes is s_w. Transit stops locate at the intersections of the routes, such that the stop spacing along an N-S route is the route spacing for the E-W routes that cross it, and vice versa.

In each peripheral quadrant, the stretch of transit routes within the service area has a hub-and-spoke structure with trunk and branch route segments (Daganzo, 2010), but these routes do not cross the boundary between peripheral quadrants (see Fig. 1).

2.1.2 The demand of transit

The entire city is decomposed into sufficiently small neighborhoods, and each neighborhood is denoted by a coordinate (x, y) in the transit network, which is the location of its centroid. The travel demand to and from each neighborhood is represented based on its centroid.

In the central district, the demand is treated as heterogeneous over space, which is defined by a predetermined demand density function and calculated by the integral calculus. Demand in the periphery is usually less than that in the central district, and the spatial variations of the peripheral demand is mild compared with the central demand. In this study, the demand is assumed to be uniformly distributed over each peripheral quadrant. Many variables and parameters in the periphery are hard to be directly calculated by the integral calculus. Instead, they can be expressed in the form of expectations based on the probability theory. For the ease of presentation, we utilize the term "cordon" which stems from Daganzo (2010) to calculate the expected values of some peripheral variables and parameters in Section 2.2. The perimeter of a rectangle with length βl and width βw (see the dotted line in Fig. 1), is termed as "cordon β" ($1 \leq \beta \leq \alpha$).

Similar to Daganzo (2010) and Ouyang et al. (2014), it is assumed that passengers: (i) use the closest stops to their origins and destinations with walking speed v_w; (ii) choose the direct route between stops, with the least possible number of transfers; (iii) randomly choose their travel direction whenever there is a tie among multiple alternatives; (iv) transfer at the first opportunity when both origins and destinations are in the peripheral region within the same hemisphere. Compared with other complex optimum rules for passenger behavior, these assumptions are simple and realistic, which lead to unique paths with near minimum travel times (Daganzo, 2010).

2.1.3 Regular service

Main routes provide the regular service in each hemisphere which run from one edge of the transit service area to the opposite edge. These main routes commute between two peripheral quadrants, vertically pass through the central district, and then travel back, which include N-S routes (N-S-N) and E-W routes (E-W-E).

Transit vehicles are assumed to have: an average cruising speed, denoted by v, taking into consideration the interruptions from traffic and pedestrian; a common dwell time, denoted by τ[3], at a stop due to the door operations and loading and unloading passengers; vehicle capacity, denoted by VC, measured by the maximum number of in-vehicle passengers. We further assume that all the vehicles stop at every stop along their routes and operate at the cruising speed with only moderate variability in the headway; see Ceder (2007) with respect to the headway control methods. The headway of the regular service is assumed to be H in the central district and become higher in the periphery after the routes branch.

2.1.4 Complementary service

Apart from the regular service, two complementary transit services are utilized to further cope with the serviceability.

(1) Local route service

The local route service is assumed to be only provided in the central district since the demand in the periphery is relatively low and the benefit of local routes there is dwarfed. In the central district, to achieve the practical layout of local routes, we require that the spacings between parallel routes comply with a power-of-two scale (Roundy, 1985; Ouyang et al., 2014).

The power-of-two scale is available to provide seamless alignment of main routes and local routes. This concept stems from the realm of inventory control to optimize the synchronization between production schedules at very small loss of optimality. Namely, when a new local route is needed in a dense area, a bifurcation node should be first added, and then the new local route serves this area in parallel with the existing route. For example, the right-hand-side of Fig. 2a shows the geometry of a multiple levels of local route structure. In this example, the main route serves the N-S hemisphere, and at each bifurcation node, one new local route is added. All the newly added local routes travel in parallel with the main route. Therefore, for the southerly direction, each bifurcation node is a branching point, while for the north direction, each bifurcation node is a converging point.

Following the power-of-two scale, the stop spacings for a neighborhood's E-W routes, $s_l(x, y)$, and its surrounding N-S routes, $s_w(x, y)$, should fulfill a discrete exponential scale as:

$$s_l(x,y) = s_l \big/ 2^{k(x,y)}, \quad s_w(x,y) = s_w \big/ 2^{k(x,y)} \tag{1}$$

For the simplicity of the model, vehicles serving the local routes are assumed to have the same headway H as the corresponding main routes. Eq. (1) contains three decision variables: the first two decision variables s_l and s_w are the base stop spacings for E-W and N-S routes, respectively; the third variable $k(x, y)$ determines the level of branching, which only takes nonnegative integer values; for instance, if $k(x, y)$ equals 0, it implies no new local route is needed at (x, y). It should be noted that the level of

[3] Though the dwelling time at a stop could be treated as a function of the count of boarding/alighting passengers, this is not done to avoid the proliferation of parameters.

branching $k(x, y)$ should only vary consecutively rather than take a sharp change (e.g., from 0 to 2 directly).

(2) Short-turn service

Unlike the regular service which travels across the entire service area, the short-turn service only commutes on a segment of the regular service in the central rectangle. The short-turn services are used to enhance the capacity and serviceability of transit routes in the central district areas with higher demand. Vehicles serving a short-turn service are assumed to operate with a headway H_s, which is generic for all the short-turn vehicles.

2.2 Continuum approximation modeling

As discussed in the Introduction, CA models can be conveniently used to predict the transit flows. Compared with other transit assignment models (e.g., strategy-based model; frequency-based model), it is more efficient for the CA model to be solved. The CA model framework of this study is elaborated as follows.

2.2.1 Transit demand

The challenges of developing a CA model mainly lay on calculating the demand rate from different areas, e.g., from the central district (denoted by C) to the peripheral service area (denoted by P) or another location in the central district. Since the transit service may not cover the entire city (i.e. $\alpha \le \mu$), we use $\hat{P}$ to denote the periphery which includes the region outside the service area.

There are four travel patterns: (i) central district - central district ($C-C$); (ii) periphery - central district ($P-C$); (iii) central district - periphery ($C-P$); (iv) periphery - periphery ($P-P$).

We first discuss about the travel pattern $C-C$. First, we use superscript "*start*" ("*end*") to denote the demand rate originating from (destining for) a neighborhood (x,y). $\delta(x,y,\bar{x},\bar{y})$ is used to denote the demand density function from the neighborhood (x,y) to another neighborhood $(\bar{x},\bar{y})$. Thus, $D_{C-C}^{start}(x,y)$ is the total amount of demand generated at (x, y) and going to other central areas; $D_{C-C}^{start}(x,y) = \iint_C \delta(x,y,\bar{x},\bar{y}) d\bar{\sigma}_C$. $D_{C-C}^{end}(x,y)$ is the total amount of demand attracted at (x,y) from other central areas; $D_{C-C}^{end}(x,y) = \iint_C \delta(\bar{x},\bar{y},x,y) d\bar{\sigma}_C$. At an aggregate level, $\iint_C D_{C-C}^{start}(x,y) d\sigma_C = \iint_C D_{C-C}^{end}(x,y) d\sigma_C = D_{C-C}$. Here, the integral domain C in $\bar{\sigma}_C$ and σ_C represent the entire central district.

For the travel pattern *P*-*C*, the total generated demand at (*x*, *y*) in the periphery (per unit area per time) to the whole central district $D_{P-C}^{start}(x,y)$ is equal to $\iint_C \delta(x,y,\bar{x},\bar{y})d\bar{\sigma}_C$, and the total attracted demand at (*x*, *y*) in the central district from the peripheral service area $D_{P-C}^{end}(x,y)$ is equal to $\iint_P \delta(\bar{x},\bar{y},x,y)d\bar{\sigma}_P$, where the integral domain $\bar{\sigma}_C$ and $\bar{\sigma}_P$ denote the entire central district and the whole peripheral service area, respectively. The total generated peripheral demand across the whole central district equals that of the total attracted central demand from the peripheral service area, namely $\iint_P D_{P-C}^{start}(x,y)d\sigma_P = \iint_C D_{P-C}^{end}(x,y)d\sigma_C = D_{P-C}$.

Different from the travel pattern *C*-*C*, the origin of a *P*-*C* trip is in the peripheral region. Since the peripheral demand is assumed to be uniformly distributed, the demand density of each peripheral quadrant is equal to that of the corresponding edge of the central district. Therefore, the demand density of four edges of the central district of *C*-*C* can be used to derive the demand density function $\delta(\cdot)$ of *P*-*C*, and give rise to the total aggregate demand $D_{P\text{-}C}$.

As to the travel patterns *C*-*P* and *P*-*P*, the total aggregate demand $D_{C\text{-}P}$ and $D_{P\text{-}P}$ can be obtained in an analogous way. Hereafter, the specific calculations of these demand rates are further discussed.

As the demand density function $\delta(\cdot)$ varies smoothly over all the neighborhoods within the central district, we can take the following two definite integral functions to calculate $D_{C-C}^{start}(x,y)$ and $D_{C-C}^{end}(x,y)$:

$$D_{C-C}^{start}(x,y) = \int_{\bar{y}=0}^{w}\int_{\bar{x}=0}^{l} \delta(x,y,\bar{x},\bar{y})d\bar{x}d\bar{y} \tag{2}$$

$$D_{C-C}^{end}(x,y) = \int_{\bar{y}=0}^{w}\int_{\bar{x}=0}^{l} \delta(\bar{x},\bar{y},x,y)d\bar{x}d\bar{y} \tag{3}$$

where $(\bar{x},\bar{y})$ stands for any neighborhood in the central district.

Then, the total aggregate demand of pattern (i), D_{C-C}, is the integral of $D_{C-C}^{end}(x,y)$ over the entire space of the central district, which equals

$$D_{C-C} = \int_{y=0}^{w}\int_{x=0}^{l} D_{C-C}^{end}(x,y)dxdy \tag{4}$$

With a little abuse of the notation, we use *N*, *S*, *E* and *W* to denote the four peripheral quadrants as well as north, south, east and west directions. Thus,

$D^{end}_{P(N)-C}(x,y)$ denotes the demand rate (per unit area per time) of the trips originating from the service area of the north peripheral quadrant and terminating at the neighborhood (*x*, *y*) within the central district. When passengers start from the peripheral service area and end at (*x*, *y*) within the central district, the demand rate $D^{end}_{P-C}(x,y)$ is the aggregated result of four peripheral quadrants as:

$$D^{end}_{P-C}(x,y) = D^{end}_{P(N)-C}(x,y) + D^{end}_{P(S)-C}(x,y) + D^{end}_{P(E)-C}(x,y) + D^{end}_{P(W)-C}(x,y) \quad \text{(5a)}$$

Proposition 1 then gives an approximation of $D^{end}_{P-C}(x,y)$ in the CA model.

Proposition 1. $D^{end}_{P-C}(x,y)$ can be approximated as:

$$D^{end}_{P-C}(x,y) \approx \frac{1}{4}(\alpha^2 - 1)\begin{bmatrix} w\int_{\bar{x}=0}^{l} \delta(\bar{x}, w, x, y) d\bar{x} + w\int_{\bar{x}=0}^{l} \delta(\bar{x}, 0, x, y) d\bar{x} \\ +l\int_{\bar{y}=0}^{w} \delta(l, \bar{y}, x, y) d\bar{y} + l\int_{\bar{y}=0}^{w} \delta(0, \bar{y}, x, y) d\bar{y} \end{bmatrix} \quad \text{(5b)}$$

The proof of Proposition 1 is given in Appendix A.1.

As previously stated in Section 2.1, several peripheral variables are calculated by the expectations based on the probability theory, which means that these peripheral variables share the same expected values for each transit trip. As per the uniform distribution of the peripheral demand, the aggregate result of these peripheral variables can be expressed as the associated expected value times the total aggregate peripheral demand. Hence, to be consistent with peripheral variables calculated by the expectations, we mainly focus on the total aggregate demand pertaining to the peripheral service area (i.e. D_{P-C}, D_{C-P}, and D_{P-P}), rather than their specific demand rates (e.g. $D^{start}_{P-C}(x,y)$, $D^{start}_{C-P}(x,y)$, $D^{end}_{P-P}(x,y)$, etc.).

The total aggregate demand of pattern (ii), D_{P-C} is the integral of $D^{end}_{P-C}(x,y)$ over the entire space of the central district, which is expressed as:

$$D_{P-C} = \int_{y=0}^{w}\int_{x=0}^{l} D^{end}_{P-C}(x,y)\,dx\,dy \quad \text{(6)}$$

As the peripheral demand is assumed to be uniformly distributed, the demand originating from the peripheral service region and destining for the entire central district $D_{\hat{P}-C}$ approximately equals:

$$D^{end}_{\hat{P}-C}(x,y) \approx \frac{1}{4}(\mu^2 - 1)\begin{bmatrix} w\int_{\bar{x}=0}^{l} \delta(\bar{x}, w, x, y) d\bar{x} + w\int_{\bar{x}=0}^{l} \delta(\bar{x}, 0, x, y) d\bar{x} \\ +l\int_{\bar{y}=0}^{w} \delta(l, \bar{y}, x, y) d\bar{y} + l\int_{\bar{y}=0}^{w} \delta(0, \bar{y}, x, y) d\bar{y} \end{bmatrix} \quad \text{(7)}$$

Then, the relationship between $D_{\hat{P}-C}$ and D_{P-C} is as:

$$D_{\hat{P}-C} = \frac{\mu^2 - 1}{\alpha^2 - 1} D_{P-C} \tag{8}$$

The term "demand ratio", denoted by κ, is used here to denote the ratio of the demand of transit trips ending at the periphery $D_{(C,\hat{P})-\hat{P}}$ and the demand of trips terminating at the central district $D_{(C,\hat{P})-C}$. Since $D_{(C,\hat{P})-\hat{P}}$ is expressed as the aggregate demand of destination trips within the whole peripheral region, and the peripheral demand is uniformly distributed, it can be obtained with given $D^{end}_{(C,\hat{P})-C}$ and κ. This assumption is applied to trips originating from both the central district and the periphery. The aggregate demand associated with the entire periphery, $D_{C-\hat{P}}$ and $D_{\hat{P}-\hat{P}}$ are calculated as follows:

$$D_{C-\hat{P}} = \kappa_1 \cdot \int_{x=0}^{l}\int_{y=0}^{w} D^{end}_{C-C}(x,y)dydx, \;\; D_{\hat{P}-\hat{P}} = \kappa_2 \cdot \int_{x=0}^{l}\int_{y=0}^{w} D^{end}_{\hat{P}-C}(x,y)dydx \tag{9}$$

where κ_1 and κ_2 are the demand ratio coefficient for the central origin trips and the peripheral origin trips, respectively. We further assume that $D_{C-\hat{P}}$ and $D_{\hat{P}-\hat{P}}$ are equally distributed among the four peripheral quadrants, and passengers from the central district or a periphery have equal probability to go to the four peripheral quadrants.

