On the optimization of the bus network design: An analytical approach based on a bi-modal Macroscopic Fundamental Diagram

Igor Dakic[a,b,*], Ludovic Leclercq[b], Monica Menendez[c]

[a]*Traffic Engineering Group, Institute for Transport Planning and Systems, ETH Zurich, Switzerland*
[b]*Univ. Gustave Eiffel, Univ. Lyon, ENTPE, LICIT, Lyon, France*
[c]*Division of Engineering, New York University Abu Dhabi, United Arab Emirates*

Abstract

Public transport systems are considered to be a crucial aspect of a sustainable urban development, as they allow more passengers to efficiently travel across an urban area at low environmental and economic costs. Multiple factors can influence the public transport level of service. All take roots in the network structure and the operating regime, i.e. how bus lines are arranged atop the street network and how the service frequency is adjusted to meet urban mobility patterns.

This is known as the bus network design problem and has been the subject of several studies. The problem is so challenging that most studies until now resort to strong assumptions such as a static description of the peak hour demand, homogeneous user behavior, and equal trip lengths. Potential effects of different types of user behavior and trip lengths patterns on the user and/or operator cost function have not been investigated whatsoever. Moreover, the existing studies have not considered the effects of the bus network structure on private car users, the level of interactions between the two modes, and the passenger mode choice that depends on the traffic conditions.

This paper aims to close this gap and provide a general framework considering multiple trip length patterns, two types of user behavior, and the effects that the bus network structure might have on the traffic performance and passenger mode choice. For modeling different trip length patterns, the proposed approach combines all origin-destination pairs with the same trip length and uses the trip length distribution as an intermediate level of abstraction. As such, it allows to solve the optimal bus network design problem in an analytical way, while considering a more realistic setting including network congestion, mixed traffic, and different mode choice decisions depending on trip lengths and walking preferences.

Numerical analysis reveals that both, the user behavior and the trip length patterns, have significant effects on the operator and user cost function. Results show that the probability of choosing any given mode is not constant across the user trip lengths, but follows certain distribution. This distribution is not unique, but varies across the trip length patterns, indicating the importance of modeling the mode choice at the trip length level. Finally, the analysis demonstrates the significance of addressing simplifications made in previous studies.

Keywords: Bus network design; Trip length distribution; Bi-modal MFD; Space allocation; Bus operations

*Corresponding author.
Email addresses: `i.dakic@ivt.baug.ethz.ch` (Igor Dakic), `ludovic.leclercq@entpe.fr` (Ludovic Leclercq), `monica.menendez@nyu.edu` (Monica Menendez)

1. Introduction and background

Multi-modality plays a critical role in mobility management. If not managed well, traffic congestion will be increasingly pervasive in urban areas. Constructing new infrastructure is an expensive solution, as the cost needed to keep pace with an increase in demand, including the induced demand, is very high (Small, 2004). Therefore, special attention should be given to the design, management, and operation of the existing road facilities for different modes. This is, however, challenging due to the complexity of the system dynamics and the lack of tools for understanding and optimizing multi-modal traffic performance under different network structures and traffic flow operations.

Motivated by the need to address the everlasting increase in travel demand, support sustainability, and preserve existing transport land use, transportation solutions are often sought in the domain of public transport systems. Public transport is often seen as a key component of a sustainable urban development. It is regarded as a public service that should provide mobility access to all citizens in an urban area.

One of the commonly explored problems in the public transport sector, more specifically in the domain of bus operations, is the bus network design. In this problem, the arrangement of bus lines atop a street network needs to be determined in a way that it provides a good level of accessibility and service between every pair of points in the city throughout the day. Several studies have looked at the optimal bus network configuration, investigating different city network structures: grid systems (Holroyd, 1967), radial systems (Byrne, 1975), systems of corridors (Wirasinghe et al., 1977), and hub-and-spoke systems (Newell, 1979). All these studies have a common objective - to determine the topological (e.g. line and stop spacings) and operational (e.g. service frequency) characteristics of the bus system that minimize the user and the operator cost.

To achieve a service and accessibility level competitive with that of the automobile at a reasonable cost, Daganzo (2010) proposed a hybrid concept, obtained by combining hub-and-spoke and grid systems. This concept has further been generalized and extended to account for more realistic network configurations (Estrada et al., 2011; Badia et al., 2014; Chen et al., 2015). The model developed by Estrada et al. (2011) also served as an inspiration to construct a real transfer-based bus network, the Nova Xarxa in Barcelona, Spain.

A recent study by Badia et al. (2017) empirically proved that such a well-designed transfer-based network could attract new users, who would not be opposed to transferring. However, most scientific literature on the bus network design problem (see e.g. Daganzo, 2010; Estrada et al., 2011; Nourbakhsh and Ouyang, 2012; Badia et al., 2014; Chen et al., 2015; Amirgholy et al., 2017) assume, up to now, that bus users choose the closest origin and destination stops such that they minimize the number of transfers, adjusting their walking distance to meet this criteria. This, in turn, may result in longer access (including the egress) time, thus longer total user time traveled, depending on the selected design parameters (e.g. stop and line spacings). Potential effects of other types of user behavior (e.g. assuming that users are willing to adjust the number of transfers in order to minimize the walking distance) on the user and/or operator cost function have not been investigated.

Furthermore, these studies have also assumed trip origins and destinations to be uniformly and independently distributed across the network, reducing thereby the complexity in the mathematical modeling. Consequently, the optimal bus network design is determined for one particular trip length pattern, imposed by the homogeneously placed origins and destinations. To the best of the authors' knowledge, the only attempt made to address this limitation can be found in Ouyang et al. (2014), who studied the optimal bus network configuration under spatially heterogeneous demand patterns.

Although the last reference investigated an irregular design of the bus network across a city, it still used a single trip length for all users in the network, similarly to all the previous studies. If, however, we look at the urban mobility patterns, we can see that such assumption is not realistic, since the trip lengths are not uniform. In addition, the authors considered only one mode (i.e. buses), without accounting for the collective effect of the analyzed topology in the global traffic performance. As a matter of fact, none of the aforementioned studies on the bus network design problem analyzed the effects of the bus network structure on private users, the level of interactions between the two modes, and the passenger mode choice that depends on the traffic conditions. A recent empirical study by Levinson (2012) found that the travel time is highly correlated with the topology of a road network. Hence, to maximize mobility in bi-modal urban systems, the impact of the bus network structure on the traffic performance should be carefully investigated.

To capture such interaction effects, we can use the recently proposed bi-modal Macroscopic Fundamental Diagram (MFD) (see e.g. Geroliminis et al., 2014; Loder et al., 2017, 2019; Dakic et al., 2019a). The bi-modal MFD allows to determine and model network-wide capacities and congestion levels, and describe city traffic macroscopically, while paying attention to multiple modes. As such, it offers new ways to analyze complex interactions between buses and cars (Dakic and Menendez, 2018), and has been mainly used to develop efficient perimeter control schemes for bi-modal urban networks (Ampountolas et al., 2017; Haitao et al., 2019; Dakic et al., 2019b, 2020). The potential to apply the bi-modal MFD for the purpose of capturing traffic conditions for the bus network design problem has not been explored whatsoever.

In this paper, we aim to close this gap and provide a general framework, capable of accounting for any trip length pattern, two types of user behavior, and the effects that the bus network structure might have on car traffic (hence the passenger mode choice). To model different trip length patterns, we propose to combine all OD pairs with the same trip length and use the trip length distribution as an intermediate level of abstraction. This way we are able to determine the optimal bus network design in an analytical way, while accounting for possibly different types of user behavior and trip length patterns for the same level of passenger demand.