In the sequel, the total aggregate demand for two O-D patterns (iii) and (iv), D_{C-P} and D_{P-P} can be expressed as:

$$D_{C-P} = \frac{\alpha^2 - 1}{\mu^2 - 1} D_{C-\hat{P}}, \;\; D_{P-P} = \left(\frac{\alpha^2 - 1}{\mu^2 - 1}\right)^2 D_{\hat{P}-\hat{P}} \tag{10}$$

2.2.2 Onboard passengers

A neighborhood's optimal route and stop spacings are not only determined by its own demand (generated and attracted) but also affected by the other onboard passengers travelling through this neighborhood. This is because the number of onboard passengers inherently affects the in-vehicle travel time. Furthermore, the vehicle size/capacity is mainly decided by the en-route stop/segment with the maximum number of onboard passengers. Hence, it is important to estimate the number of onboard passengers on each vehicle.

As the demand mostly locates in the central district, here we mainly focus on the flow of onboard passengers passing through the central district from four travel directions. For example, consider the eastbound direction, the rate of onboard passengers (per unit area per time) in Model 1, denoted by $F_W(x,y)$, is computed as follows:

$$F_W(x,y) = f_{W(C)}(x,y) + f_{W(P)}(x,y) \tag{11a}$$

where $f_{W(C)}(x,y)$ denotes trips that originate from the central district; $f_{W(P)}(x,y)$ denotes trips originating from the periphery. Proposition 2 then gives the expression of these two terms in Eq. (11a).

Proposition 2. Two terms in $F_W(x,y)$ can be expressed as:

$$\begin{aligned} f_{W(C)}(x,y) &= f^1_{W(C)}(x,y) + f^2_{W(C)}(x,y) \\ &= \frac{1}{2}\left(1+\frac{\alpha^2-1}{\mu^2-1}\kappa_1\right)\left(\begin{aligned} &\int_{x_1=0}^{x}\left[\int_{x_2=x}^{l}\int_{y_2=0}^{w}\delta(x_1,y,x_2,y_2)dy_2dx_2\right]dx_1 \\ &+\int_{x_2=x}^{l}\left[\int_{x_1=0}^{x}\int_{y_1=0}^{w}\delta(x_1,y_1,x_2,y)dy_1dx_1\right]dx_2 \end{aligned}\right) \end{aligned} \tag{11b}$$

$$\begin{aligned} &f_{W(P)}(x,y) = f^1_{W(P)}(x,y) + f^2_{W(P)}(x,y) \\ &\approx \left(1+\frac{\alpha^2-1}{\mu^2-1}\kappa_2\right)\left(\begin{aligned} &\int_{x_2=x}^{l}\int_{y_2=0}^{w}\frac{1}{w}D^{end}_{P(W)-C}(x_2,y_2)dy_2dx_2 \\ &+\int_{x_2=x}^{l}\left[D^{end}_{P(W)-C}(x_2,y) + \frac{x}{l}\left(D^{end}_{P(N)-C}(x_2,y) + D^{end}_{P(S)-C}(x_2,y)\right)\right]dx_2 \end{aligned}\right) \end{aligned} \tag{11c}$$

The proof of Proposition 2 is provided in Appendix A.2. The coefficient 1/2 in Eq. (11b) denotes the passengers' random choice of travel direction when there is a tie between two perpendicular routes. The approximation of the onboard rate components in Eq. (11c) is conservative because the peripheral quadrant is temporarily treated as a rectangle instead of a trapezoid for the sake of simplicity when calculating the peripheral onboard demand passing through (x, y). By multiplying $F_W(x, y)$ in Eq. (11a) with the route spacing $s_W(x, y)$ and headway, we can readily get the in-vehicle passengers of an eastbound directional vehicle in the neighborhood of (x, y).

In Model 2, there are some differences between the regular vehicles and the short-turn vehicles when passengers' origins or destinations are in the periphery. In order to select the route with the least number of transfers, a portion of these passengers may only wait the first arriving regular vehicle. Take the eastbound direction as an instance. In the neighborhood of (x, y) of the central district, the rate of onboard passengers (per unit area per time) boarding the regular vehicles and the short-turn vehicles, denoted by $F^1_W(x,y)$ and $F^2_W(x,y)$ respectively, are as:

$$\begin{aligned}F_W^1(x,y)=&\frac{1}{2}\frac{H_s}{H+H_s}(1+\frac{\alpha^2-1}{\mu^2-1}\kappa_1)\int_{x_1=0}^{x}\left[\int_{x_2=x}^{l}\int_{y_2=0}^{w}\delta(x_1,y,x_2,y_2)dy_2dx_2\right]dx_1+\\&\frac{1}{2}(\frac{H_s}{H+H_s}+\frac{\alpha^2-1}{\mu^2-1}\kappa_1)\int_{x_2=x}^{l}\left[\int_{x_1=0}^{x}\int_{y_1=0}^{w}\delta(x_1,y_1,x_2,y)dy_1dx_1\right]dx_2+\\&(1+\frac{\alpha^2-1}{\mu^2-1}\kappa_2)\int_{x_2=x}^{l}\int_{y_2=0}^{w}\frac{1}{w}D_{P(W)-C}^{end}(x_2,y_2)dy_2dx_2+\\&(\frac{H_s}{H+H_s}+\frac{\alpha^2-1}{\mu^2-1}\kappa_2)\int_{x_2=x}^{l}\left[D_{P(W)-C}^{end}(x_2,y)+\frac{x}{l}\left(D_{P(N)-C}^{end}(x_2,y)+D_{P(S)-C}^{end}(x_2,y)\right)\right]dx_2\end{aligned}\tag{11d}$$

$$\begin{aligned}F_W^2(x,y)=&\frac{1}{2}\frac{H}{H+H_s}(1+\frac{\alpha^2-1}{\mu^2-1}\kappa_1)\int_{x_1=0}^{x}\left[\int_{x_2=x}^{l}\int_{y_2=0}^{w}\delta(x_1,y,x_2,y_2)dy_2dx_2\right]dx_1+\\&\frac{1}{2}\frac{H}{H+H_s}\int_{x_2=x}^{l}\left[\int_{x_1=0}^{x}\int_{y_1=0}^{w}\delta(x_1,y_1,x_2,y)dy_1dx_1\right]dx_2+\\&\frac{H}{H+H_s}\int_{x_2=x}^{l}\left[D_{P(W)-C}^{end}(x_2,y)+\frac{x}{l}\left(D_{P(N)-C}^{end}(x_2,y)+D_{P(S)-C}^{end}(x_2,y)\right)\right]dx_2\end{aligned}\tag{11e}$$

The sum of $F_W^1(x,y)$ and $F_W^2(x,y)$ is identical to $F_W(x,y)$ in Eq. (11a). Passengers of four travel patterns have diverse behaviors of route choice between the regular vehicles and the short-turn vehicles. Eq. (11d) and Eq. (11e) are the result of transit assignment between two services. By multiplying $F_W^1(x,y)$ ($F_W^2(x,y)$) with the route spacing $s_W(x, y)$ and headway H (H_s), we can obtain the number of in-vehicle passengers of an eastbound regular (short-turn) vehicle in the neighborhood of (x, y). The flow of onboard passengers in three other directions can be similarly achieved based on the symmetric feature.

2.2.3 Passenger transfers

Passenger transfers are considered in the optimization objectives, which are expressed as the sum of an extra waiting time during transferring and a transfer penalty representing the inconvenience. The number of transfers of each trip is intimately related with the location of its origin and destination.

In the CA model, two transfer patterns are needed: directional transfer (denoted by D^{d_trans}) and spacing transfer (denoted by D^{s_trans}). The directional transfer occurs when passengers change their travel direction at the intersection of two perpendicular routes within the central district. For one trip, if either its origin or destination is in the central district, then one directional transfer is assumed to be needed. This directional transfer assumption is conservative because it ignores the cases that a trip's origin and destination are on the same bus route.[4] Furthermore, for a small part of passengers, if

[4] Ouyang et al. (2014) present that the expected number of directional transfers per trip whose origin and destination are in a rectangular city is in the range of 0.8-0.9. The approximation error from this conservative assumption will be trivial as the transit network size increases.

their origin and destination are in two opposite peripheral quadrants, then at most two directional transfers are needed.

Proposition 3. The number of directional transfers D^{d_trans} can be expressed as:

$$D_C^{d\_trans}(x,y)=\frac{1}{2}(1+\frac{\alpha^2-1}{\mu^2-1}\kappa_1)\int_{\bar{x}=0}^{l}\int_{\bar{y}=0}^{w}\left[\delta(x,\bar{y},\bar{x},y)+\delta(\bar{x},y,x,\bar{y})\right]d\bar{y}d\bar{x} \quad (12a)$$

$$D_P^{d\_trans}=D_{P-C}+\frac{3}{2}D_{P-P} \quad (12b)$$

where $D_C^{d\_trans}(x,y)$ denotes the number of directional transfers per unit area per time that the origin (x, y) is located in the central district; $D_P^{d\_trans}$ denotes and the total aggregate number of directional transfers that the origin of trips is in the periphery. The proof of Proposition 3 is given in Appendix A.3.

For transit network with local route services, passengers in the central district may have spacing transfer D^{s_trans} at a bifurcation or convergence node as shown in the right-hand-side of Fig. 2b. In a worst case, three transfers are needed: some passengers take a local route from their origins and then transfer to a main route, and meanwhile after performing a directional transfer, they have to make the second spacing transfer into a local route to their destinations. The number of spacing transfer is associated with the variation of $k(x, y)$ along the x and y directions. The variation of $k(x, y)$ is a discrete change which is hard to compute unless giving the entire shape of $k(x, y)$ in the central region. Ouyang et al. (2014) proposed a simplified approach that $k(x, y)$ is treated as a continuous variable, and described the variations via its “local” partial derivatives. Thus, the number of spacing transfers per unit area per time $D_C^{s\_trans}(x,y)$ in the CA model is the sum of spacing transfers from four travel directions:

$$D_C^{s\_trans}(x,y)=d_W^{s\_trans}(x,y)+d_E^{s\_trans}(x,y)+d_S^{s\_trans}(x,y)+d_N^{s\_trans}(x,y)$$

(13a)

Proposition 4. In the CA model the four terms in $D_C^{s\_trans}(x,y)$ can be approximated as:

$$d_W^{s\_trans}(x,y)\approx\frac{1}{2}\left(f_{W(C)}^2(x,y)+f_{W(P)}^2(x,y)+f_{E(C)}^1(x,y)+f_{E(P)}^1(x,y)\right)\left[\frac{\partial}{\partial x}k(x,y)\right]^+ \quad (13b)$$

$$d_E^{s\_trans}(x,y)\approx\frac{1}{2}\left(f_{E(C)}^2(x,y)+f_{E(P)}^2(x,y)+f_{W(C)}^1(x,y)+f_{W(P)}^1(x,y)\right)\left[\frac{\partial}{\partial x}k(x,y)\right]^- \quad (13c)$$

$$d_S^{s\_trans}(x,y)\approx\frac{1}{2}\left(f_{S(C)}^2(x,y)+f_{S(P)}^2(x,y)+f_{N(C)}^1(x,y)+f_{N(P)}^1(x,y)\right)\left[\frac{\partial}{\partial y}k(x,y)\right]^+ \quad (13d)$$

$$d_N^{s\_trans}(x,y) \approx \frac{1}{2}\left(f_{N(C)}^2(x,y)+f_{N(P)}^2(x,y)+f_{S(C)}^1(x,y)+f_{S(P)}^1(x,y)\right)\left[\frac{\partial}{\partial y}k(x,y)\right]^- \quad (13e)$$

Here, the partial derivatives in bracket means that $[\cdot]^+ = \max\{0,\cdot\}$ and $[\cdot]^- = -\min\{0,\cdot\}$. The proof of Proposition 4 is provided in Appendix A.4. The approximation in Eqs. (13b)-(13e) is conservative, because it reflects all the local changes of the continuous variable $k(x, y)$, but some of them may not have a value change after rounding.

Note that the spacing transfer formulae are also applicable to the transit network with short-turn service in Model 1, because $k(x, y)$ in the short-turn case is equal to a constant (i.e. 0), and the value of $D_C^{s\_trans}(x,y)$ equals zero.

2.3 An optimization model for optimal transit network design

With the CA model framework proposed in Section 2.2, we proceed to propose two CA models for the two complementary services, where Model 1 addresses the local route service and Model 2 accounts for the short-turn strategy. Costs associated with passengers and operating agency are first presented.

2.3.1 User travel costs

Four important components associated with the passengers' travel costs are considered: the walking access time, *AT*; the waiting time at stops, including those at transfers, *WT*; the in-vehicle travel time, *IVT*; and the extra penalties pertinent to transfers, *Pen*.

(1) Walking access time

As we assume that the city has a dense rectangular mesh of two-way streets, the boarding and alighting demand at each stop is yielded from a $s_l \times s_w$ rectangle at the peripheral stop or a $s_l(x,y)\times s_w(x,y)$ rectangle centered at the stop in the central district. Hence, the expected walking access distance in the peripheral service area is $\frac{1}{4}(s_l+s_w)$, and at (x, y) within the central district is $\frac{1}{4}(s_l(x,y)+s_w(x,y))$. These are applicable to both Model 1 and Model 2. The access time of passengers per unit area $AT_C(x,y)$ in the neighborhood (x, y) of the central district and the total access time in the periphery AT_P are:

$$AT_C(x,y)=\frac{1}{4v_w}\left[s_l(x,y)+s_w(x,y)\right]\left[(1+\frac{\alpha^2-1}{\mu^2-1}\kappa_1)D_{C-C}^{start}(x,y)+D_{C-C}^{end}(x,y)+D_{P-C}^{end}(x,y)\right] \tag{14a}$$

$$AT_P=\frac{1}{4v_w}(s_l+s_w)\left[D_{C-P}+D_{P-C}+2D_{P-P}\right] \tag{14b}$$

(2) Waiting time

The service headways of all routes are assumed to be constant and are not synchronized among perpendicular routes. Hence, the expected waiting time of stops at (x, y) is H/2 for transit network with local route service (i.e. Model 1), as in Daganzo (2010) and Ouyang et al. (2014). In Model 2, passengers of four travel patterns at a stop near (x, y) have various route choices. Most passengers board the first arriving vehicle, such that we can use the combined expected waiting time for transit network with short-turn service, i.e. $1/[2(1/H+1/H_s)]$, as in Spiess and Florian (1989) and Verbas and Mahmassani (2015). Meantime, a portion of passengers only wait the regular vehicle and their expected waiting time is H/2. The waiting time of passengers per unit area $WT_C(x, y)$ in the neighborhood (x, y) of the central district and the total waiting time in the periphery WT_P are:

$$WT_C(x,y)=\frac{1}{2}H\left[(1+\frac{\alpha^2-1}{\mu^2-1}\kappa_1)D_{C-C}^{start}(x,y)+D_C^{d\_trans}(x,y)+D_C^{s\_trans}(x,y)\right] \tag{15a}$$

$$\begin{aligned}WT_C(x,y)=&\frac{1}{2}\frac{HH_s}{H+H_s}(1+\frac{\alpha^2-1}{\mu^2-1}\kappa_1)D_{C-C}^{start}(x,y)+\\&\frac{1}{4}(\frac{HH_s}{H+H_s}+\frac{\alpha^2-1}{\mu^2-1}\kappa_1 H)\int_{\bar{x}=0}^{l}\int_{\bar{y}=0}^{w}\left[\delta(x,\bar{y},\bar{x},y)+\delta(\bar{x},y,x,\bar{y})\right]d\bar{y}d\bar{x}\end{aligned} \tag{15b}$$

$$WT_P=\frac{1}{3}\frac{\alpha^2+\alpha+1}{\alpha+1}H\left[D_{P-C}+D_{P-P}\right]+\frac{1}{2}HD_P^{d\_trans} \tag{15c}$$

$$WT_P=\frac{1}{3}\frac{\alpha^2+\alpha+1}{\alpha+1}H\left[D_{P-C}+D_{P-P}\right]+\frac{1}{2}\frac{HH_s}{H+H_s}D_{P-C}+\frac{1}{4}(2H+\frac{HH_s}{H+H_s})D_{P-P} \tag{15d}$$

where Eqs. (15a) and (15b) underscore the waiting time in the neighborhood of (x, y) per unit area for Model 1 and Model 2, respectively. The total waiting time in the peripheral region for Model 1 and Model 2 is shown in Eq. (15c) and Eq. (15d) respectively, which is derived in Appendix A.5.

(3) In-vehicle travel time

In the central district, the in-vehicle travel time of onboard passengers $IVT_C(x, y)$ in the neighborhood of (x, y) per unit area can be separated into two components: (i) vehicles in four travel directions move a unit longitudinal distance at the average cruising speed, v; (ii) vehicles dwell at stops when they load and unload passengers.

For transit network with local route service in Model 1, it has an extra in-vehicle travel time that local vehicles make lateral movements near the transferring nodes, and the lateral distance is approximately half of the local route spacing in the traveling direction.

$$
\begin{aligned}
IVT_C(x,y) &= \frac{1}{v}\left(F_E(x,y)+F_W(x,y)+F_S(x,y)+F_N(x,y)\right) \\
&+\tau\left(\frac{F_E(x,y)+F_W(x,y)}{s_l(x,y)}+\frac{F_S(x,y)+F_N(x,y)}{s_w(x,y)}\right) \\
&+\frac{1}{4v}\left(\left(F_E(x,y)+F_W(x,y)\right)s_w(x,y)\left|\frac{\partial}{\partial x}k(x,y)\right|+\left(F_S(x,y)+F_N(x,y)\right)s_l(x,y)\left|\frac{\partial}{\partial y}k(x,y)\right|\right)
\end{aligned} \tag{16a}
$$

As $k(x, y)$ is a constant value for the transit network with short-turn service, the third term vanishes and Eq. (16a) is also applicable to Model 2.

In the periphery, the total in-vehicle travel time IVT_P for two models is expressed in an approximately aggregated pattern, which is derived in Appendix A.6.

$$IVT_P = \frac{1}{8}\frac{(2\alpha+1)(\alpha-1)}{\alpha+1}\left(l(1/v+\tau/s_l)+w(1/v+\tau/s_w)\right)\left(D_{C-P}+D_{P-C}+2D_{P-P}\right) \tag{16b}$$

(4) Transfer penalty

The extra penalties pertinent to transfers $Pen_C(x, y)$ in the neighborhood (x, y) of the central district and Pen_P in the periphery are shown as follows:

$$Pen_C(x,y) = \theta\left(D_C^{d\_trans}(x,y)+D_C^{s\_trans}(x,y)\right) \tag{17a}$$

$$Pen_P = \theta D_P^{d\_trans} \tag{17b}$$

where θ is the extra penalty cost per unit time pertaining to the inconvenience of each transfer, and is in addition to the waiting time at a transfer stop. Eq. (17a) is also applicable to Model 2 because $D_C^{s\_trans}(x,y)$ here equals zero.

2.3.2 Operating costs

The operating cost is associated with the total vehicular time traveled, including the cruising time along the route and dwelling time at the stops. It is reasonable to assume that the operating cost is proportionally related to the total travel time of the transit vehicles. Hence, the operating costs can be expressed per time as follows:

$$O_C(x,y) = \frac{\phi_o 2^{k(x,y)}}{H s_l s_w}\left[\frac{2}{v}(s_l+s_w)+\frac{s_l s_w}{2^{k(x,y)}v}\left(\left|\partial k(x,y)/\partial x\right|+\left|\partial k(x,y)/\partial y\right|\right)+2^{k(x,y)+2}\tau\right] \tag{18a}$$

$$O_C(x,y) = \frac{\phi_o}{s_l s_w}\left(1/H+1/H_s\right)\left[2\left(s_l+s_w\right)/v+4\tau\right] \tag{18b}$$

$$O_P = \phi_o \frac{3(\alpha-1)lw}{s_l s_w H}\left[\left(s_l + s_w\right)/v + 2\tau\right] \quad (18c)$$

where ϕ_o is the unit value of transit operating cost. Eqs. (18a) and (18b) denote the operating costs in the neighborhood of (x, y) per unit area for Model 1 and Model 2, respectively. Eq. (18c) underscores the total operating cost in the periphery. The detailed derivations of these formulae are provided in Appendix A.7.

Remark 1. If the hub-and-spoke structure is replaced by grids in the peripheral service area, it will influence the behaviors of some passengers whose origins or destinations are in the periphery. In such case, some formulae pertinent to user costs need to be adjusted, such as the number of directional transfers $D_P^{d-trans}$ in Eq. (12b), the waiting time WT_P in Eq. (15c) and the in-vehicle travel time IVT_P in Eq. (16b). Meanwhile, the operating cost O_P in Eq. (18c) also needs to be modified. These characteristics of citywide grid structure can be easily incorporated in the proposed CA modeling framework.

2.3.3 Objectives

In this study, we introduce two different objective functions. Firstly, we consider a special case that the transit service covers the citywide space (i.e. $\alpha = \mu$). In the special case, the transit service area contains all the peripheries of the city. The objective function is the minimization of the generalized cost with respect to both passengers and transit operators per unit time for two CA models.

$$\begin{aligned} Z_1 &= Z_C + Z_P \\ &= \int_{x=0}^{l}\int_{y=0}^{w} Z_C(x,y)dydx + Z_P \\ &= \int_{x=0}^{l}\int_{y=0}^{w}\left[\phi_a AT_C(x,y) + \phi_w WT_C(x,y) + \phi_v IVT_C(x,y) + \phi_w Pen_C(x,y) + O_C(x,y)\right]dydx \\ &\quad + \left(\phi_a AT_P + \phi_w WT_P + \phi_v IVT_P + \phi_w Pen_P + O_P\right) \end{aligned} \quad (19)$$

where ϕ_a, ϕ_w, and ϕ_v denote the unit values of passenger access time, waiting time and in-vehicle travel time, respectively. Eq. (19) contains two terms denoting the overall cost in the central district Z_C and that in the periphery Z_P.

In the generic case, the city area sometimes is sufficiently broad, and transit routes may be too long to span the entire city. Hence, the citywide transit network is no longer an opportune option. We use the variable α ($1\leq \alpha \leq\mu$) to determine the optimal service boundary in ways that the network design is most cost-effective. Then, the objective function in Eq. (19) needs to be adjusted in order to adapt to the generic case.

The first objective considers all the passengers whose trips are located in the transit service area. The objective function is to minimize the average generalized cost for each passenger served by the transit network.

$$Z_2 = Z_1 \big/ (D_{C-C} + D_{P-C} + D_{C-P} + D_{P-P}) \tag{20}$$

The second objective takes account of passengers whose origins or destinations are in the periphery but outside the service area if $\alpha < \mu$. Such network design gives rise to inconvenience for these peripheral passengers who have to use other modes of transportation such as walking, biking, feeder buses, or park and ride to access the nearest main routes. Then, the objective function is to minimize the total social cost:

$$Z_3 = Z_1 + \rho(D_{\hat{P}-C} + D_{C-\hat{P}} + D_{\hat{P}-\hat{P}} - D_{P-C} - D_{C-P} - D_{P-P}) \tag{21}$$

where ρ denotes the penalty cost that represents a passenger whose origin or destination is outside the service area; it should be not less than the average travel cost of passengers who are in the periphery when $\alpha = \mu$. If the local authority wishes to increase transit patronage and further enhance transit attractiveness, the penalty coefficient ρ can be set a larger weight. Eq. (21) contains two terms: the first term is the generalized cost of transit operators and passengers within the service area, and the second term is the penalty cost due to unsatisfied passengers who are outside the service area.

2.3.4 Constraints

(1) Capacity constraint

Vehicle capacity constraint requires that the number of in-vehicle passengers in the critical stop (i.e., where vehicle occupancy is maximum) should be no more than the given capacity *VC*. As the demand is mostly concentrated in the central district, the critical stops always occur in this area. For instance, the in-vehicle passengers of an eastbound directional vehicle in the neighborhood of (x, y) within the central district:

$$F_W(x,y)s_w(x,y)H \leq VC \tag{22a}$$

$$F_W^1(x,y)s_w(x,y)H \leq VC,\ \ F_W^2(x,y)s_w(x,y)H_s \leq VC \tag{22b}$$

where Eqs. (22a) and (22b) underscore the vehicle occupancy constraints for Model 1 and Model 2, respectively. Vehicles traveling in other three directions have the similar vehicle occupancy constraints.

(2) Fleet size constraint

Vehicle operating could have a fleet size constraint, which ensures the available fleet needed to run the transit system.

$$\frac{1}{\phi_o}\int_{y=0}^{w}\int_{x=0}^{l} O_C(x,y)dxdy + \frac{3(\alpha+1)}{2(\alpha^2+\alpha+1)\phi_o} O_P \le FS \tag{23}$$

where *FS* denotes the maximum fleet size. Model 1 and Model 2 shares the same fleet size constraint. The fleet size of two models in the left-hand side of Eq. (23) is derived in Appendix A.8.

(3) Headway constraint

The policy headway constraints are considered. The upper bound policy headway H_P limits the average waiting time of passengers according to regulations that require a minimal service level.

$$H \le H_P \tag{24a}$$

$$H \le H_P, \quad H_s \le H_P \tag{24b}$$

where Eqs. (24a) and (24b) underscore the headway constraints for Model 1 and Model 2, respectively.

2.4 Solution procedures

The solution procedures of the two CA models are discussed in this section. We will consider Model 1 first and then Model 2.

2.4.1 Solution method for Model 1

The decision variables of Model 1 are s_l, s_w, H, $k(x,y)$ and α. For any given s_l, s_w, H and α, Z_P in Eq. (19) can be determined, while $Z_C(x,y)$ in the neighborhood of (*x*, *y*) within the central district can be generally optimized as a function of *k*(*x*, *y*). The determination of *k*(*x*, *y*) is on the basis of the previous work of Ouyang et al. (2014). *k*(*x*, *y*) is first treated as a continuous function. Then, $Z_C(x,y)$ can be expressed as the following pattern:

$$\begin{aligned} Z_C(x,y) = \Omega(x,y)\cdot 4^{k(x,y)} + \Gamma(x,y)\cdot 2^{k(x,y)} + \Pi(x,y)\cdot 2^{-k(x,y)} \\ + \Theta(x,y) + \Lambda\left(x,y,\{k(x,y)\}_{\forall(x,y)}\right) \end{aligned} \tag{25}$$

where the coefficient of each term is:

$$\Omega(x,y) = 4\phi_o\tau\cdot H^{-1}s_l^{-1}s_w^{-1}$$

$$\Gamma(x,y) = \frac{2\phi_o}{vH}\left(\frac{1}{s_l}+\frac{1}{s_w}\right) + \tau\left(\frac{F_E(x,y)+F_W(x,y)}{s_l} + \frac{F_S(x,y)+F_N(x,y)}{s_w}\right)$$

$$\Pi(x,y)=\frac{1}{4v_w}(s_l+s_w)\left[(1+\frac{\mu^2-1}{\alpha^2-1}\kappa_1)D_{C-C}^{start}(x,y)+D_{C-C}^{end}(x,y)+D_{P-C}^{end}(x,y)\right]$$

$$\Theta(x,y)=\frac{1}{2}H\left[(1+\frac{\mu^2-1}{\alpha^2-1}\kappa_1)D_{C-C}^{start}(x,y)+D_C^{d\_trans}(x,y)\right]+\theta D_C^{d\_trans}(x,y)$$
$$+\frac{1}{v}\left(F_E(x,y)+F_W(x,y)+F_S(x,y)+F_N(x,y)\right)$$

$$\Lambda(x,y,\{k(x,y)\}_{\forall(x,y)})=\frac{1}{4v}\left(\left(F_E(x,y)+F_W(x,y)\right)s_w(x,y)\left|\frac{\partial}{\partial x}k(x,y)\right|\right)$$
$$+\frac{1}{4v}\left(\left(F_S(x,y)+F_N(x,y)\right)s_l(x,y)\left|\frac{\partial}{\partial y}k(x,y)\right|\right)$$
$$+\frac{\phi_o}{Hv}\left(\left|\frac{\partial}{\partial x}k(x,y)\right|+\left|\frac{\partial}{\partial x}k(x,y)\right|\right)+\left(\frac{H}{2}+\theta\right)D_C^{s\_trans}(x,y)$$

Given s_l, s_w, H and α, the first four terms of Eq. (25) is a convex function of $2^{k(x,y)}$. Meanwhile, the exponential function $2^{k(x,y)}$ per se is monotonic. Note that the fifth term of Eq. (25) is non-negative. Therefore, we can procure a lower bound of the optimal value of Eq. (25) by ignoring its last term. Then, employ an interval reduction methods (such as the golden section method) to search the $k^*(x,y)$ for each cell that optimizes the first four terms of Eq. (25) as the lower bound value. An upper bound can be procured by calculating the fifth term based on the value of $k^*(x,y)$ and adding it to the first four terms. This upper bound value provides a feasible (near-optimum) solution to $Z_C(x,y)$, as the cost posed by spacing transfers occupies only a trivial part of all five terms of Eq. (25). Finally, we can round $\left\{k^*(x,y),\forall(x,y)\right\}$ to the nearest integer to form the final design in the central district. The decision variables s_l, s_w, H and α in Eqs. (20) and (21) can be solved numerically by a grid search. Combined with the local optimization of $\left\{k^*(x,y),\forall(x,y)\right\}$, the final transit network for Model 1 is determined.

2.4.2 Solution method for Model 2

For Model 2, the decision variables are s_l, s_w, H, H_s and α. The solution procedures are similar. The decision variables s_l, s_w and α are solved by a grid search. As regards the other decision variables H and H_s, we can easily verify that the objective is convex in H and H_s when two spacing variables and α are fixed. Here, we can use some descent methods such as Newton's method to obtain the optimal headways of two services.