Overall, the contributions of this research are sixfold. First, we account for spatially non-uniform network topology and distribution of the passenger demand across cardinal directions. Second, we use the distribution of the user trip lengths as an intermediate level of abstraction to determine the optimal design parameters. Such a level of abstraction allows to not only account for different trip length patterns (per and across cardinal directions) for the same level of passenger demand, but also to capture more accurately the network loading for all modes. Third, we consider the passenger mode choice, thus include the travel costs for both, the bus and the car mode, into the objective function. This way we optimize the performance of the whole network, while taking into account all transport modes. Fourth, we consider the influence of the passenger demand and the network topology on the traffic performance, which, in turn, affects the passenger mode choice (i.e. the demand). In other words, we use a full feedback loop to model all aforementioned dependencies. Fifth, we incorporate different lane allocations into the bus network design problem, and account for their different influences on car traffic. Six, we investigate the effects of user behavior and trip length patterns on the optimal bus network design.

For the readers convenience, Table 1 provides the list of the most important notation used in this paper.

The rest of the paper is organized as follows. Section 2 describes in detail the proposed methodological framework used to determine the optimal design parameters. In Section 3, we discuss the effects of user behavior and trip length patterns on the optimal network configuration and the passenger mode choice. Concluding remarks and future research directions are given in Section 4.

Table 1: Nomenclature.

$\mathcal{D}$	set of decision variables
$\mathcal{M}$	set of modes, indexed by $m \in \mathcal{M}$
$\mathcal{A}$	set of possible lane allocations, indexed by $a \in \mathcal{A}$
$\mathcal{P}$	set of cardinal directions of travel, indexed by $p \in \mathcal{P}$
ϕ_x/ϕ_y	network length along E-W/N-S directions
ψ_x/ψ_y	street spacing along E-W/N-S directions
s_x/s_y	stop spacing along E-W/N-S directions
l_x/l_y	line spacing along E-W/N-S directions
H_x/H_y	bus headway along E-W/N-S directions
N_x/N_y	number of corridors along E-W/N-S directions
$N_{x,b}/N_{y,b}$	number of bus lines along E-W/N-S directions
$\zeta_{x,m,a}/\zeta_{y,m,a}$	fraction of corridors with lane allocation a on which mode m operates along E-W/N-S directions
ℓ_x/ℓ_y	trip length realization along E-W/N-S directions
$f(\ell_x, \ell_y)$	joint probability density function of the user trip lengths
$TTT_{m\|\ell_x,\ell_y}$	total time traveled for mode m, for a given combination of ℓ_x and ℓ_y
$IVTT_{m\|\ell_x,\ell_y}$	in-vehicle time traveled for mode m, for a given combination of ℓ_x and ℓ_y
$\Pr(m\|\ell_x, \ell_y)$	probability of choosing mode m for a given combination of ℓ_x and ℓ_y
A_m	access (including the egress) time for mode m
W_m	waiting time for mode m
u_m	free flow speed of mode m
$\tau_{p,m,a}$	experienced car delay along a corridor with lane allocation a in direction p
$v_{p,m,a}$	operating speed of mode m for lane allocation a in direction p
$v_{p,m}$	expected speed of mode m in direction p
O_p	maximum bus occupancy in direction p
w	average walking speed
φ	expected number of transfers
α	average value of time
θ	parameter of the Logit model
δ	fixed penalty for transfers expressed in terms of an equivalent walking distance
ω	time lost per stop due to required door operations and deceleration/acceleration maneuver
ω'	time added per boarding/alighting passenger
Ω	average dwell time in the network
C	bus capacity
L_a	infrastructure length of lane allocation a
B	size of the bus fleet
V	total vehicular distance traveled by buses per hour of operation
Λ	trip generation rate during the loading time of the peak hour
λ	trip generation rate during the unloading time of the peak hour and the off-peak periods
$Z_U/Z_O/Z$	user/operator/total system cost function

2. Methodological framework

2.1. General network settings

Considered here is a bi-modal, bi-directional urban network with an average trip generation rate Λ during the loading time of the peak period, and an average trip generation rate λ during the off-peak periods and unloading time of the peak period (see Fig. 1). Notice that, unlike the previous studies on the bus network design problem, we do not consider a mean value for the passenger demand over the full peak period, but distinguish between the loading and unloading phase. This is because we are interested in analyzing how the congestion emerges across the network, allowing more realistic modeling of traffic conditions.

Let $\mathcal{M} = \{b, c\}$ be the set of modes, indexed by $m \in \mathcal{M}$, such that b denotes the bus and c denotes the car mode. Index cardinal directions of travel by $p \in \mathcal{P} = \{eb, wb, nb, sb\}$, where eb, wb, nb, and sb stand for the eastbound, westbound, northbound, and southbound directions, respectively. Similarly to previous studies (Daganzo, 2010; Ouyang et al., 2014; Chen et al., 2015; Badia et al., 2016; Amirgholy et al., 2017), we assume that the network has a grid-like structure. To make a grid representation as realistic as possible, we assume a rectangular shape with sides ϕ_x and ϕ_y, as depicted in Fig. 2. Such a rectangular form is inspired by many cities (e.g. Barcelona, Manhattan, Buenos Aires, Oslo, Helsinki, Miami, and Washington D.C) that are elongated in shape (Estrada et al., 2011). The spacings between the streets along E-W (including eb and wb) and N-S (including nb and sb) directions are ψ_x and

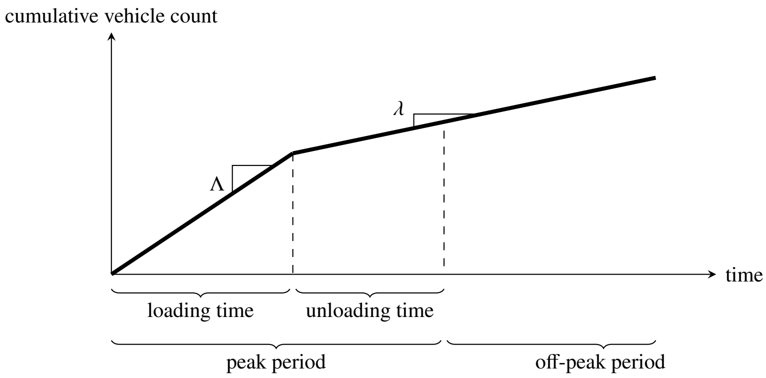

Figure 1: Modeling trip generation rate during the peak and off-peak periods.

ψ_y, respectively. Atop this street structure is the bus network. The bus network consists of two types of bus lines: (i) horizontal lines, with stop spacing s_x and line spacing l_y; and (ii) vertical lines, with stop spacing s_y and line spacing l_x (see Fig. 2).

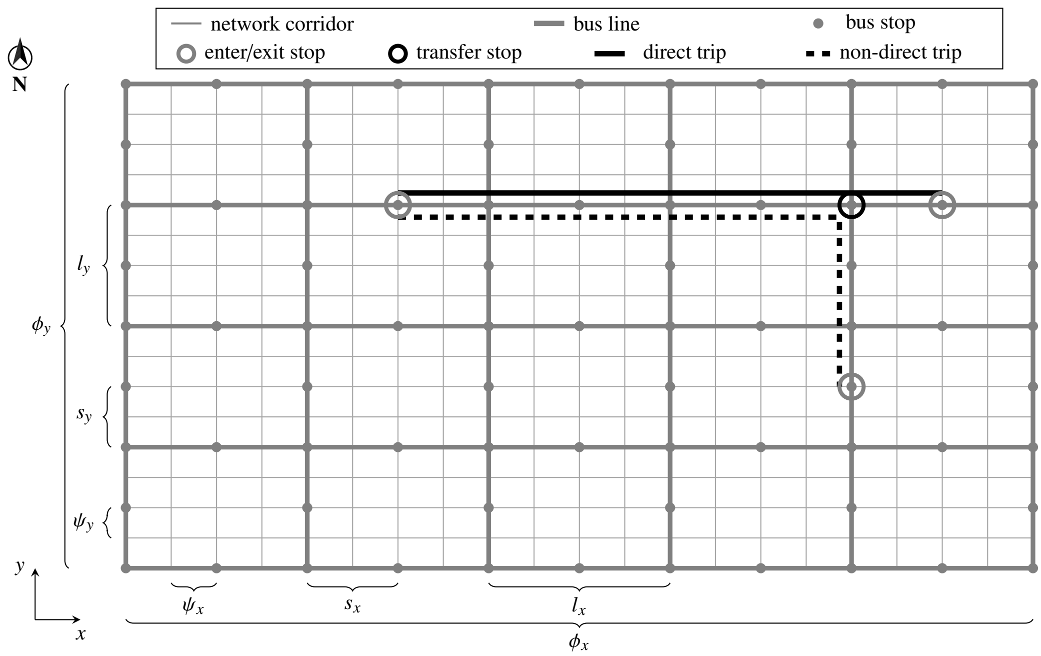

Figure 2: City structure and the corresponding bus network.