3. Numerical Examples

The proposed CA models and solution methods described in previous sections are verified based on a square city network. In this section, we consider two cases: (i) a small city where the transit service is assumed to cover the citywide space ($\alpha = \mu$); (ii) a relatively large city where α becomes a decision variable ($1 \le \alpha \le \mu$) to determine the optimal service area.

In both cases, the entire city is decomposed into sufficiently small neighborhoods, each of size 0.1 km by 0.1 km. The length and width of the central district are set as 10 km. Travel demand in the central district is heterogeneous over space and expressed as the demand density. For instance, the demand from a neighborhood encompassing coordinates (x_1, y_1) to another neighborhood that encompasses (x_2, y_2) has a density $\delta(x_1, y_1, x_2, y_2)$ which follows this form:

$$\delta(x_1, y_1, x_2, y_2) = \prod_{i=1}^{2}\left(a_1 + a_2 \exp\left[-(a_3 x_i - a_{4i})^2 - (a_5 y_i - a_{6i})^2\right]\right) \tag{26}$$

Two numerical examples focus on an in-depth discussion of the mono-centric city, where both the origins and the destinations of travel demand are gathered in the city center. The other city topologies and associated spatial demand distributions (e.g. the twin city in Ouyang et al. (2014)) can also be obtained by altering the parameters of the demand density function in Eq. (26).

The parameter values used are summarized as follows: the average cruising speed of vehicles v is taken as 25 km/h and access walking speed v_w is 3.5 km/h. The dwelling time at a stop τ is set to be 30 s. The transfer penalty coefficient θ is taken as 1 min. Three unit values of travel time in Eq. (19) are taken as ϕ_a = 30 \$/h, ϕ_w = 30 \$/h, and ϕ_v = 20 \$/h. The unit value of transit operating cost is set as ϕ_o = 120 \$/h.

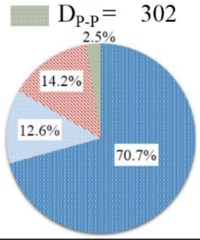
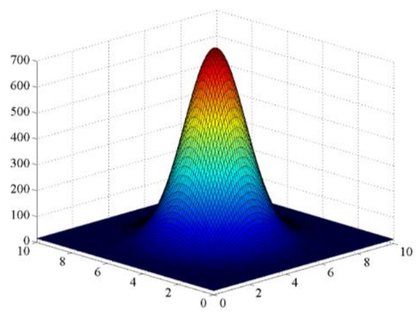
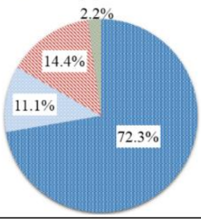
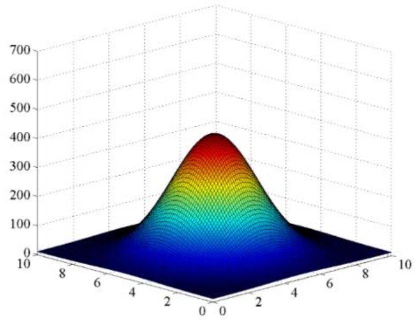
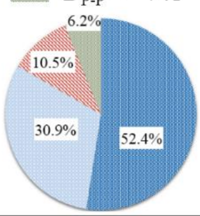
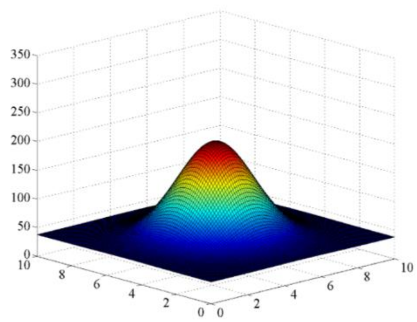
The solution methods are coded in *Matlab* R2013b and implemented on a personal computer with Inter Core i7-4700 CPU @ 2.40 GHz, 2.40 GHz and 8.00 G RAM. The convergence criterion for the golden section method in Model 1 and the Newton's method in Model 2 is based on the marginal contribution of successive iterations, with the predetermined tolerance of 10^{-2}.

We compare the performance of three heterogeneous network designs which are labeled as "A", "B" and "C" respectively. Network A is the original hybrid network firstly proposed by Daganzo (2010) with grids in the central district and hub-and-spoke structure in the periphery, but with no complementary services. The outcomes of Network A are viewed as benchmarks. Apart from the basic hybrid network, each of Network labeled "B" and "C" incorporates one complementary service, which corresponds to Model 1 and Model 2 respectively. Network B considers the local route service during the strategic planning stage, while the short-turn service from the perspective of tactical planning stage is embedded into Network C.

3.1. Case 1: α is fixed

In Case 1, the side length of city boundary is set as 14 km, i.e. $L = W = 14$. Since the peripheral region is not large, we assume that the transit service covers the citywide space ($\alpha = \mu = 1.4$). Z_1 in Eq. (19) is selected as the objective function. Three typical demand scenarios are explored. The demand density parameters in Table 1 were set so that the entire city's total base demand rate of transit trip-making is the same (12,000/hr) in each of the three scenarios.

Table 1 Input Parameters of Three Demand Scenarios in Case 1

Demand scenario	Demand density parameters	Aggregate demand of four O-D patterns unit: pass/h	Demand distribution in the central district unit: pass/ (h·km^2)
I	$a_1 = 0.0017$ $a_2 = 0.0803$ $a_3 = 0.58$ $a_{41} = a_{42} = 2.9$ $a_5 = 0.58$ $a_{61} = a_{62} = 2.9$	$D_{C\text{-}C} = 8{,}489$ $D_{P\text{-}C} = 1{,}511$ $D_{C\text{-}P} = 1{,}698$ $D_{P\text{-}P} = 302$  	
II	$a_1 = 0.0012$ $a_2 = 0.0437$ $a_3 = 0.41$ $a_{41} = a_{42} = 2.05$ $a_5 = 0.41$ $a_{61} = a_{62} = 2.05$	$D_{C\text{-}C} = 8{,}674$ $D_{P\text{-}C} = 1{,}326$ $D_{C\text{-}P} = 1{,}735$ $D_{P\text{-}P} = 265$  	
III	$a_1 = 0.0048$ $a_2 = 0.0209$ $a_3 = 0.46$ $a_{41} = a_{42} = 2.3$ $a_5 = 0.46$ $a_{61} = a_{62} = 2.3$	$D_{C\text{-}C} = 6{,}295$ $D_{P\text{-}C} = 3{,}705$ $D_{C\text{-}P} = 1{,}259$ $D_{P\text{-}P} = 741$  	

The combined boarding and alighting demand is obtained for each neighborhood by inserting Eq. (26) into Eqs. (2)-(10). The demand ratio coefficients in Eq. (9) are taken as $\kappa_1 = \kappa_2 = 0.2$. Hence, the demand rate of trips going to the central district and the periphery are 10,000 pax/h and 2,000 pax/h, respectively. The aggregate demand of four OD patterns and the resulting demand distributions in the central district for each demand scenario are also illustrated in Table 1. The values of constraint parameters are presented as: the vehicle capacity *VC* is taken as 90. The maximum fleet size *FS* is set as 420. The policy headway H_P is taken as 30 min.

3.1.1. Optimal network designs

This subsection assumes that vehicles can accommodate the whole city's travel demand. The proposed CA models are used to obtain the optimal transit network configuration plan (including route setting, spacing and headway) of Network A, B and C, respectively. Table 2 presents the key network design characteristics and the associated performance of three networks for each demand scenarios. The resulting local route service in the central district of Network B is depicted in Fig. 3.

Table 2 Transit Network Design Characteristics and Performance

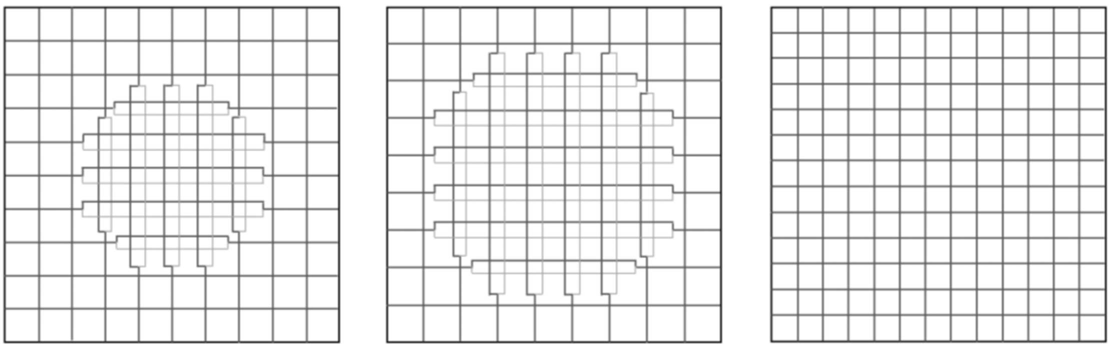
	Demand scenario (I)			Demand scenario (II)			Demand scenario (III)		
	Network A	Network B	Network C	Network A	Network B	Network C	Network A	Network B	Network C
s_l , s_w (km)	0.714	1.000	0.714	0.714	1.111	0.714	0.769	0.769	0.769
H (min)	7.4	6.4	10.9	7.4	6.5	11.3	6.9	6.9	8.2
H_s (min)	-	-	15.5	-	-	14.6	-	-	30
AT ($/h)	73,502	68,232	73,502	73,582	72,943	73,582	79,323	79,323	79,323
WT ($/h)	45,285	42,886	44,031	45,214	43,135	43,800	43,770	43,770	43,923
IVT ($/h)	68,828	71,476	68,828	73,258	76,064	73,258	112,819	112,819	112,819
Pen ($/h)	6,078	7,241	6,078	6,076	7,014	6,076	6,201	6,201	6,201
User ($/h)	193,693	189,835	192,439	198,130	199,156	196,716	242,113	242,113	242,266
Operating ($/h)	45,164	42,895	43,980	45,157	42,197	43,837	43,921	43,921	43,662
Total generalized cost ($/h)	238,857	232,730	233,017	243,294	241,353	240,553	286,245	286,245	285,928
Total generalized cost improvement (%)	-	2.56	1.02	-	0.79	1.13	-	0.00	0.01
Average travel cost ($/pax)	15.64	15.22	15.53	16.00	16.01	15.89	19.66	19.66	19.66
Average travel time (min/pax)	37.0	36.4	36.8	38.1	38.4	37.9	48.7	48.7	48.7

Note: Network A is the original hybrid network in Daganzo (2010) with no complementary service; Network B incorporates the local route service; Network C incorporates the short-turn service.

In the first demand scenario, the majority of transit trips are clustered in the city center of the central district; see Table 1. The highest demand density in the central district and the average demand in the periphery differ by a factor of around 50. In this case, Network B and Network C outperform Network A with over 1.0% reduction in the generalized cost. These cost savings arise in the benefits of two complementary transit services, which provide extra services for the high-demand central district. Though the spacing of main routes s_l and s_w in Network B is appreciably larger as compared against the other two designs (Network A and C), the local route service makes the city center of Network B have denser routes and stops as shown in Fig. 3a.

In the second demand scenario, the central district still aggregates most of the travel demand. Compared with the scenario I, the demand density in the central district changes more smoothly over space in the scenario II, and hence the central demand is distributed in a wider area. The highest demand density in the central district and the average demand in the periphery differ by a factor of about 35. In this case, the transit system still benefits from two complementary transit services. The short-turn service in Network C leads to a 1.13% decline in the generalized cost, while the local route service in Network B decreases the generalized cost by 0.79%. It is because the short-turn service covers the entire central district, and meantime the central demand of scenario II are distributed in a spacious area. Therefore, more passengers can be served by the short-turn service. As the service area of local routes increases (see Fig. 3b), the total cost raises quickly and the contribution of the local route service drops in this scenario.

Note that in the previous two demand scenarios, local routes in Network B are generically placed at areas with high demand density. Results show that the proportion of passengers who need three transfers is below 6%. At the same time, all the value of $k(x, y)$ in two demand scenarios never exceed one after rounding, which means the route spacings in the central district of Network B never vary by more than a factor of two. This conclusion is identical to Ouyang et al. (2014). They attribute this result to the economic order quantity (Harris, 1990) type of trade-offs between user and operating costs: the generalized cost is insensitive to values of the decision variables when the solution is near optimal (e.g. Daganzo, 2005). We find that the other possible reason is due to the inconvenience of transfers. The number of spacing transfers will increase rapidly when the value of $k(x, y)$ in some neighborhoods equals two, because many passengers have to experience two spacing transfers during their transit trips; see the right-side of Fig. 2b.



(a) Scenario I (b) Scenario II (c) Scenario III

Fig. 3. The local route service in the central district of Network B.

The third demand scenario is a bit different in that the central demand is not significantly larger than the peripheral demand. The highest demand density in the central district and the average demand in the periphery here differ only by a factor of about 5. An intriguing result is shown in Fig. 3c that Network B shares the same network configuration and headway with benchmark Network A. In this demand scenario, the advantage of two complementary transit services fades away. Transit is a shared passenger transport service, which is particularly apt to the city topologies with centripetal demand distributions (Daganzo, 2010). However, compared with the previous two demand scenarios, demand scenario III better fits the cities with a spatially homogeneous demand density throughout. Such a homogeneous demand penalizes transit and impairs the benefit of two complementary transit services.

3.1.2. Sensitivity analysis: effect of transit demand

Once the route configurations (route settings and spacings) are determined, the temporal variations in transit demand can be addressed by adjusting headways that vary with time of day. Our CA models are sufficiently flexible as we can take further constraints into account. For example, we may consider the limited values of vehicle capacity, a maximum fleet size and a maximum policy headway in Section 2.3.4.

With the obtained route configurations of Network A, B and C in Section 3.1.1, we conduct a sensitivity analysis on the effect of transit demand. A constant demand multiplier (*DL*) is utilized to reflect the variations of demand, i.e., *DL*=1 is the base demand in Section 3.1.1, and *DL*=2 means the demand is doubled in each neighborhood of the central district and in the periphery. The values of constraint parameters are presented as: the vehicle capacity *VC* is taken as 90. The maximum fleet size *FS* is set as 420. The policy headway H_P is taken as 30 min. Three constraints in Section 2.3.4 are considered in this subsection. Fig. 4 provides the key features of three network designs under different demand levels of three scenarios. The term *service capability* is utilized to denote the maximum number of transit trips that can be accommodated when all the constraints are simultaneously satisfied. We proceed to explore for each demand scenario:

In the first demand scenario, the service capability of Network B is appreciably higher than the other two networks, which could almost accommodate thirty thousand transit trips per hour. The local routes in Network B are placed at the city center as shown in Fig. 3a, and most of city's travel demand is distributed in this relatively small area. Thereby, Network B makes better use of its network configuration to cater for the heterogeneous demand in scenario I. In the second demand scenario, Network B still owns the highest service capability compared with Network A and C. We find that the vehicle occupancy constraint is the last possible binding constraint as the demand level increases incrementally. The service capability is obtained when the vehicle occupancy constraint is no longer satisfied. Hence, the local route service in Network B outperforms the short-turn service in Network C in coping with the alleviation of the potential in-vehicle congestion issue. However, the local routes in Network B bring about only modest savings in the generalized cost as compared to the benchmark Network A, which is explained in the previous subsection. In comparison, Network C is a better option to cope with a city's spatial variations in travel demand like scenario II. In the third demand scenario, three networks have similar performance in terms of the average passenger travel time and the average generalized cost per passenger.

In general, three networks exhibit their own advantages under a range of particular demand scenarios. From the spatial perspective, Network B with the local route service is opportune when most of city's travel demand is clustered in one or several regions and meanwhile these regions only occupy a relatively small area of the central district. When the majority of transit trips are still in the central district while they are distributed in a wider region, Network C with the short-turn strategy becomes a suitable option. When demand is further spread over a more spacious area, the central demand is not significantly larger than the peripheral demand. Since the transit system is not competitive under the circumstance of uniformly-distributed travel demand, two complementary transit services cannot take full of their advantages. Yet, Daganzo (2010) presents that the hybrid structure of Network A performs better than the transit network with only the grid structure or the hub-and-spoke structure when the demand density is spatially homogeneous throughout. From the temporal perspective, the benefits of two complementary services become incrementally prominent when the city's overall demand increases and the central district aggregates a large proportion of transit trips.

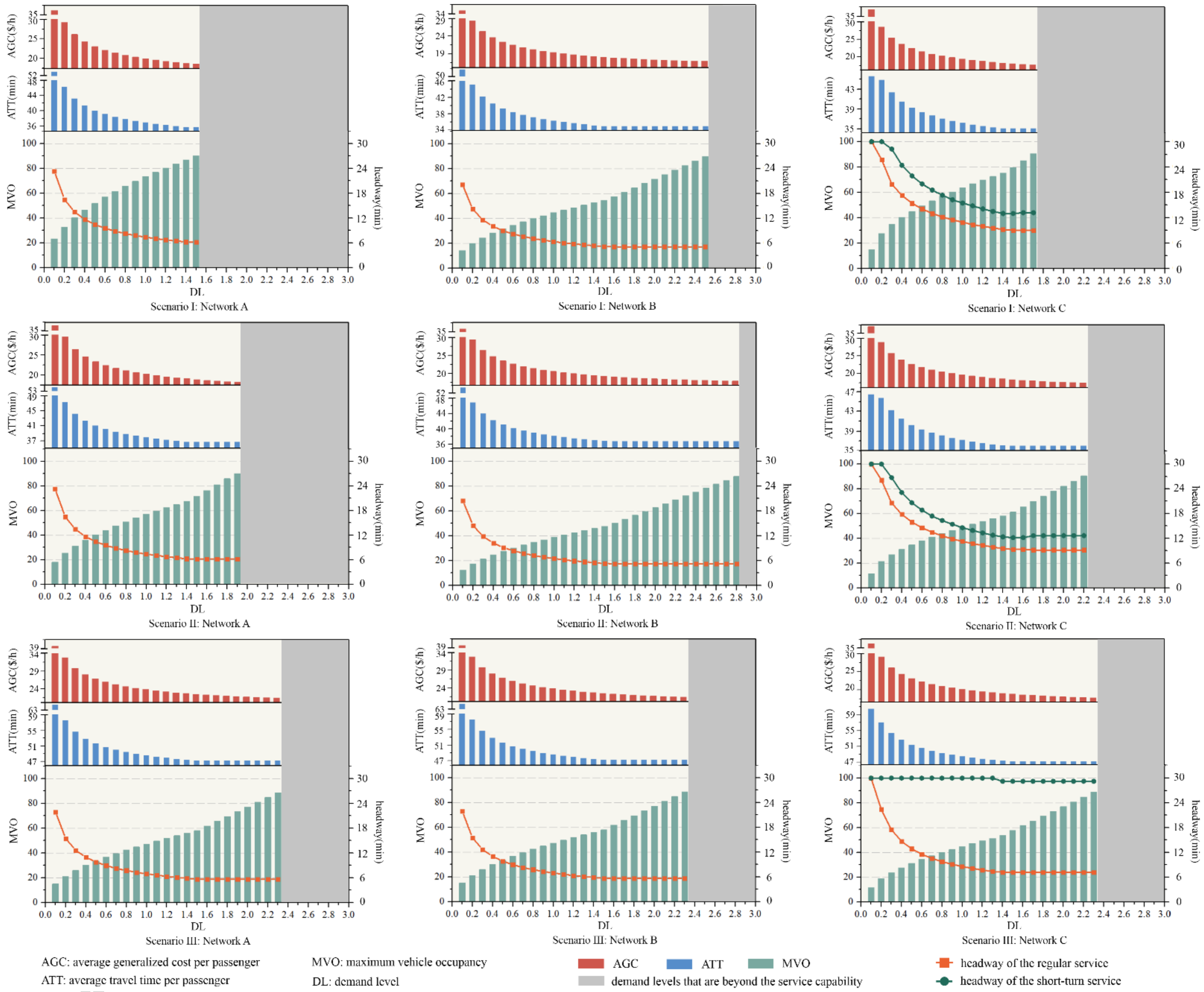


Fig. 4. Sensitivity analysis of the transit demand with all the constraints.

3.2. Case 2: α is variable

Case 2 is a relatively large city with the side length of city boundary equal to 30 km, i.e. L=W=30. The variable α is not fixed, which ranges between 1 and μ. Z_2 in Eq. (20)

and Z_3 in Eq. (21) are selected as two objective functions. The penalty cost ρ in Eq. (21) is taken as 40 $/h.

The entire city's total demand rate of transit trips is set as 20,000 pax/h. The demand density parameters in Eq. (26) are taken as a_1=0.0005, a_2=0.0632, a_3=0.48, a_{41}= a_{42}=2.4, a_5=0.48, a_{61}= a_{62}=2.4. The demand ratio coefficients in Eq. (9) are set to be $\kappa_1 = \kappa_2 = 0.6$. Then, the resulting aggregate demand of four OD patterns (pax/h): $D_{C-C} = 8{,}291$, $D_{\hat{P}-C} = 4{,}209$, $D_{C-\hat{P}} = 4{,}975$ and $D_{\hat{P}-\hat{P}} = 2{,}525$.

If all the constraints in Section 2.3.4 are considered, these restrictions may possibly influence the benefit of three heterogeneous network designs that can be achieved. In this subsection, we only consider capacity constraint to ensure that the maximum occupancy never exceeds vehicle capacity, and analyze the results of Network A, B and C. The vehicle capacity is set as 90.

3.2.1. Sensitivity analysis: effect of variable α

The variable α determines the outermost service boundary which makes the transit network design most cost-effective. Two objectives are considered to compare the performance of Network A, B and C, as Fig. 5 shows.

Fig. 5a presents the relation between variable α and objective values of Z_2 for three networks. As expected, Network B outperforms Network A and C with over 4.8% reduction in the average generalized cost per passenger when transit network only serves the central district (i.e. α = 1.0). It is because in this case, main routes have no hub-and-spoke stretch in the periphery. The short-turn service in Network C is identical to the regular service, whereas the local route service in Network B still configures denser routes and stops in areas with higher demand. With the value of variable α increasing, the benefits of two complementary services in Network B and C gradually become pronounced. For example, the average generalized costs for Network B and C drop by 5.3% and 5.4% respectively as compared to the benchmark case of Network A when α = 3.0.

Furthermore, objective values of Z_2 for three networks monotonically increase with a growth of variable α. Hence, the objective function Eq. (20) implies that the optimal service area is only to serve the central district. This is inappropriate and the reasons are twofold: (i) Eq. (20) focuses on transit trips which are in the service area. When α equals 1.0, the average trip length of passengers in the central district is relatively short. As the variable α increases, the trip length of newly added passengers whose origins or destinations approach the outermost service boundary significantly increases. Then, the average generalized cost per passenger will definitely rise up due to these new peripheral passengers. (ii) the OD pattern $D_{C\text{-}C}$ only occupies 41% of the entire city's transit trips. The case of α equal to 1.0 will cause a large reduction of potential ridership and greatly impair transit attractiveness.

The objective function Z_3 in Eq. (21) considers the inconvenience for passengers who are beyond the service area. It tries to find a trade-off between transit patronage and the generalized cost to passengers and operators combined. Fig. 5b shows the objective values of Z_3 for three networks related with changes in variable α. With the variable α increasing, the objective value of Z_3 for each network first decreases and then increases. It reaches the lowest value when variable α equals 2.3, 2.5 and 2.6 for Network A, B and C, respectively. It means that the optimal service boundary of three networks is various, and main routes in Network C serve the largest peripheral area. Compared with benchmark Network A, the local route service in Network B and the short-turn service in Network C contribute to a reduction of 2.73% and 2.21% respectively in the total social cost.

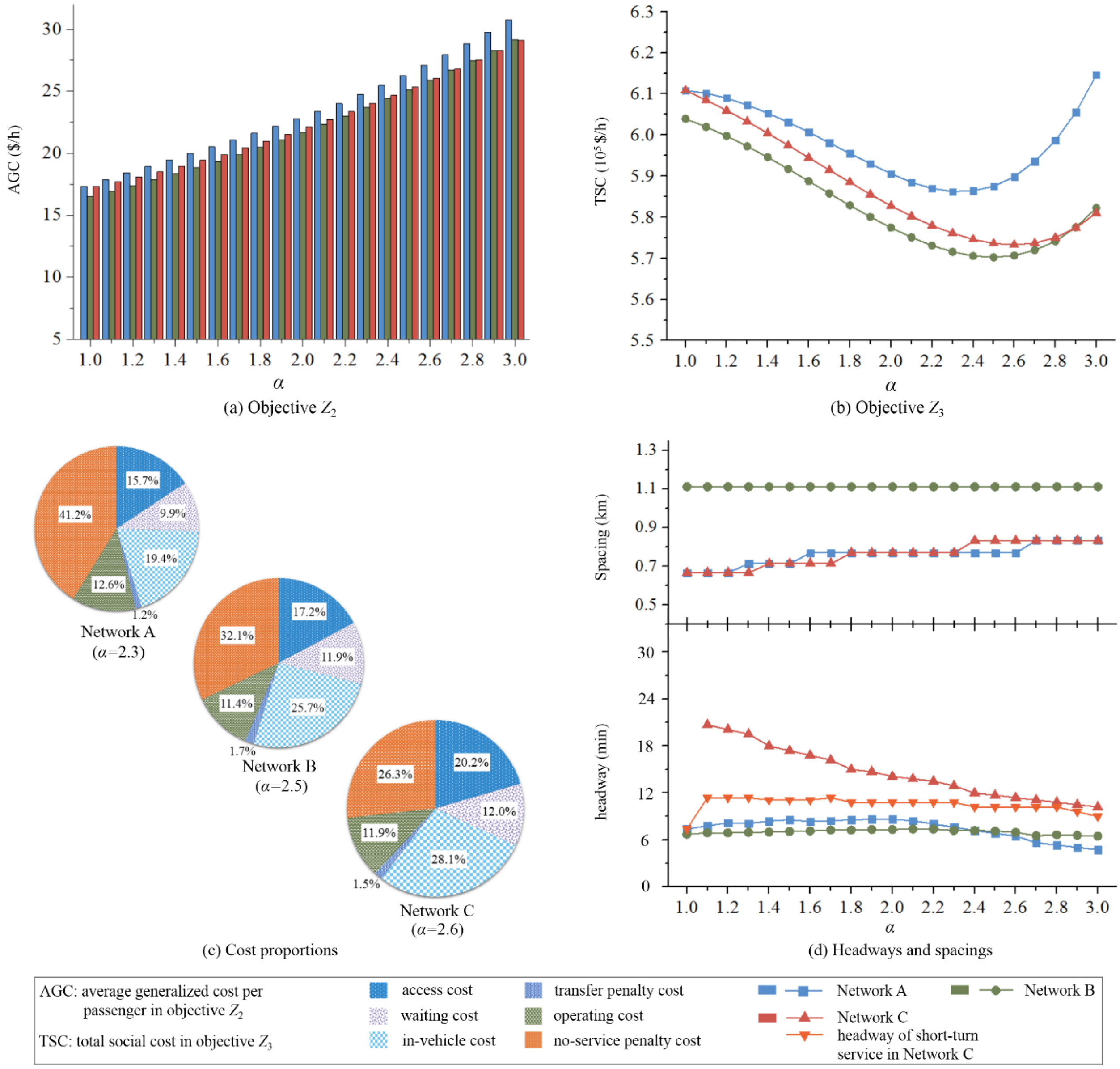


Fig. 5. Transit network design characteristics with respect to changes in variable α.

Fig. 5c intuitively presents the percentage of six cost components when Network A, B and C are in their respective optimal designs. Though there remain some differences in cost proportions of three networks, the general trends are analogous. For instance, the cost pertinent to passengers constitutes the major part of total social cost (over 85%). Figure 6d depicts the corresponding route spacings and headways of Network A, B and

C under different values of variable α. Compared with Network A and C, route spacings in Network B have lower sensitivity with respect to variable α. Headways of three networks are flexibly adjusted to handle the spatial/temporal variations in demand when the outermost service boundary changes.

3.2.2. Sensitivity analysis: effect of cost parameters

A further sensitivity analysis is conducted on the effect of cost parameters on the transit network design, as shown in Table 3. In this subsection, we use Z_3 in Eq. (21) as the objective function.

Access cost is a highly sensitive parameter for finding the optimal spacing of routes and stops. Assuming all other unit costs to be equal, if the unit value of access cost ϕ_a is decreased, the optimal spacing is increased significantly, meaning that passengers' access costs become cheaper and, thus, it is less efficient to provide a large number of transit routes. Meanwhile, as the transit routes get sparse, headways of three networks tend to decline in order to maintain a certain level of service.

Waiting cost mainly has a negative correlation with the optimal headway. With the unit value of waiting cost ϕ_a increasing, vehicles need to be dispatched more frequently to ensure that passengers' waiting time is not too long. When the headway decreases, the optimal spacing of three networks simultaneously experiences a slight ascent which may be under consideration of reducing the operating cost.

As observed from Table 3, the cost of in-vehicle time is an important factor in determining the optimal service boundary of transit network. The variable α is quite sensitive to the unit value of in-vehicle time ϕ_v. For example, if the value of ϕ_v is doubled, the optimal service area of all three networks would be curtailed to only the central district (i.e. α = 1.0). It is because as the service area extends to the farther peripheral region, the trip length of new passengers who are close to the outermost service boundary gets longer, which yields remarkably influence on passengers' in-vehicle travel cost.