Following the approach by Estrada et al. (2011) and Fan et al. (2018), and for simplification purposes, we assume that the stop spacing includes an integer number of street segments, such that:

$$s_x = z_{s,x}\psi_x, \quad z_{s,x} \in \mathbb{N}, \tag{1a}$$
$$s_y = z_{s,y}\psi_y, \quad z_{s,y} \in \mathbb{N}. \tag{1b}$$

Likewise, the line spacing includes an integer number of stop spacings to facilitate transfers between the E-W and the N-S directions, i.e.:

$$l_x = z_{l,x}s_x = z_{l,x}z_{s,x}\psi_x, \quad z_{l,x}, z_{s,x} \in \mathbb{N}, \tag{2a}$$

$$l_y = z_{l,y} s_y = z_{l,y} z_{s,y} \psi_y, \quad z_{l,y}, z_{s,y} \in \mathbb{N}. \tag{2b}$$

Let $\{f(x_1, y_1, x_2, y_2) : x_1, x_2 \in [0, \phi_x], y_1, y_2 \in [0, \phi_y]\}$ be the probability density function describing the spatial distribution of the passenger demand across the network, generated during the peak hour. Each pair of coordinates (x_1, y_1) and (x_2, y_2) represents the corresponding origin and destination, respectively. For each OD pair, we find the user trip length as the minimum Manhattan distance for that OD:

$$\ell = |\ell_x| + |\ell_y| = |x_2 - x_1| + |y_2 - y_1|, \tag{3}$$

where ℓ_x and ℓ_y denote the user trip lengths along the E-W and N-S directions, respectively. Notice that $\ell_x \geq 0$ for *eb*, $\ell_x < 0$ for *wb*, $\ell_y \geq 0$ for *nb*, and $\ell_y < 0$ for *sb*. Then, the probability density function of the user trip lengths $\{f(\ell_x, \ell_y) : \ell_x \in [-\phi_x, \phi_x], \ell_y \in [-\phi_y, \phi_y]\}$ can be obtained by grouping all OD pairs with the same ℓ_x and ℓ_y. We consider the transformation $f(x_1, y_1, x_2, y_2) \to f(\ell_x, \ell_y)$ to be out of the scope for this study, thus assume that $f(\ell_x, \ell_y)$ is exogenously given.

It should be noted here that with the transformation $f(x_1, y_1, x_2, y_2) \to f(\ell_x, \ell_y)$ we lose the information about the locations of the trip initialization. Consequently, very spatially-heterogeneous demand distributions cannot be accurately modeled. That being said, the proposed framework is valid for all distributions of the user origins that are slowly-varying in space. The transformation allows to achieve not only a more realistic modeling of urban mobility patterns compared to the uniform distribution of user origins and destinations that comes with a unique trip distance, but also the analytical tractability of the problem.

Also note that $f(\ell_x, \ell_y)$ has no time dependency. This is reasonable, as we design our network according to the demand in the peak period, during which we consider the average traffic conditions for the loading and unloading phase, hence the trip length distribution does not vary in time. This implies that for different peak periods (i.e. morning and afternoon) we can use different trip length distributions and determine the the corresponding optimal bus network configurations. The final bus network design can then be found as the best out of the two.

In this study we consider two types of user behavior: users adjust the walking distance at the origin and the destination to minimize the number of transfers while respecting a maximum threshold for the walking distance (Type 1); and users adjust the number of transfers to minimize the walking distance at the origin and the destination (Type 2). In both cases, we assume that users arrive at the stop independently of the schedule, thus they wait on average half of the bus headway (as in Daganzo, 2010; Ouyang et al., 2014). This assumption is reasonable for relatively short headways.

In each direction of travel, the bus demand is served by buses operating with a constant headway (H_x along E-W; H_y along N-S) and identical passenger capacity C, such that $C_x = C_y = C$ (i.e. all N-S lines have the same headway which is not necessarily the same as that of the E-W lines; and all buses are the same size). Furthermore, each mode of transport can operate in either dedicated or mixed lane fashion. For the car traffic, we can further distinguish between the corridors with a dedicated bus lane and car-only lanes. These operating regimes directly affect the level of service provided by each individual mode, thus the passenger mode choice. Let $\mathcal{A}_m$ be the set of possible lane allocations for mode m, indexed by $a \in \mathcal{A}_m$. Assuming that cars can drive along all corridors in the network (i.e. the number of lanes along a corridor in any given direction p is $\eta_p \geq 2$), we set $\mathcal{A}_b = \{db, mb\}$ and $\mathcal{A}_c = \{co, db, mb\}$, where *co*, *db*, and *mb* stand for a corridor with car-only lanes, a corridor with a dedicated bus lane, and a corridor with a mixed lane, respectively. Denote by $N_{x,b}$ the number of E-W corridors with a bus line and by $N_{y,b}$ the number of N-S corridors with a bus line, such that:

$$N_{x,b} = \sum_{a \in \mathcal{A}_b} N_{x,b,a} = \lfloor \phi_y / l_y \rfloor + 1, \tag{4a}$$

$$N_{y,b} = \sum_{a \in \mathcal{A}_b} N_{y,b,a} = \lfloor \phi_x / l_x \rfloor + 1, \tag{4b}$$

where $\lfloor \cdot \rfloor$ indicates the floor function; $N_{x,b,a}$ and $N_{y,b,a}$ represent the total number of corridors with a bus line operating with lane allocation $a \in \mathcal{A}_b$ along E-W and N-S directions, respectively. Then, assuming that the lane allocation remains the same along the whole corridor and for both travel directions (*eb* and *wb* along E-W; *nb* and *sb* along N-S), the fraction of bus lines with each lane allocation type along E-W and N-S directions can simply be obtained as:

$$\zeta_{x,b,a} = N_{x,b,a} / N_{x,b}, \quad a \in \mathcal{A}_b \tag{5a}$$

$$\zeta_{x,b,a} = N_{y,b,a}/N_{y,b}, \quad a \in \mathcal{A}_b. \tag{5b}$$

Similarly, the fraction of corridors available for car traffic operating with lane allocation $a \in \mathcal{A}_c$ along E-W ($\zeta_{x,c,a}$) and N-S ($\zeta_{y,c,a}$), can be computed using Eq. 6, where N_x is the number of E-W corridors and N_y is the number of N-S corridors in the network (Eq. 7).

$$\zeta_{x,c,a} = \begin{cases} 1 - N_{x,b}/N_x, & \text{if } a = co, \\ \zeta_{x,b,a}N_{x,b}/N_x, & \text{if } a \in \{db, mb\}, \end{cases} \tag{6a}$$

$$\zeta_{y,c,a} = \begin{cases} 1 - N_{y,b}/N_y, & \text{if } a = co, \\ \zeta_{y,b,a}N_{y,b}/N_y, & \text{if } a \in \{db, mb\}, \end{cases} \tag{6b}$$

with

$$N_x = \phi_y/\psi_y + 1, \tag{7a}$$

$$N_y = \phi_x/\psi_x + 1. \tag{7b}$$

Notice from Eqs. 5–6 that we do not specify which corridors have a dedicated bus line. Instead, we stick to the fraction of dedicated bus lines along E-W and N-S directions as the global parameters for our design problem.

2.2. *Mathematical formulation of the objective function*

Here we formulate the objective function used to determine the optimal design parameters. The set of decision variables $\mathcal{D} = \{z_{s,x}, z_{s,y}, z_{l,x}, z_{l,y}, H_x, H_y, \zeta_{x,b,db}, \zeta_{y,b,db}\}$ includes stop spacings ($z_{s,x}$ and $z_{s,y}$), line spacings ($z_{l,x}$ and $z_{l,y}$), headways (H_x and H_y), and fractions of dedicated bus lines ($\zeta_{x,b,db}$ and $\zeta_{y,b,db}$) along E-W and N-S directions. The objective function is given by Eq. 8 and, similarly to the previous studies, it consists of two components: agency cost $Z_O(\mathcal{D})$; and user cost $Z_U(\mathcal{D})$.

$$\min_{\mathcal{D}} Z(\mathcal{D}) = Z_O(\mathcal{D}) + Z_U(\mathcal{D}), \tag{8}$$

$$\text{s.t. } H_x, H_y \geq H_{\min}, \tag{9}$$

$$z_{s,x}, z_{s,y}, z_{l,x}, z_{l,y} \in \mathbb{N}, \tag{10}$$

$$0 \leq \zeta_{x,b,db} \leq 1, \ 0 \leq \zeta_{y,b,db} \leq 1, \tag{11}$$

$$O_p \leq C, \ \forall p \in \mathcal{P}. \tag{12}$$

where O_p is the maximum bus occupancy in direction p.

Note that all decision variables need to be strictly positive for physical reasons. In addition, the headway in each cardinal direction should not be below the predefined threshold $H_{\min}$ (Eq. 9). The stop spacings should include an integer number of street segments, whereas the line spacings should consist of an integer number of stop spacings (Eq. 10). Moreover, the number of dedicated bus lines cannot exceed the number of bus corridors (Eq. 11). Finally, the last constraint (Eq. 12) indicates that the maximum bus occupancy in each cardinal direction p needs to be lower than the bus passenger capacity C.

The agency cost metrics include the infrastructure length of each lane allocation type $\{L_a(\mathcal{D}) : a \in \mathcal{A}_b\}$, the size of the bus fleet $B(\mathcal{D})$, and the total vehicular distance traveled by buses per hour of operation $V(\mathcal{D})$ (Eq. 13). The associated parameters $\{\pi_{L,a} : a \in \mathcal{A}_b\}$, π_B, and π_V are the corresponding unit costs. To put all the cost components in the objective function (Eq. 8) under the same unit (i.e. hours per passenger), we convert agency money into an equivalent user riding time with parameter α denoting an average value of time, and divide it by the total number of bus passengers, $\Pr(b)\lambda$. Note that we use lower trip generation rate value (λ) in Eq. 13 as is gives us the maximum bound for the operator cost, which corresponds to the worst-case scenario (see Daganzo (2010) for more details).

$$Z_O(\mathcal{D}) = [\pi_{L,db}L_{db}(\mathcal{D}) + \pi_{L,mb}L_{mb}(\mathcal{D}) + \pi_B B(\mathcal{D}) + \pi_V V(\mathcal{D})]/[\alpha \Pr(b)\lambda], \tag{13}$$

with

$$\Pr(m) = \int_{-\phi_x}^{\phi_x}\int_{-\phi_y}^{\phi_y} \Pr(m\,|\,\ell_x, \ell_y) f(\ell_x, \ell_y)\,\mathrm{d}\ell_y\,\mathrm{d}\ell_x, \quad m \in \mathcal{M}, \tag{14}$$

where $\Pr(m\,|\,\ell_x, \ell_y)$ stands for the probability of choosing mode m for a given combination of ℓ_x and ℓ_y. This probability can be computed by applying the Logit model (Bhat and Guo, 2004), as given by Eq. 15, where θ is the parameter of the Logit model, and $TTT(\mathcal{D})_{m\,|\,\ell_x,\ell_y}$ denotes the total time traveled for a given mode m and combination of ℓ_x and ℓ_y.

$$\Pr(m\,|\,\ell_x, \ell_y) = \frac{\exp(-\theta \cdot TTT(\mathcal{D})_{m\,|\,\ell_x,\ell_y})}{\sum_{m'\in\mathcal{M}} \exp(-\theta \cdot TTT(\mathcal{D})_{m'\,|\,\ell_x,\ell_y})}, \quad m \in \mathcal{M}. \tag{15}$$

On the other hand, the user cost metric of interest is the expected total time traveled, computed across both modes and the entire set of the user trip lengths (Eq. 16).

$$Z_U(\mathcal{D}) = \int_{-\phi_x}^{\phi_x}\int_{-\phi_y}^{\phi_y} TTT(\mathcal{D})_{m\,|\,\ell_x,\ell_y} \Pr(m\,|\,\ell_x, \ell_y) f(\ell_x, \ell_y)\,\mathrm{d}\ell_y\,\mathrm{d}\ell_x. \tag{16}$$

In the following, we derive the components of the operator cost function. We start by formulating the infrastructure length (Eq. 17). This variable can simply be obtained by multiplying the length associated with one bus line by the total number of bus lines with a given lane allocation along E-W and N-S directions.

$$L_a(\mathcal{D}) = \begin{cases} \zeta_{x,b,db} N_{x,b}\phi_x + \zeta_{y,b,db} N_{y,b}\phi_y, & \text{if } a = db, \\ (1-\zeta_{x,b,db}) N_{x,b}\phi_x + (1-\zeta_{y,b,db}) N_{y,b}\phi_y, & \text{if } a = mb. \end{cases} \tag{17}$$

The bus fleet consists of buses operating in all cardinal directions (Eq. 18). In each direction of travel, the number of required buses can be approximated by multiplying the number of buses required for one bus line by the total number of bus lines.

$$B(\mathcal{D}) = \sum_{p\in\{eb,wb\}} N_{x,b}\phi_x/(H_x v_{p,b}) + \sum_{p\in\{nb,sb\}} N_{y,b}\phi_y/(H_y v_{p,b}), \tag{18}$$

where $v_{p,b}$ stands for the expected bus speed (including the dwell and acceleration/deceleration time) in direction p. This expected bus speed can be determined using Eq. 19, where $v_{p,m,a}$ denotes the operating speed of mode m for lane allocation a in direction p.

$$v_{p,m} = \begin{cases} \sum_{a\in\mathcal{A}_m} v_{p,m,a}\zeta_{x,m,a}, & \text{if } p \in \{eb, wb\}, \\ \sum_{a\in\mathcal{A}_m} v_{p,m,a}\zeta_{y,m,a}, & \text{if } p \in \{nb, sb\}, \end{cases} \quad m \in \mathcal{M} \tag{19}$$

Note that the operating speeds $v_{p,m,a}$ are not predefined, but depend on the network loading and the topology in a given cardinal direction. This will be further elaborated in Section 2.4. Following the approach by Daganzo (2010), we formulate the bus operating speed as follows:

$$v_{p,b,a} = \begin{cases} [1/u_b + \omega/s_x + \Omega]^{-1}, & \text{if } p \in \{eb, wb\},\ a = db \\ [1/u_b + \omega/s_y + \Omega]^{-1}, & \text{if } p \in \{nb, sb\},\ a = db, \\ [1/v_{p,c,mb} + \omega/s_x + \Omega]^{-1}, & \text{if } p \in \{eb, wb\},\ a = mb, \\ [1/v_{p,c,mb} + \omega/s_y + \Omega]^{-1}, & \text{if } p \in \{nb, sb\},\ a = mb, \end{cases} \tag{20}$$

where u_b is the bus free-flow speed; $v_{p,c,mb}$ denotes the operating car speed along a mixed lane in direction p; ω represents the time lost per stop due to required door operations and deceleration/acceleration maneuvers; Ω stands for the average dwell time in the network. To estimate Ω, we first need to define the last agency cost metric, the total vehicular distance traveled by buses per hour of operation $V(\mathcal{D})$. This distance can be obtained by multiplying the total number of buses operating in all cardinal directions by the corresponding expected bus speeds (Eq. 21).