Different from cost parameter ϕ_v, the variable α shows lower sensitivity with respect to the unit value of transit operating cost ϕ_o. Even the value of ϕ_o is doubled, its influence on the optimal service area is limited, e.g., the optimal value of variable α for Network A decreases from 2.7 to 2.3. The reason is that passengers' travel cost is the main cost component of total social cost. On the contrary, the operating cost only occupies around 12% of total social cost, and hence poses a relatively small effect on the route configuration of three networks. Nevertheless, the optimal headway is higher sensitive to cost parameter ϕ_o. If ϕ_o is increased, transit routes operating with high frequency are costly, and hence the optimal headway is obviously increased.

Regarding cost parameter ρ, it mainly influences the service boundary of three transit networks. The higher the value of penalty cost ρ, the larger the hub-and-spoke stretch of main routes covers in the peripheral region. To ensure the level of service, ρ should be no less than the average travel cost of passengers whose origins or destinations are in the city's periphery when transit network covers the citywide space.

Table 3 Sensitivity analysis of Cost Parameters

Cost parameters		Network A					Network B					Network C						
		fixed α [a]		variable α [b]			fixed α		variable α			fixed α			variable α			
		s_l, s_w (km)	H (min)	α	s_l, s_w (km)	H (min)	s_l, s_w (km)	H (min)	α	s_l, s_w (km)	H (min)	s_l, s_w (km)	H (min)	H_s (min)	α	s_l, s_w (km)	H (min)	H_s (min)
ϕ_a (\$/h)	10	1.43	2.76	2.8	1.43	3.12	2.00	3.96	3.0	2.00	3.96	1.67	6.90	3.62	3.0	1.67	6.90	3.62
	20	1.00	3.96	2.5	1.00	5.28	1.43	5.40	2.8	1.43	5.88	1.11	8.70	6.04	2.9	1.11	8.70	6.60
	30	0.83	4.74	2.3	0.77	7.62	1.11	6.48	2.5	1.11	7.14	0.83	10.2	9.06	2.6	0.83	11.4	10.2
	40	0.71	5.58	2.2	0.67	9.64	1.00	6.84	2.2	0.91	7.92	0.71	11.1	11.1	2.3	0.67	14.1	12.0
	50	0.63	6.36	1.9	0.59	10.2	0.91	7.32	2.0	0.83	8.88	0.67	11.7	11.4	2.0	0.59	16.8	12.6
ϕ_w (\$/h)	10	0.71	5.58	2.4	0.63	8.88	0.91	8.76	2.8	0.91	9.78	0.71	19.5	7.80	2.9	0.71	19.8	8.40
	20	0.77	5.16	2.4	0.71	7.80	1.11	7.02	2.7	1.11	8.04	0.77	13.5	8.42	2.7	0.77	14.1	10.8
	30	0.83	4.74	2.3	0.77	7.62	1.11	6.48	2.5	1.11	7.14	0.83	10.2	9.06	2.6	0.83	11.4	10.2
	40	0.91	4.38	2.3	0.83	7.02	1.25	4.86	2.3	1.11	6.24	0.91	8.40	8.08	2.4	0.83	10.5	8.70
	50	0.91	4.38	2.2	0.91	6.06	1.25	4.76	2.1	1.11	5.76	1.00	7.20	6.92	2.3	0.91	9.04	7.50
ϕ_v (\$/h)	10	0.71	5.58	3.0	0.71	5.58	0.91	7.26	3.0	0.91	7.26	0.77	10.8	9.90	3.0	0.77	10.8	9.90
	20	0.83	4.74	2.3	0.77	7.62	1.11	6.48	2.5	1.11	7.14	0.83	10.2	9.06	2.6	0.83	11.4	10.2
	30	0.91	4.38	1.6	0.77	8.34	1.25	6.12	1.7	1.25	6.84	0.91	9.94	7.80	1.7	0.77	15.6	10.8
	40	0.91	4.38	1.0	0.71	7.08	1.43	5.58	1.0	1.25	6.42	1.00	10.8	6.34	1.2	0.71	26.7	8.42
	50	1.00	3.96	1.0	0.71	7.08	1.43	5.22	1.0	1.25	6.30	1.11	8.70	6.02	1.0	0.71	7.08	-
ϕ_o (\$/h) [c]	30	0.67	3.30	2.8	0.67	3.72	1.00	3.42	2.9	1.00	3.54	0.71	5.72	5.44	3.0	0.71	5.72	5.44
	60	0.77	4.02	2.7	0.71	5.10	1.11	4.44	2.8	1.00	5.10	0.77	7.50	7.50	2.8	0.71	8.10	7.52
	120	0.83	4.74	2.3	0.77	7.62	1.11	6.48	2.5	1.11	7.14	0.83	10.2	9.06	2.6	0.83	11.4	10.2
	180	0.91	5.34	2.1	0.83	9.54	1.25	7.26	2.3	1.11	8.82	0.91	11.7	9.90	2.4	0.91	14.1	12.0
	240	0.91	6.30	2.0	0.91	10.8	1.25	8.58	2.1	1.25	8.34	1.00	13.2	10.2	2.2	0.91	17.1	13.8
ρ (\$/h)	30	0.83	4.74	1.0	0.67	7.38	1.11	4.74	1.0	1.11	6.72	0.83	10.2	9.06	1.0	0.67	7.38	-
	35	- [d]	-	1.9	0.77	8.64	-	-	1.9	1.11	7.32	-	-	-	2.0	0.77	14.1	10.8
	40	-	-	2.3	0.77	7.62	-	-	2.5	1.11	7.14	-	-	-	2.6	0.83	11.4	10.2
	45	-	-	2.7	0.83	5.64	-	-	3.0	1.11	6.48	-	-	-	3.0	0.83	10.2	9.06
	50	-	-	3.0	0.83	4.74	-	-	3.0	1.11	6.48	-	-	-	3.0	0.83	10.2	9.06

[a] the transit service covers the citywide space, i.e., $\alpha = 3.0$.
[b] the parameter α is a decision variable, which ranges between 1.0 and 3.0.
[c] the unit operating cost ϕ_o is related with vehicle capacity VC, when ϕ_o equals 30, 60, 120, 180 and 240, the corresponding values of vehicle capacity are assumed to be 50, 70, 90, 110 and 130, respectively.
[d] the network configuration plan is invariable with respect to the penalty cost ρ when α equals 3.0.

4. Conclusions

A continuum approximation (CA) modeling framework was formulated in this study to tackle the transit network design problem. The basic transit network followed the hybrid network in Daganzo (2010), with grids in the central district and hub-and-spoke structure in the periphery. Based on the CA, two models were developed to consider two complementary transit services at different planning stages. From the strategic stage, Model 1 incorporated the local route service built on the previous work of Ouyang et al. (2014). A grid of main routes is first placed at relatively large spacings in the whole central district, along with one or more grids of closely-spaced local routes that further supply the high demand area. It allows spacings and headways to vary optimally over an entire city, and extends the model in Ouyang et al. (2014) by allowing for a hub-and-spoke stretch of main routes in the periphery. Model 2 considered the short-turn tactical strategy that the main routes provide a regular service that span the entire service area, while the short-turn service allows some vehicles only to serve a part of route segment of main routes that cover areas of high demand.

The models and solution methods were numerically computed by a mono-centric city. The first numerical example analyzed a small city where transit service covers the citywide space. We compared the two CA models with the model with no complementary service under three typical demand scenarios. Results showed that the local route service and the short-turn strategy are potentially viable and effective approaches to accommodate a city's spatial variations in travel demand. The second example discussed a relatively large city where the transit service does not need cover the entire city, and hence service boundary becomes a decision variable. It was found that the performances of CA models with different complementary services are different as the outermost service boundary varies. Later, a sensitivity analysis was performed on cost parameters. The changes in cost parameters, such as access cost, were shown to greatly influence the network configurations.

Admittedly, our proposed network configurations come with some limitations, and the following improvements are suggested: (i) It could be considered to formulate the CA models that are adapt to more urban shapes and capable to simultaneously optimize the size of central district and the boundary of peripheral service area; (ii) The transit network alteration and heterogeneous passenger behaviors could be considered to better represent the practical circumstances; (iii) Research is needed to explore how to incorporate fleet management (e.g. holding and stop-skipping strategies) into the CA framework, so that transit vehicle schedules can be coordinated which yields a reduction of transfer waiting time. The authors recommend that future studies could focus on these issues.

Acknowledgements

The authors would like to thank the Editor and the anonymous reviewers for their valuable comments and suggestions to improve the quality of the article. This study is supported by the General Projects (No. 71771050) and the Youth Projects (No. 51508122) of the National Natural Science Foundation of China, the Natural Science Foundation of Jiangsu Province in China (BK20180402), and the China Postdoctoral Science Foundation (2017M621596).

Appendix A. Proofs

A.1. The demand starting from four peripheral quadrants and ending at (x, y) in the central district $D_{P-C}^{end}(x, y)$ is given by Eq. (5b).

Proof. The peripheral demand is usually sparse and its spatial variation is not distinct compared with the central demand. As abovementioned, travel demand originating from the periphery is assumed to be uniformly distributed over each peripheral quadrant. The north peripheral quadrant is first set as an example. As we assume that demand is uniformly distributed over the north peripheral quadrant, the demand density of the north edge of the central rectangle is equal to that in the entire north peripheral quadrant. Hence, the demand rate starting from the north peripheral quadrant and ending at (x, y) per unit area in the central district is approximately calculated as follows:

$$D^{end}_{P(N)-C}(x,y) \approx \frac{1}{2}\cdot\frac{1}{2}(W-w)(L+l)\cdot\frac{1}{l}\int_{\bar{x}=0}^{l}\delta(\bar{x},w,x,y)d\bar{x} = \frac{1}{4}(\alpha^2-1)w\int_{\bar{x}=0}^{l}\delta(\bar{x},w,x,y)d\bar{x}$$

$D^{end}_{P(S)-C}(x,y)$, $D^{end}_{P(E)-C}(x,y)$, $D^{end}_{P(W)-C}(x,y)$ in the other three directions are also approximately obtained as $\frac{1}{4}(\alpha^2-1)w\int_{\bar{x}=0}^{l}\delta(\bar{x},0,x,y)d\bar{x}$, $\frac{1}{4}(\alpha^2-1)l\int_{\bar{y}=0}^{w}\delta(l,\bar{y},x,y)d\bar{y}$, $\frac{1}{4}(\alpha^2-1)l\int_{\bar{y}=0}^{w}\delta(0,\bar{y},x,y)d\bar{y}$, respectively. Then, summing up these four terms proves Proposition 1. □

A.2. Two terms of the rate of onboard passengers for the eastbound direction within the central district $F_W(x,y)$ are given by Eqs. (11b)-(11c).

Proof. $f_{W(C)}(x,y)$ can be further divided into two OD patterns as shown in Fig. A1, which are denoted by $f^1_{W(C)}(x,y)$ and $f^2_{W(C)}(x,y)$ respectively. For trips with origins in the elementally-thin horizontal swath and destinations in the wider, shaded central rectangular region in Fig. A1(a), the flow of onboard passengers near (*x*, *y*) is $\frac{1}{2}\int_{x_1=0}^{x}\left[\int_{x_2=x}^{l}\int_{y_2=0}^{w}\delta(x_1,y,x_2,y_2)dy_2dx_2\right]dx_1$. The coefficient 1/2 denotes the passengers' arbitrary choice of travel direction when there is a tie between two perpendicular routes. The ratio of demand starting from (*x*, *y*) and ending at the periphery to demand starting from (*x*, *y*) and ending at the central district is assumed to be a constant value κ_1. Hence, for trips that start from the elementally-thin horizontal slice and end at the shaded peripheral service area, the onboard rate near (*x*, *y*) is the product of coefficient κ_1, $\frac{\alpha^2-1}{\mu^2-1}$ and the onboard rate with the same origins but destined for the central district.

Then, $f^1_{W(C)}(x,y)$ is equal to $\frac{1}{2}(1+\frac{\alpha^2-1}{\mu^2-1}\kappa_1)\int_{x_1=0}^{x}\left[\int_{x_2=x}^{l}\int_{y_2=0}^{w}\delta(x_1,y,x_2,y_2)dy_2dx_2\right]dx_1$.

Similarly, for trips with origins in the wider, shaded central region and destinations in the elementally-thin horizontal swath as shown in Fig. A1(b), $f^2_{W(C)}(x,y)$ is obtained in an analogous way: $\frac{1}{2}(1+\frac{\alpha^2-1}{\mu^2-1}\kappa_1)\int_{x_2=x}^{l}\left[\int_{x_1=0}^{x}\int_{y_1=0}^{w}\delta(x_1,y_1,x_2,y)dy_1dx_1\right]dx_2$. Add these two terms, we get the expression of $f_{W(C)}(x,y)$ in Eq. (11b).

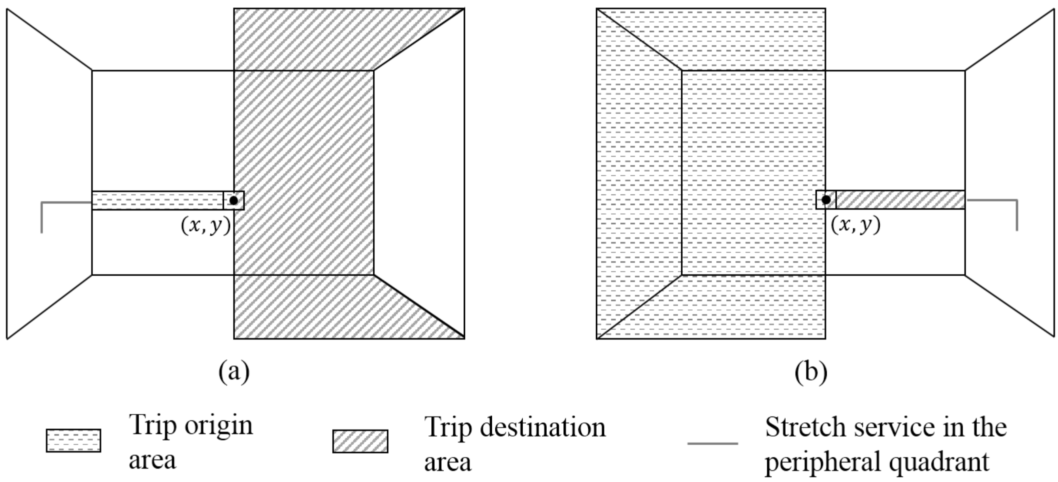


Fig. A1. The flux of onboard passengers $F_W(x, y)$ near location (x, y).

$f_{W(P)}(x,y)$ is also divided into two OD patterns as shown in Fig. A1, which are denoted by $f^1_{W(P)}(x,y)$ and $f^2_{W(P)}(x,y)$ respectively. For passengers that start from the west peripheral quadrant, they use the stretch service which passes through (x, y), and end at the shaded central district in Fig. A1(a). The number of E-W routes that pass through location (x, y) per unit distance is $1/s_W$, and we approximate that the demand on these routes which provide eastbound stretch service and end at (x, y) is $(1/s_w)/(w/s_w)\cdot D^{end}_{P(W)-C}(x,y)$. Then, the number of onboard passengers using the eastbound stretch service to pass through (x, y) and terminating at the central district is $\int_{x_2=x}^{l}\int_{y_2=0}^{w}\frac{1}{w}D^{end}_{P(W)-C}(x_2,y_2)dy_2dx_2$. Considering the constant coefficient κ_2 and Eq. (10), trips whose both origins and destinations are in the periphery of this case is acquired. Then, $f^1_{W(P)}(x,y)$ equals $(1+\frac{\alpha^2-1}{\mu^2-1}\kappa_2)\int_{x_2=x}^{l}\int_{y_2=0}^{w}\frac{1}{w}D^{end}_{P(W)-C}(x_2,y_2)dy_2dx_2$.

When passengers originate from the wide shaded peripheral quadrants and terminate at the elementally-thin horizontal swath in Fig. A1(b), these trips can be divided into three subcategories based on the locations of their origins. The west origin subcategory is first computed as $\int_{x_2=x}^{l}D^{end}_{P(W)-C}(x_2,y)dx_2$. The north and south origin subcategories are a bit different because only trips in the shaded region of these two peripheral quadrants pass through (x, y). Here, we assume that the percentage of shaded-region demand in the total quadrant demand is approximately equal to x/l, and the onboard rate of two subcategories is $\int_{x_2=x}^{l}\frac{x}{l}\left(D^{end}_{P(N)-C}(x_2,y)+D^{end}_{P(S)-C}(x_2,y)\right)dx_2$.
Multiplied by κ_2 and $\frac{\alpha^2-1}{\mu^2-1}$, trips that both origins and destinations are in the

peripheral service area of this case is acquired. Then, $f_{W(P)}^{2}(x,y)$ is equal to $(1+\frac{\alpha^2-1}{\mu^2-1}\kappa_2)\int_{x_2=x}^{l}\left[D_{P(W)-C}^{end}(x_2,y)+x/l\left(D_{P(N)-C}^{end}(x_2,y)+D_{P(S)-C}^{end}(x_2,y)\right)\right]dx_2$. Adding the two terms $f_{W(P)}^{1}(x,y)$ and $f_{W(P)}^{2}(x,y)$, we get the expression of $f_{W(P)}(x,y)$ in Eq. (11c).

□

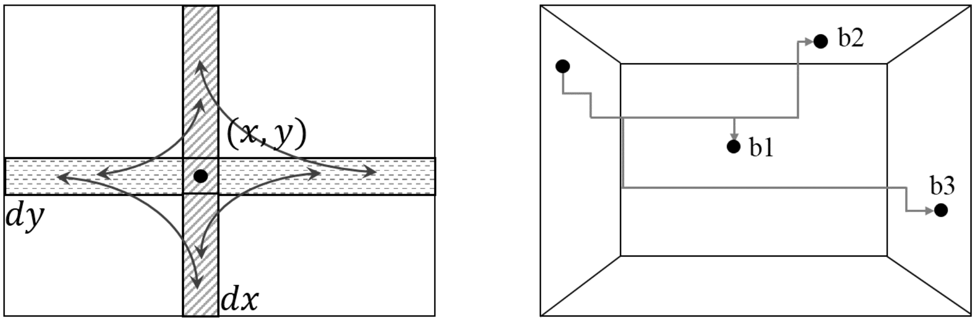
A.3. The number of directional transfers D^{d_trans} are given by Eqs. (12a) and (12b).

Proof. According to the location of trip's origin, we distinguish the following two cases.

Case I. The origin of trips is located in the central district.

The first subcategory in this case is that both the origin and destination are in the central district. Fig. A2(a) presents the directional transfer per unit area per time at (*x*, *y*) (Ouyang et al., 2014). For trips that originate within the elementally-thin vertical swath and are destined for the horizontal swath, the transfer rate equals $\frac{1}{2}\int_{\bar{x}=0}^{l}\int_{\bar{y}=0}^{w}\delta(x,\bar{y},\bar{x},y)d\bar{y}d\bar{x}$. The coefficient 1/2 again denotes the passengers' random choice of initial travel direction. Similarly, the transfer rate at (*x*, *y*) for trips with origins in the horizontal swath and destinations in the vertical one is equal to $\frac{1}{2}\int_{\bar{x}=0}^{l}\int_{\bar{y}=0}^{w}\delta(\bar{x},y,x,\bar{y})d\bar{y}d\bar{x}$.

Trips in the second subcategory are those ending at the peripheral service area. Note that passengers in this subcategory only have one choice of their travel direction if they want to complete their trips with one transfer. We previously assume that the ratio of demand starting from (*x*, *y*) and ending at the periphery to demand starting from (*x*, *y*) and ending at the central region is a constant value (i.e. κ_1). Based on this assumption and Eq. (10), the directional transfer rate for the second subcategory is the product of coefficient κ_1 , $\frac{\alpha^2-1}{\mu^2-1}$ and the transfer rate of the first subcategory, namely $\frac{1}{2}\frac{\alpha^2-1}{\mu^2-1}\kappa_1\int_{\bar{x}=0}^{l}\int_{\bar{y}=0}^{w}\left[\delta(x,\bar{y},\bar{x},y)+\delta(\bar{x},y,x,\bar{y})\right]d\bar{y}d\bar{x}$. Finally, Eq. (12a) is obtained by adding the transfer rate of two subcategories.



(a) Case I (b) Case II

Fig. A2. Directional transfer in two cases.

Case II. The origin of trips is located in the peripheral service area.

Trips in this case are classified into three subcategories as shown in Fig. A2(b). Subcategory (b1) presents trips that start from one quadrant and end at the central district. Each passenger in (b1) only needs one direction transfer to arrive at the destination, and the total number of direction transfers for subcategory (b1) is D_{P-C}. Trip ends in subcategory (b2) and (b3) are all in the peripheral quadrants. The difference is that trip end in (b3) is in the same hemisphere with its trip origin, whereas trip end in (b2) is in a different hemisphere. Trips in (b2) require one directional transfer but in (b3) two are required for passengers to complete their trips. Passengers are previously assumed to travel to the four peripheral quadrants with equal probability, and hence the demand of (b2) and (b3) is equal (i.e. $\frac{1}{2}D_{P-P}$). The total numbers of directional transfers for (b2) and (b3) are $\frac{1}{2}D_{P-P}$ and D_{P-P}, respectively. Finally, Eq. (12b) is acquired by adding the total number of directional transfers of three subcategories in Case II. □

A.4. The approximate numbers of spacing transfers D^{s_trans} for four travel directions are given by Eqs. (13b)-(13e).

Proof. A portion of passengers make a spacing transfer when there exists a bifurcation (or convergence) node for four travel directions. For instance, the case in which route bifurcation occurs in the eastbound direction at (x, y) is first considered. Two scenarios arise: (i) some onboard passengers traveling along the eastbound route whose O-D pattern is shown in Fig. A1(b) will face a route bifurcation at (x, y); (ii) some onboard passengers along the westbound route whose O-D pattern is similar to that in Fig. A1(a) but the origin/destination areas are horizontally flipped. These passengers will experience a route convergence. Approximately half of the onboard passengers in the two scenarios will perform a spacing transfer at (x, y). Hence, the transfer number in the first scenario is approximately $\frac{1}{2}\left(f_{W(C)}^{2}(x,y) + f_{W(P)}^{2}(x,y) \right)$, and in the second scenario is about $\frac{1}{2}\left(f^1_{E(C)}(x,y)+f^1_{E(P)}(x,y)\right)$.[5] As the value of $k(x,y)$ is expressed as an approximately continuous pattern, $[\partial k(x,y)/\partial x]^+$ is used to denote the case of route bifurcation in the eastbound direction at (x, y). Then, $d_W^{s-trans}(x,y)$ in Eq. (13b) is obtained.

Thanks to symmetry, the number of spacing transfers can be similarly acquired when route bifurcation occurs in the other three directions $d_E^{s-trans}(x,y)$, $d_N^{s-trans}(x,y)$ and $d_S^{s-trans}(x,y)$), which are given in Eqs. (13c)-(13e). □

A.5. The total waiting time in the periphery for Model 1 and Model 2 is given by Eq. (15c) and Eq. (15d), respectively.

Proof. The north peripheral quadrant is first calculated. Recall that the number of routes emerging from the edge of the central area is l/s_l and the frequency per route is $1/H$. Meanwhile, the number of routes crossing cordon β is approximately equal to $\beta l/s_l$, $1\le\beta\le\alpha$. Therefore, the average headway across cordon β is the reciprocal of the frequency per route $(\beta l/s_l)/[(l/s_l)/H]=\beta H$, and the associated waiting time is approximately equal to $\frac{1}{2}\beta H$. This result is common in the other three quadrants. As the distribution of passenger flow is uniform, and β has a triangular distribution in $[1,\alpha]$ with p.d.f., $2\beta/(\alpha^2-1)$, the expected waiting time per passenger in the peripheral region is $\int_{\beta=1}^{\alpha}(\frac{1}{2}\beta H)\cdot 2\beta/(\alpha^2-1)\,d\beta=\frac{\alpha^2+\alpha+1}{3(\alpha+1)}H$. Multiplied by the travel demand $D_{P-C}+D_{P-P}$, the waiting time of passengers originating from the periphery is obtained.

For Model 1, adding the waiting time resulted from the directional transfers for peripheral passengers $\frac{1}{2}HD_P^{d-trans}$, the total waiting time of peripheral passengers WT_P is achieved in Eq. (15c). For Model 2, the directional transfers for peripheral passengers in Fig. A2(b) have three types. In type b1, passengers have one directional transfer and they board the first arriving vehicle. The associated waiting time is $\frac{1}{2}\left(1/H+1/H_s\right)^{-1}D_{P-C}$. In type b2, passengers have one directional transfer and they

[5] The coefficient 1/2 here is still conservative because it ignores the case that some passengers whose origin and destination are located at the same main route. They pass through a local grid but don't not need to make a spacing transfer.

only wait the regular vehicle. The associated waiting time is equal to $\frac{1}{2}\cdot\frac{1}{2}HD_{P-P}$. In type b3, passengers have two directional transfers and their waiting time is $\frac{1}{2}\cdot(H+\frac{HH_s}{H+H_s})\frac{1}{2}D_{P-P}$. Adding the above four terms, the total waiting time of peripheral passengers WT_P is acquired in Eq. (15d). □

A.6. The total expected in-vehicle travel time in the periphery is given by Eq. (16b).

Proof. The expected in-vehicle travel time in the periphery per trip is used to represent those trips whose origins or destinations are located in the peripheral region. This can occur both inbound from the peripheral origin stop to an entry node for the central district, and outbound from an exit node of the central region to the peripheral destination stop. Thanks to symmetry, these inbound and outbound distances are described by the same random variable, R_P.

Here, the north peripheral quadrant is first considered, i.e., $R_{P(N)}$. If the origin stop is the north barrier of cordon β, the perpendicular distance to the north edge of the central region is $(1/2)(\beta\text{-}1)w$, and the expected lateral displacement is approximately one-half of the distance $(1/2)(\beta\text{-}1)l$. Hence, $E(R_{P(N)}|\beta)=(1/4)(\beta\text{-}1)(2w+l)$. As β is triangularly distributed in $[1,\alpha]$ with p.d.f., $2\beta/(\alpha^2\text{-}1)$, the expectation $E(R_{P(N)})=(1/12)[(2\alpha+1)(\alpha\text{-}1)/(\alpha+1)](2w+l)$, and the in-vehicle cruising time in the north quadrant is $E(R_{P(N)})/v$. When the origin stop is on the cordon β, the number of stops that passengers traveling in the perpendicular direction is $(1/2)(\beta\text{-}1)w/s_w$. Meanwhile, approximately one-half of vehicles are displaced laterally, and the expected number of stops that passengers traveling in the lateral direction is around one-half of the value $(1/2)(\beta\text{-}1)w/s_w$, such that the total number of stops is $(3/4)(\beta\text{-}1)w/s_w$. Since β has a triangular distribution between 1 and α, the expected in-vehicle dwelling time is equal to $(1/4)[(2\alpha+1)(\alpha\text{-}1)/(\alpha+1)]\tau w/s_w$.

The case in the south quadrant is identical with that in the north quadrant. The in-vehicle cruising time and dwelling time in other two peripheral quadrants can be similarly obtained, i.e., $(1/12)[(2\alpha+1)(\alpha\text{-}1)/(\alpha+1)](2l+w)/v$ and $(1/4)[(2\alpha+1)(\alpha\text{-}1)/(\alpha+1)]\tau l/s_l$. In this study, the peripheral demand is assumed to be equally distributed among four peripheral quadrants. Therefore, the average in-vehicle time in the periphery per trip is $(1/8)[(2\alpha+1)(\alpha\text{-}1)/(\alpha+1)](l+w)/v+(1/8)[(2\alpha+1)(\alpha\text{-}1)/(\alpha+1)]\tau(w/s_w+l/s_l)$. Passengers in D_{C-P} has one outbound trip, and in D_{P-C} has one inbound trip. For passengers in D_{P-P}, each of them has both the inbound trip and outbound trip. Thus, multiplied by $D_{C-P}+D_{P-C}+2D_{P-P}$, the total expected in-vehicle travel time in the peripheral region in Eq. (16b) is obtained. □

A.7. The operating costs for the transit network with two services are given by Eqs.

(18a)-(18c).

Proof. For transit network with local service in Model 1, the spatial density of stops at (x, y) is $s_l^{-1}(x,y)s_w^{-1}(x,y)=2^{2k(x,y)}s_l^{-1}s_w^{-1}$, and the dwelling time of four directional vehicles per unit area is $2^{2k(x,y)+2}\tau/s_l s_w$. The cruising time includes the longitudinal travel time of all vehicles and the lateral travel time of local vehicles. The longitudinal travel time of four directional vehicles per unit area is $2(s_l^{-1}(x,y)+s_w^{-1}(x,y))/v$. The lateral travel time is conservative here, which is assumed to be approximately half of the local route spacing in the perpendicular direction. The lateral travel time for E-W routes per unit area is $s_w^{-1}(x,y)\cdot\frac{1}{2}s_w(x,y)\cdot\left|\partial k(x,y)/\partial x\right|/v=\frac{1}{2v}\left|\partial k(x,y)/\partial x\right|$. and for N-S routes, the value is $\frac{1}{2v}\left|\partial k(x,y)/\partial y\right|$. Multiplied by ϕ_o, the operating cost at location (x, y) per unit area per time for Model 1, $O_C(x, y)$ in Eq. (18a) is acquired by adding three terms. It is noteworthy that in most cases, the longitudinal distance traveled by all vehicles are usually far more than the lateral distance of local buses. For the sake of simplicity, we can ignore the lateral travel time and further approximate the value of $O_C(x,y)$ in Eq. (18b) as $\frac{\phi_o 2^{k(x,y)}}{Hs_l s_w}\left[2\left(s_l+s_w\right)+2^{k(x,y)+2}\tau\right]$.