$$V(\mathcal{D}) = 2[N_{x,b}\phi_x/H_x + N_{y,b}\phi_y/H_y]. \tag{21}$$

Then, the average dwell time in the network can be determined using Eq. 22, where ω' denotes the time added per boarding/alighting passenger and $\varphi(\mathcal{D})$ is the expected number of transfers. To get the maximum bound for the user cost function, we use higher trip generation rate value (Λ) in Eq. 22, similarly to Daganzo (2010) and Estrada et al. (2011).

$$\Omega = \omega' \Pr(b)\Lambda[1 + \varphi(\mathcal{D})]/V(\mathcal{D}). \tag{22}$$

Finally, the last term in the objective function is the total time traveled for a given mode m and combination of the user trip lengths. It consists of the expected access (including the egress) time $A_m(\mathcal{D})$, the expected waiting time $W_m(\mathcal{D})$, the in-vehicle time traveled $IVTT(\mathcal{D})_{m\,|\,\ell_x,\ell_y}$ for a given ℓ_x and ℓ_y, and some additional time components (Eq. 23). For the bus mode, these additional components include the transferring time, whereas for the car mode, the additional time traveled is related to the equivalent mileage costs of an auto trip (as in Daganzo, 2010).[1]

$$\begin{aligned} TTT(\mathcal{D})_{m\,|\,\ell_x,\ell_y} &= A_m(\mathcal{D}) + W_m(\mathcal{D}) + IVTT(\mathcal{D})_{m\,|\,\ell_x,\ell_y} \\ &\quad + \mathbb{1}_{\{m=b\}}\varphi(\mathcal{D})\delta/w + \mathbb{1}_{\{m=c\}}(|\ell_x| + |\ell_y|)\pi_D/\alpha, \quad m \in \mathcal{M}, \end{aligned} \tag{23}$$

with

$$IVTT(\mathcal{D})_{m\,|\,\ell_x,\ell_y} = |\ell_x|/v_{\xi(\ell_x),m} + |\ell_y|/v_{\xi(\ell_y),m}, \quad m \in \mathcal{M}, \tag{24}$$

where δ is a fixed penalty for transfers expressed in terms of an equivalent walking distance; w denotes the average walking speed (accounting for the delays that users encounter when crossing streets); π_D represents the auto cost per unit distance; $\mathbb{1}_{\{condition\}}$ is an indicator function that return the value of 1 if *condition* is satisfied; $\xi(\cdot)$ stands for a function that maps each trip length component (ℓ_x and ℓ_y) to a cardinal direction in which it is traversed, such that: $\xi(\ell_x \geq 0) = eb$; $\xi(\ell_x < 0) = wb$; $\xi(\ell_y \geq 0) = nb$; and $\xi(\ell_y < 0) = sb$.

The following equations approximate the network-level components of the user cost function for the bus mode (Eq. 23), including the access, the waiting, and the transferring time, for each type of user behavior. For Type 1, Eqs. 25–27 represent an extension of those in Estrada et al. (2011) for more general bus network configurations. On the other hand, for Type 2, Eqs. 28–30 require a more detailed explanation. Their derivation is provided in the Appendix.

Type 1. (adjusting the walking distance to minimize the number of transfer)

$$A_{1,b}(\mathcal{D}) = [l_x + s_y + l_y + s_x]/[4w]. \tag{25}$$

$$\varphi_1(\mathcal{D}) = \Pr(\varphi > 0) = 1 - \Pr(\varphi = 0) = 1 - [l_y\phi_x + l_x\phi_y - l_xl_y]/[\phi_x\phi_y]. \tag{26}$$

$$W_{1,b}(\mathcal{D}) = [1 + \varphi_1(\mathcal{D})]\bar{H}/2. \tag{27}$$

where $\Pr(\varphi = 0)$ and $\Pr(\varphi > 0)$ are the probabilities for making zero and non-zero transfers, respectively; $\bar{H} = [N_{b,x}H_x + N_{b,y}H_y]/[N_{b,x} + N_{b,y}]$ denotes the weighted (by the number of bus lines in the corresponding cardinal directions) bus headway in the network.

Type 2. (adjusting the number of transfer to minimize the walking distance)

$$A_{2,b}(\mathcal{D}) = \begin{cases} [6l_xl_y + 6l_xs_x - 2l_y^2 - 3l_ys_x + 3l_ys_y]/[12l_xw], & \text{if } l_x \geq l_y, \\ [6l_xl_y + 6l_ys_y - 2l_x^2 - 3l_xs_y + 3l_xs_x]/[12l_yw], & \text{if } l_x < l_y. \end{cases} \tag{28}$$

$$\varphi_2(\mathcal{D}) = \begin{cases} \Pr(\varphi > 0)[4l_x^2l_y^2 - 2l_xl_y^3 + l_y^4]/[2l_x^2l_y^2], & \text{if } l_x \geq l_y, \\ \Pr(\varphi > 0)[4l_x^2l_y^2 - 2l_x^3l_y + l_x^4]/[2l_x^2l_y^2], & \text{if } l_x < l_y. \end{cases} \tag{29}$$

$$W_{2,b}(\mathcal{D}) = \begin{cases} \Pr(\varphi = 0)\bar{H}/2 + \Pr(\varphi > 0)[H_x(8l_x^2l_y^2 - 4l_xl_y^3 + l_y^4) + H_y(4l_x^2l_y^2 + l_y^4)]/[8l_x^2l_y^2], & \text{if } l_x \geq l_y, \\ \Pr(\varphi = 0)\bar{H}/2 + \Pr(\varphi > 0)[H_x(4l_x^2l_y^2 + l_x^4) + H_y(8l_x^2l_y^2 - 4l_x^3l_y + l_x^4)]/[8l_x^2l_y^2], & \text{if } l_x < l_y. \end{cases} \tag{30}$$

[1] Alternatively, one can also incorporate the fare for using the bus mode and/or congestion tolls, parking, and garaging costs associated with the car mode.

2.3. Derivation of the maximum bus occupancy

Now that we have derived all components of the objective function, we can compute the maximum bus occupancy O_p in each cardinal direction p. This variable is important to ensure that the bus service provided is enough to accommodate the passenger demand generated in the most restrictive links during the peak hour.

To compute O_p, we have to evaluate the maximum number of on-board passengers Q_p across all points in the corresponding cardinal direction. For this purpose, we use vertical and horizontal cordons denoted as $\beta\phi_x$ and $\beta\phi_y$, respectively, such that $\beta \in [0, 1]$ (see Fig. 3). Intuitively, vertical cordons are used for Q_{eb} and Q_{wb}, whereas the horizontal ones are used for Q_{nb} and Q_{sb}. Recall that due to the transformation $f(x_1, y_1, x_2, y_2) \to f(\ell_x, \ell_y)$ we do not know the locations of the trip initialization. Hence, for each cordon, we have to determine the probability that bus users cross the cordon for a given ℓ_x (in case of a vertical cordon) or ℓ_y (in case of a horizontal cordon), in which case they are considered as part of the total number of on-board passengers in the corresponding cardinal direction. These probabilities are denoted as $\Pr(J_\beta(\ell_x))$ and $\Pr(J_\beta(\ell_y))$, where $J_\beta(\ell_x)$ is the event that bus users travel across vertical cordon $\beta\phi_x$ for a given ℓ_x, and $J_\beta(\ell_y)$ is the event that bus users travel across horizontal cordon $\beta\phi_y$ for a given ℓ_y. Given the assumption regarding the slowly-varying spatial distribution of the user origins, the aforementioned probabilities can be computed as:

$$\Pr(J_\beta(\ell_x)) = \begin{cases} |\ell_{x,b}|/(\phi_x - |\ell_{x,b}|), & \text{if } |\ell_{x,b}| \le \min\{\beta\phi_x, (1-\beta)\phi_x\}, \\ \min\{\beta\phi_x, (1-\beta)\phi_x\}/(\phi_x - |\ell_{x,b}|), & \text{if } \min\{\beta\phi_x, (1-\beta)\phi_x\} < |\ell_{x,b}| \le \max\{\beta\phi_x, (1-\beta)\phi_x\}, \\ 1, & \text{if } |\ell_{x,b}| > \max\{\beta\phi_x, (1-\beta)\phi_x\}, \end{cases} \tag{31a}$$

$$\Pr(J_\beta(\ell_y)) = \begin{cases} |\ell_{y,b}|/(\phi_y - |\ell_{y,b}|), & \text{if } |\ell_{y,b}| \le \min\{\beta\phi_y, (1-\beta)\phi_y\}, \\ \min\{\beta\phi_y, (1-\beta)\phi_y\}/(\phi_y - |\ell_{y,b}|), & \text{if } \min\{\beta\phi_y, (1-\beta)\phi_y\} < |\ell_{y,b}| \le \max\{\beta\phi_y, (1-\beta)\phi_y\}, \\ 1, & \text{if } |\ell_{y,b}| > \max\{\beta\phi_y, (1-\beta)\phi_y\}, \end{cases} \tag{31b}$$

where $\phi_x - |\ell_x|$ and $\phi_y - |\ell_y|$ represent the domains of valid user origins for the corresponding ℓ_x and ℓ_y (see Fig. 3).

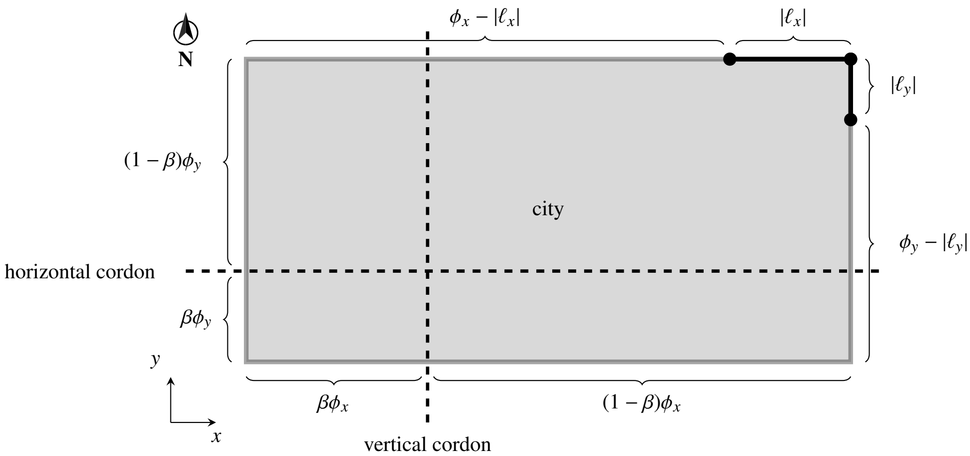

Figure 3: A city and its vertical and horizontal cordons.

Then, the maximum bus occupancy can simply be obtained by dividing the maximum number of on-board passengers across all bus lines and all cordons in a given cardinal direction by the total number of bus lines and the respective bus frequency (Eq. 32). To get the maximum number of on-board passengers, we use the higher trip generation rate in Eq. 33, i.e. the value of Λ observed during the loading time of the peak period.

$$O_p = \begin{cases} Q_p H_x / N_{x,b}, & \text{if } p \in \{eb, wb\}, \\ Q_p H_y / N_{y,b}, & \text{if } p \in \{nb, sb\}, \end{cases} \tag{32}$$

with

$$Q_{eb} = \max_{\beta\in[0,1]} \left\{ \int_{0}^{\phi_x} \int_{-\phi_y}^{\phi_y} \Pr(J_\beta(\ell_x))\Lambda \Pr(b\,|\,\ell_x,\ell_y) f(\ell_x,\ell_y)\,\mathrm{d}\ell_y\,\mathrm{d}\ell_x \right\}, \tag{33a}$$

$$Q_{wb} = \max_{\beta\in[0,1]} \left\{ \int_{-\phi_x}^{0} \int_{-\phi_y}^{\phi_y} \Pr(J_\beta(\ell_x))\Lambda \Pr(b\,|\,\ell_x,\ell_y) f(\ell_x,\ell_y)\,\mathrm{d}\ell_y\,\mathrm{d}\ell_x \right\}, \tag{33b}$$

$$Q_{nb} = \max_{\beta\in[0,1]} \left\{ \int_{0}^{\phi_y} \int_{-\phi_x}^{\phi_x} \Pr(J_\beta(\ell_y))\Lambda \Pr(b\,|\,\ell_x,\ell_y) f(\ell_x,\ell_y)\,\mathrm{d}\ell_x\,\mathrm{d}\ell_y \right\}, \tag{33c}$$

$$Q_{sb} = \max_{\beta\in[0,1]} \left\{ \int_{-\phi_y}^{0} \int_{-\phi_x}^{\phi_x} \Pr(J_\beta(\ell_y))\Lambda \Pr(b\,|\,\ell_x,\ell_y) f(\ell_x,\ell_y)\,\mathrm{d}\ell_x\,\mathrm{d}\ell_y \right\}. \tag{33d}$$

Notice from Eq. 32 the importance of providing different bus headways along E-W and N-S directions. In other words, a unique bus headway for the entire network might lead to an efficient bus service (measured by the ability to serve the total generated demand with the available bus capacity) in some directions. However, other directions might experience a lower level of service, i.e. significantly higher bus occupancy, due to its less favorable design parameters (e.g. less number of bus lines, lower fraction of dedicated bus lines, or longer stop spacing), constrained by the geometrical characteristics of the network. In other words, a unique bus headway, as it has been previously proposed in most of the literature (see e.g. Daganzo, 2010; Estrada et al., 2011; Ouyang et al., 2014; Amirgholy et al., 2017), might not lead to the optimal design parameters in all cardinal directions of travel. Therefore, by considering different bus headways, we balance the bus level of service provided in each direction, preventing such scenario.

To verify Eqs. 31–33, we show that the solution for the maximum bus occupancy in case of uniformly and independently distributed origins and destinations is the same to that of Daganzo (2010) (see the Appendix for more details).

2.4. *Derivation of the operating car speed for each lane allocation type*

As mentioned before, in the problem investigated in this study, there are three types of corridor in each cardinal direction of travel p for which the operating speed $v_{p,m,a}$ needs to be estimated: (i) corridors with car-only lanes; (ii) corridors with a dedicated bus lane; and (iii) corridors with a mixed lane. For corridors with a dedicated bus lane, the operating bus speed $v_{p,b,db}$ is not affected by the traffic conditions, i.e. interactions with the car mode (assuming that a dedicated bus lane operates with a transit signal priority). Hence, it can simply be modeled as the free-flow bus speed u_b (including the dwell and acceleration/deceleration time) (Eq. 20). Likewise, assuming that buses and cars drive at the same free-flow speed (as in e.g. Daganzo, 2010; Zheng and Geroliminis, 2013; Dakic and Menendez, 2018), i.e. $u_b = u_c$, the operating bus speed for corridors with a mixed lane, $v_{p,b,mb}$, can be modeled as the operating car speed along a mixed lane in the corresponding direction of travel, $v_{p,c,mb}$ (including the dwell time and acceleration/deceleration time) (Eq. 20). As a result, for each cardinal direction, we only need to estimate the operating car speeds.