For transit network with short-turn service in Model 2, the average headway of stops at (x, y) within the central district is $(1/H+1/H_S)^{-1}$. The average cruising speed is assumed to be v, and in each headway, the total vehicular cruising time of four directional vehicles per unit area is $2(s_l^{-1}+s_w^{-1})/v$. In the neighborhood of (x, y), the spatial density of stops is $s_l^{-1}s_w^{-1}$, and in each headway, the total dwelling time of four directional vehicles per unit area is $4\tau/s_l s_w$. Multiplied by ϕ_o, the operating cost in the neighborhood of (x, y) per unit area per time for Model 2, $O_C(x, y)$ in Eq. (18b) is obtained.

The operating cost of the periphery is calculated by two different hemispheres. For the north peripheral quadrant of N-S hemisphere, the cordons are separated by s_w, and when vehicles go from one cordon to the next one, they have to travel s_w distance units perpendicularly. Also approximately half of vehicles are displaced laterally by a distance s_l when traveling along each cordon, and hence the average lateral travel distance for a vehicle is $s_l/2$. Since there are $\frac{1}{2}(W-w)/s_w=\frac{1}{2}(\alpha-1)w/s_w$ cordons in the north quadrant, the average travel time by a vehicle is $\frac{1}{2v}(\alpha-1)w(\frac{1}{2}s_l+s_w)/s_w$ and the dwelling time approximately equals $\frac{1}{2}(\alpha-1)(1+\frac{1}{2})\tau w/s_w=\frac{3}{4}(\alpha-1)\tau w/s_w$.

The number of routes emerging from the central district is l/s_l and the frequency per route is 1/*H*. The result, $(l/s_l)/H$ is then multiplied by ϕ_o and the average cruising time plus dwelling time per vehicle to obtain the total operating cost per unit time in the north quadrant: $\phi_o \frac{(\alpha-1)lw}{2s_l s_w H}\left[(\frac{1}{2}s_l + s_w)/v + \frac{3}{2}\tau\right]$.

The operating cost in south quadrant is identical to that in the north quadrant, and they are covered in two directions, then the total peripheral operating cost in the N-S hemisphere is $\phi_o \frac{2(\alpha-1)lw}{s_l s_w H}\left[(\frac{1}{2}s_l + s_w)/v + \frac{3}{2}\tau\right]$. The total peripheral operating cost in the E-W hemisphere is similarly obtained: $\phi_o \frac{2(\alpha-1)lw}{s_l s_w H}\left[(\frac{1}{2}s_w + s_l)/v + \frac{3}{2}\tau\right]$. Through adding two terms, we get O_P in Eq. (18c). Note that the peripheral operating cost for Model 1 and Model 2 shares the same formula in Eq. (18c). □

A.8. The fleet size of the two models is equal to the left-hand side of Eq. (23).

Proof. For Model 1, the total cycle time for all the vehicles traveling in the central district is equal to $\frac{1}{\phi_o}H\left(\int_{y=0}^{w}\int_{x=0}^{l}O_C(x,y)dxdy\right)$, and the headway is *H*. Note that the layover time in the depots is not considered in this study, and this is applied in the following cases. In the periphery, the total cycle time for all the vehicles providing the regular service is $\frac{1}{\phi_o}HO_P$. The peripheral headway is higher than *H* due to routes branch. In Appendix A.5, we present that the average headway across cordon β in the periphery is βH, and β has a triangular distribution in [1,α] with p.d.f., $2\beta/(\alpha^2-1)$, the approximate headway of transit vehicles in the peripheral region is $\int_{\beta=1}^{\alpha}(\beta H)\cdot 2\beta/(\alpha^2-1)\,d\beta = \frac{2}{3}\frac{\alpha^2+\alpha+1}{\alpha+1}H$. Dividing the total cycle time of two city districts by the respective headway and adding these two terms, the fleet size is acquired which is the left-hand side of Eq. (23).

For Model 2, in the central district, the total cycle time for both the short-turn vehicles and the regular vehicles is equal to $\frac{1}{\phi_o}(1/H + 1/H_s)^{-1}\left(\int_{y=0}^{w}\int_{x=0}^{l}O_C(x,y)dxdy\right)$, and the headway for these two vehicle types are *H* and H_s, respectively. The peripheral total cycle time and headway is identical with that in Model 1. Divided by the associated headway, the fleet of the short-turn service and the regular service required in the central district are obtained. Then, add the fleet of the peripheral region, and the total fleet size is obtained which is also equal to the left-hand side of Eq. (23). □

Appendix B. Notation

List of Notation

L, W	length and width of the entire city, respectively
l, w	length and width of the central district, respectively
v	average vehicle cruising speed
v_w	average walking speed of access trip
τ	average dwell time at a stop
H	headway of vehicles with the regular service and the local route service in the central district
H_s	headway of vehicles with the short-turn complementary service
H_P	policy headway
VC	vehicle capacity
FS	the maximum fleet size
s_l, s_w	base route spacing between N-S routes and E-W routes, respectively
$s_l(x, y), s_w(x, y)$	route spacing between N-S routes and E-W routes for a neighborhood centered at (x, y), respectively
$k(x, y)$	a power-of-two parameter with respect to $s_l(x, y)$ and $s_w(x, y)$
$\delta(x_1, y_1, x_2, y_2)$	demand density from a neighborhood (x_1, y_1) to another neighborhood (x_2, y_2)
$C, P, \hat{P}$	scripts that denote the central district, the peripheral service area and the periphery (including the region outside the service area), respectively
N, S, E, W	scripts that denote four peripheral quadrants as well as four travel directions, which are differentiated as North (N), South (S), East (E) and West (W).
$C-C, P-C$ $P-C, P-P$	scripts that represent four travel patterns, e.g. $P-C$ denotes the travel pattern which originates from the peripheral service area and destines for the central district
$D^{start}(x, y)$	the total generated demand starting from a neighborhood (x, y), and subscripts ($C-C, P-C, C-P$, and $P-P$) are used to represent four travel patterns
$D^{end}(x, y)$	the total attracted demand ending at a neighborhood (x, y), and subscripts ($C-C, P-C, C-P$, and $P-P$) are used to represent four travel patterns
D_{C-C}, D_{P-C} D_{C-P}, D_{P-P}	the total generated (attracted) demand of four travel patterns, e.g., D_{P-C} denotes the total peripheral demand across the whole central district
κ_1, κ_2	the demand ratio coefficient for the central origin trips and the peripheral origin trips, respectively
D^{d_trans}	the number of directional transfers
D^{s_trans}	the number of spacing transfers

$F(x,y)$	the number of onboard passengers per unit area for a neighborhood (x,y), and subscripts (N, S, E, and W) are used to represent four travel directions
AT	the walking access time
WT	the waiting time at stops, including those at transfers
IVT	the in-vehicle travel time
Pen	the extra penalties pertinent to transfers
O	the operating cost
μ	a geometric parameter, $\mu = L/l = W/w$ $(\mu > 1)$
α	a variable defining the transit service boundary, $1 \le \alpha \le \mu$
ϕ_a, ϕ_w, ϕ_v	the unit values of passenger access time, waiting time and in-vehicle travel time, respectively
θ	the penalty cost due to the inconvenience of each transfer
ρ	the penalty cost due to unsatisfied passengers who are outside the service area
ϕ_o	the unit value of transit operating cost

References

Amirgholy, M., Shahabi, M., Gao, H.O., 2017. Optimal design of sustainable transit systems in congested urban networks: A macroscopic approach. Transportation Research, Part E 103, 261-285.

Bie, Y., Gong, X., Liu, Z., 2015. Time of day intervals partition for bus schedule using GPS data. Transportation Research, Part C 60, 443-456.

Ceder, A., Stern, H.I., 1981. Deficit function bus scheduling with deadheading trip insertion for fleet size reduction. Transportation Science 15(4), 338-363.

Ceder, A., 2007. Public Transit Planning and Operation: Theory, Modeling and Practice. Elsevier, Butterworth-Heinemann, Oxford, UK.

Chen, J., Liu, Z., Zhu, S., Wang, W., 2015. Design of limited-stop bus service with capacity constraint and stochastic travel time. Transportation Research, Part E 83, 1-15.

Chen, J., Wang, S., Liu, Z., Chen, X., 2018. Network-level optimization of bus stop placement in urban areas. KSCE Journal of Civil Engineering, 22(4), 1446-1453.

Chen, P.W., Nie, Y.M., 2017. Analysis of an idealized system of demand adaptive paired-line hybrid transit. Transportation Research, Part B 102, 38-54.

Chen, P.W., Nie, Y.M., 2018. Optimal design of demand adaptive paired-line hybrid transit: Case of radial route structure. Transportation Research, Part E 110, 71-89.

Cipriani, E., Gori, S., Petrelli, M., 2012. Transit network design: A procedure and an application to a large urban area. Transportation Research, Part C 20(1), 3-14.

Cortés, C.E., Jara-Díaz, S., Tirachini, A., 2011. Integrating short turning and deadheading in the optimization of transit services. Transportation Research, Part A 45 (5), 419-434.

Daganzo, C.F., 2005. Logistics systems analysis. Springer-Verlag, Heidelberg, Germany.

Daganzo, C.F., 2010. Structure of competitive transit networks. Transportation Research, Part B 44(4): 434-446.

Estrada, M., Roca-Riu, M., Badia, H., Robusté, F., Daganzo, C.F., 2011. Design and implementation of efficient transit networks: procedure, case study and validity test. Transportation Research, Part A 45(9), 935-950.

Furth, P.G., Day, F.B., 1985. Transit routing and scheduling strategies for heavy demand corridors. Transportation Research Record 1011, 23-26.

Furth, P.G., 1987. Short turning on transit routes. Transportation Research Record 1108, 42-52.

Guan, J., Yang, H., Wirasinghe, S.C., 2003. Simultaneous optimization of transit line configuration and passenger line assignment. Transportation Research, Part B 40, 885-902.

Harris, F.W., 1990. How many parts to make at once. Operations Research 38 (6), 947-950.

Huang, D., Liu, Z., Fu, X., Blythe, P. T., 2018. Multimodal transit network design in a hub-and-spoke network framework. Transportmetrica A: Transport Science, 8, 706-735.

Ibarra-Rojas, O. J., Delgado, F., Giesen, R., Muñoz, J. C., 2015. Planning, operation, and control of bus transport systems: A literature review. Transportation Research, Part B 77, 38-75.

Li, Z.C., Lam, W.H., Wong, S.C., 2011. On the allocation of new lines in a competitive transit network with uncertain demand and scale economies. Journal of Advanced Transportation 45(4), 233-251.

Li, Z.C., Lam, W.H., Wong, S.C., Sumalee, A., 2012. Design of a rail transit line for profit maximization in a linear transportation corridor. Transportation Research, Part E 48(1), 50-70.

Lin, C.C., Chen, S.H., 2008. An integral constrained generalized hub-and-spoke network design problem. Transportation Research, Part E 44(6), 986-1003.

Liu, Z., Yan, Y., Qu, X., Zhang, Y., 2013. Bus stop-skipping scheme with random travel time. Transportation Research Part C: Emerging Technologies, 35, 46-56.

Nourbakhsh, S.M., Ouyang, Y., 2012. A structured flexible transit system for low demand areas. Transportation Research, Part B 46 (1), 204-216.

Obeng, K., Sakano, R., Naanwaab, C., 2016. Understanding overall output efficiency in public transit systems: The roles of input regulations, perceived budget and input subsidies. Transportation Research, Part E 89, 133-150.

Ouyang, Y., 2007. Design of vehicle routing zones for large-scale distribution systems. Transportation Research, Part B 41(10), 1079-1093.

Ouyang Y., Nourbakhsh S.M., Cassidy M.J., 2014. Continuum approximation approach to bus network design under spatially heterogeneous demand. Transportation Research, Part B 68, 333-344.

Petit, A., Ouyang, Y., Smith, R., 2016. Flexible-route versus fixed-route bus networks under spatially heterogeneous demand. In Transportation Research Board 95th Annual Meeting (No. 16-0534).

Roundy, R., 1985. 98%-effective integer-ratio lot-sizing for one-warehouse multiretailer systems. Management Science 31 (11), 1416-1430.

Saidi, S., Wirasinghe, S.C., Kattan, L., 2014. Rail transit: exploration with emphasis on networks with ring lines. Transportation Research Record 2419, 23-32.

Saidi, S., Wirasinghe, S.C., Kattan, L., 2016. Long-term planning for ring-radial urban rail transit networks. Transportation Research, Part B 86, 128-146.

Sheth, C., Triantis, K., Teodorović, D., 2007. Performance evaluation of bus routes: A provider and passenger perspective. Transportation Research, Part E 43(4), 453-478.

Spiess, H., Florian, M., 1989. Optimal strategies: a new assignment model for transit networks. Transportation Research, Part B 23(2), 83-102.

Szeto, W.Y., Wu, Y., 2011. A simultaneous bus route design and frequency setting problem for Tin Shui Wai, Hong Kong. European Journal of Operational Research 209, 141-155.

Szeto, W.Y., Jiang, Y., 2014. Transit route and frequency design: Bi-level modeling and hybrid artificial bee colony algorithm approach. Transportation Research, Part B 67, 235-263.

Verbas, İ.Ö., Mahmassani, H.S., 2015. Exploring trade-offs in frequency allocation in a transit network using bus route patterns: Methodology and application to large-scale urban systems. Transportation Research, Part B 81, 577-595.

Wan, Q., Lo, H., 2003. A mixed integer formulation for multiple-route transit network design. Journal of Mathematical Modelling and Algorithms 2, 299-308.

Wirasinghe, S.C., Seneviratne, P.N., 1986. Rail line length in an urban transportation corridor. Transportation Science 20(4), 237-245.

Yan, Y., Liu, Z., Bie, Y., 2016. Performance evaluation of bus routes using automatic vehicle location data. Journal of Transportation Engineering, 142(8), 04016029.

Yao, X., 2007. Where are public transit needed-Examining potential demand for public transit for commuting trips. Computers, Environment and Urban Systems 31(5), 535-550.

Yu, B., Yang, Z.Z., Jin, P.H., Wu, S.H., Yao, B.Z., 2012. Transit route network design-maximizing direct and transfer demand density. Transportation Research, Part C 22, 58-75.

Zhang, W., Xu, W., 2017. Simulation-based robust optimization for the schedule of single-direction bus transit route: The design of experiment. Transportation Research, Part E 106, 203-230.