Since corridors have regular configuration along E-W and N-S directions, the operating car speed can be determined using the bi-modal Macroscopic Fundamental Diagram (MFD) (see e.g. Geroliminis et al., 2014; Loder et al., 2017, 2019; Dakic et al., 2019a). The bi-modal MFD is defined as $P_{p,c,a} = P_{p,c,a}(n_{p,c,a})$, where $P_{p,c,a}$ is the total travel production and $n_{p,c,a}$ is the total car accumulation along a corridor with lane allocation a in direction p (Fig. 4). Its analytical approximation is based on the variational formulation of kinematic wave theory (VT) (Daganzo, 2005a,b; Daganzo and Menendez, 2005; Daganzo and Geroliminis, 2008; Leclercq and Geroliminis, 2013). VT uses a shortest path (or least cost) formulation to solve complex kinematic wave problems given a concave flow-density relationship. As the description of VT and the method of cuts applied to the bi-modal MFD estimation (see Fig. 4) is out of the scope for this study, we refer the readers to Dakic et al. (2019a) for more details.

To estimate $v_{p,c,a}$, we first need to determine whether the total generated car demand $\Lambda_{p,c,a}$ (Eq. 34) along a corridor with allocation a in direction p during the loading time T of the peak period (Fig. 5) is smaller than the corresponding corridor capacity $\mu_{p,c,a}$ ($\mu_{p,c,a} = P^*_{p,c,a}/\phi_x$ along E-W; $\mu_{p,c,a} = P^*_{p,c,a}/\phi_y$ along N-S). From that perspective, we can distinguish the following two cases: $\Lambda_{p,c,a} \le \mu_{p,c,a}$; and $\Lambda_{p,c,a} > \mu_{p,c,a}$. The first case implies uncongested traffic conditions, in which the corridor capacity is enough to serve the total generated demand, hence no delays due to the

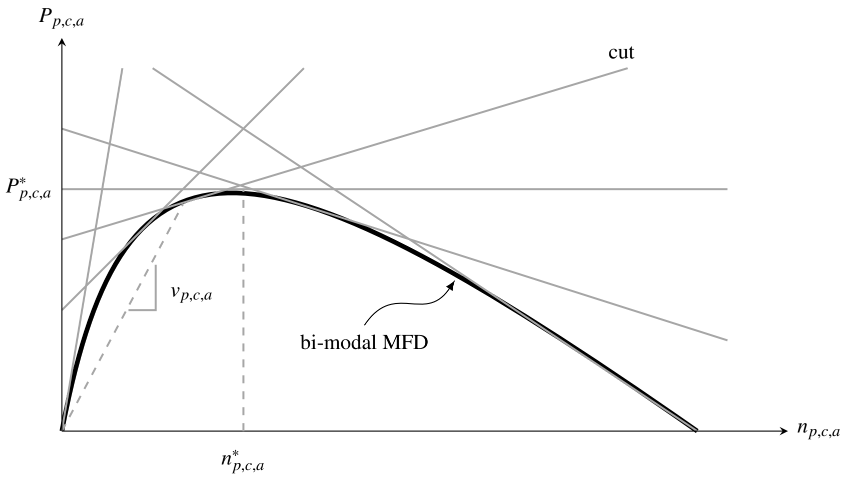

Figure 4: A family of cuts and the bi-modal MFD-based operating car speed for a given traffic state and lane allocation.

congestion are experienced in the system. Consequently, there is a direct match between the total generated demand and the mean travel production along a corridor. In the second case, however, such a direct match between the demand and the mean travel production can no longer be established, as the mean production might be limited to the corridor capacity during the time when congestion is active. Therefore, to tackle this problem, we follow the approach by Amirgholy et al. (2017) and use a simple bottleneck model that considers the entire corridor as a single bottleneck with a mean capacity (see Amirgholy et al. (2017) for more details). Below we summarize how the operating car speed is computed for each case separately.

$$\Lambda_{eb,c,a} = \int_0^{\phi_x} \int_{-\phi_y}^{\phi_y} [\Lambda / N_x(\eta_p - \mathbb{1}_{\{a=db\}})] \Pr(c \,|\, \ell_x, \ell_y) f(\ell_x, \ell_y) \, \mathrm{d}\ell_y \, \mathrm{d}\ell_x, \tag{34a}$$

$$\Lambda_{wb,c,a} = \int_{-\phi_x}^{0} \int_{-\phi_y}^{\phi_y} [\Lambda / N_x(\eta_p - \mathbb{1}_{\{a=db\}})] \Pr(c \,|\, \ell_x, \ell_y) f(\ell_x, \ell_y) \, \mathrm{d}\ell_y \, \mathrm{d}\ell_x, \tag{34b}$$

$$\Lambda_{nb,c,a} = \int_0^{\phi_y} \int_{-\phi_x}^{\phi_x} [\Lambda / N_y(\eta_p - \mathbb{1}_{\{a=db\}})] \Pr(c \,|\, \ell_x, \ell_y) f(\ell_x, \ell_y) \, \mathrm{d}\ell_x \, \mathrm{d}\ell_y, \tag{34c}$$

$$\Lambda_{sb,c,a} = \int_{-\phi_y}^{0} \int_{-\phi_x}^{\phi_x} [\Lambda / N_y(\eta_p - \mathbb{1}_{\{a=db\}})] \Pr(c \,|\, \ell_x, \ell_y) f(\ell_x, \ell_y) \, \mathrm{d}\ell_x \, \mathrm{d}\ell_y. \tag{34d}$$

In Eq. 34 we assume that the car capacity for a corridor with a dedicated bus line is reduced by one lane compared to other lane allocation types.

Case 1. ($\Lambda_{p,c,a} \leq \mu_{p,c,a}$): the network operates in the uncongested regime. The operating car speed can be determined by finding the point on the bi-modal MFD in which the total travel production is equivalent to that obtained by applying the queuing formula of Little (1961):

$$P_{p,c,a} = \Lambda_{p,c} \bar{\ell}_p, \quad p \in \mathcal{A}_c, \; p \in \mathcal{P}. \tag{35}$$

where $\bar{\ell}_p$ is the average trip length in direction p (Eq. 36).

$$\bar{\ell}_{eb} = \frac{\int_0^{\phi_x} \int_{-\phi_y}^{\phi_y} |\ell_x| f(\ell_x, \ell_y) \, \mathrm{d}\ell_y \, \mathrm{d}\ell_x}{\int_0^{\phi_x} \int_{-\phi_y}^{\phi_y} f(\ell_x, \ell_y) \, \mathrm{d}\ell_y \, \mathrm{d}\ell_x}, \tag{36a}$$

$$\bar{\ell}_{wb} = \frac{\int_{-\phi_x}^{0} \int_{-\phi_y}^{\phi_y} |\ell_x| f(\ell_x, \ell_y) \, \mathrm{d}\ell_y \, \mathrm{d}\ell_x}{\int_{-\phi_x}^{0} \int_{-\phi_y}^{\phi_y} f(\ell_x, \ell_y) \, \mathrm{d}\ell_y \, \mathrm{d}\ell_x}, \tag{36b}$$

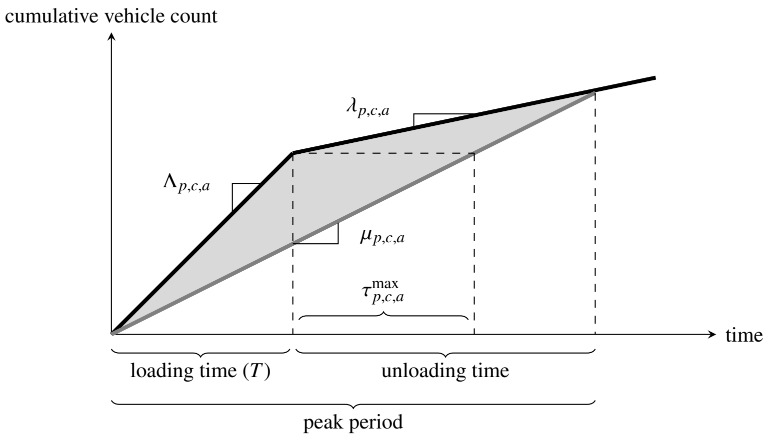

Figure 5: Computing the average car delay experienced along a corridor with allocation a in direction p.

$$\bar{\ell}_{nb} = \frac{\int_0^{\phi_y} \int_{-\phi_x}^{\phi_x} |\ell_y| f(\ell_x, \ell_y)\, \mathrm{d}\ell_x\, \mathrm{d}\ell_y}{\int_0^{\phi_y} \int_{-\phi_x}^{\phi_x} f(\ell_x, \ell_y)\, \mathrm{d}\ell_x\, \mathrm{d}\ell_y}, \tag{36c}$$

$$\bar{\ell}_{sb} = \frac{\int_{-\phi_y}^{0} \int_{-\phi_x}^{\phi_x} |\ell_y| f(\ell_x, \ell_y)\, \mathrm{d}\ell_x\, \mathrm{d}\ell_y}{\int_{-\phi_y}^{0} \int_{-\phi_x}^{\phi_x} f(\ell_x, \ell_y)\, \mathrm{d}\ell_x\, \mathrm{d}\ell_y}. \tag{36d}$$

Case 2. ($\Lambda_{p,c,a} > \mu_{p,c,a}$): the network operates at capacity. Due to the congestion, cars experience certain delay $\tau_{p,c,a}$, which needs to be taken into account when computing $v_{p,c,a}$. Following the approach by Amirgholy et al. (2017), we can compute this delay as half of the maximum car delay $\tau^{\max}_{p,c,a}$ experienced during the peak hour, i.e. $\tau_{p,c,a} = \tau^{\max}_{p,c,a}/2 = [T(\Lambda_{p,c,a} - \mu_{p,c,a})/\mu_{p,c,a}]/2$ (see Fig. 5). The operating car speed is then given as:

$$v_{p,c,a} = [1/v^*_{p,c,a} + \tau_{p,c,a}]^{-1} = [n^*_{p,c,a}/P^*_{p,c,a} + \tau_{p,c,a}]^{-1}, \quad p \in \mathcal{P},\ a \in \mathcal{A}_c \tag{37}$$

3. Numerical results

3.1. Solution algorithm

For the purpose of computing the optimal design parameters, we follow the approach of previous studies and use a grid search (see e.g. Daganzo, 2010; Estrada et al., 2011; Badia et al., 2016). To speed up the searching process over the feasible region, we perform a Monte Carlo sampling for a given trip length pattern. The result is the set of trip length values along E-W ($\mathcal{L}_x$) and N-S ($\mathcal{L}_y$) directions. Then we find the network-level parameters of the bus user cost function according to the selected type of user behavior (i.e. minimizing the number of transfers or the walking distance). To compute the total travel cost for each mode, we then need to find the operating speeds. These operating speeds depend on the mode choice, which is, on the other hand, a function of the utility cost function (i.e. the travel time) for each mode. In other words, the total travel costs are determined by solving a fixed point problem, as shown by the flow chart in Fig. 6. This flow chart summarizes the methodological steps.

Considering that the set $\mathcal{R}$ of potential mode choice values (Pr(m)) discretized by 1% is limited, we use a simple grid search method rather than a conventional convergence algorithm to solve the fixed point problem. For each mode choice value $r \in \mathcal{R}$, we compute its absolute difference Δ_r to the mode choice value given by the Logit model (Eqs. 14–15), and find the one (r^*) with the minimum difference, i.e. $r^* = \arg\min_r\{\Delta_r\}$. Since we discretize our set $\mathcal{R}$ by 1%, the computed difference Δ_{r^*} is always less than 1%. Once the fixed point problem is solved, we compute the operator and user costs, hence the total system cost.

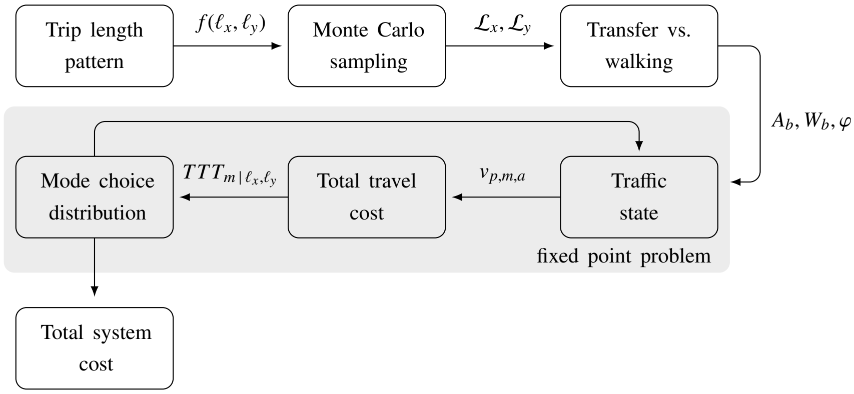

Figure 6: Schematic representation of the solution algorithm.

3.2. *Optimal design parameters and the effects of user behavior and trip length patterns*

Here we investigate the effects of user behavior and trip length patterns on the optimal bus network design and passenger mode choice (i.e. traffic performance) by applying the previously described framework to the Barcelona network. Table 2 summarizes the input parameters for this case study. The values are similar to those in the literature (see e.g. Daganzo, 2010; Estrada et al., 2011; Badia et al., 2014, 2016). Similarly to Ouyang et al. (2014), we explore four distinct trip length patterns, reflecting different spatial demand distributions: (i) a uniform city (Fig. 7a), where origins and destinations are uniformly distributed across the entire network; (ii) a mono-centric city (Fig. 7c), where most of the origins and destinations are located within the city center, resulting in shorter trip lengths compared to (i) (see Fig. 7d); (iii) a commuter city (Fig. 7e), where most of the origins and destinations are located in two regions positioned in the opposite corners of the network, resulting in longer trip lengths compared to (i) (see Fig. 7f); and (iv) a twin city (Fig. 7g), where most of the origins and destinations are located in two nearly-adjacent regions near the city center.

Table 2: Experimental settings.

Input parameter	Variable	Units	Value
Network length	ϕ_x/ϕ_y	km	10/4.95
Street spacings	ψ_x/ψ_y	km	0.25/0.15
Equivalent penalty distance per transfer	δ	km	0.03
Trip generation rate during the peak hour	Λ	pax/hr	75000
Trip generation rate during the off-peak hour	λ	pax/hr	30000
Loading time of the peak period	T	hr	1
Bus free-flow speed	u_b	km/hr	40
Average walking speed	w	km/hr	2
Time lost per stop	ω	sec	30
Boarding and alighting time per passenger	ω'	sec/pax	1
Unit dedicated-infrastructure cost	$\pi_{L,db}$	\$/km-hr	90
Unit mixed-infrastructure cost	$\pi_{L,mb}$	\$/km-hr	9
Unit vehicle cost	π_B	\$/veh-hr	40
Unit distance cost	π_V	\$/veh-km	2
Unit mileage cost	π_D	\$/km	0.3
Value of time	α	\$/pax-hr	20
Minimum headway	$H_{\min}$	min	3
Bus passenger capacity	C	pax/veh	150
Average car waiting time	W_c	min	7
Parameter of the Logit model	θ	$\emptyset$	1.5

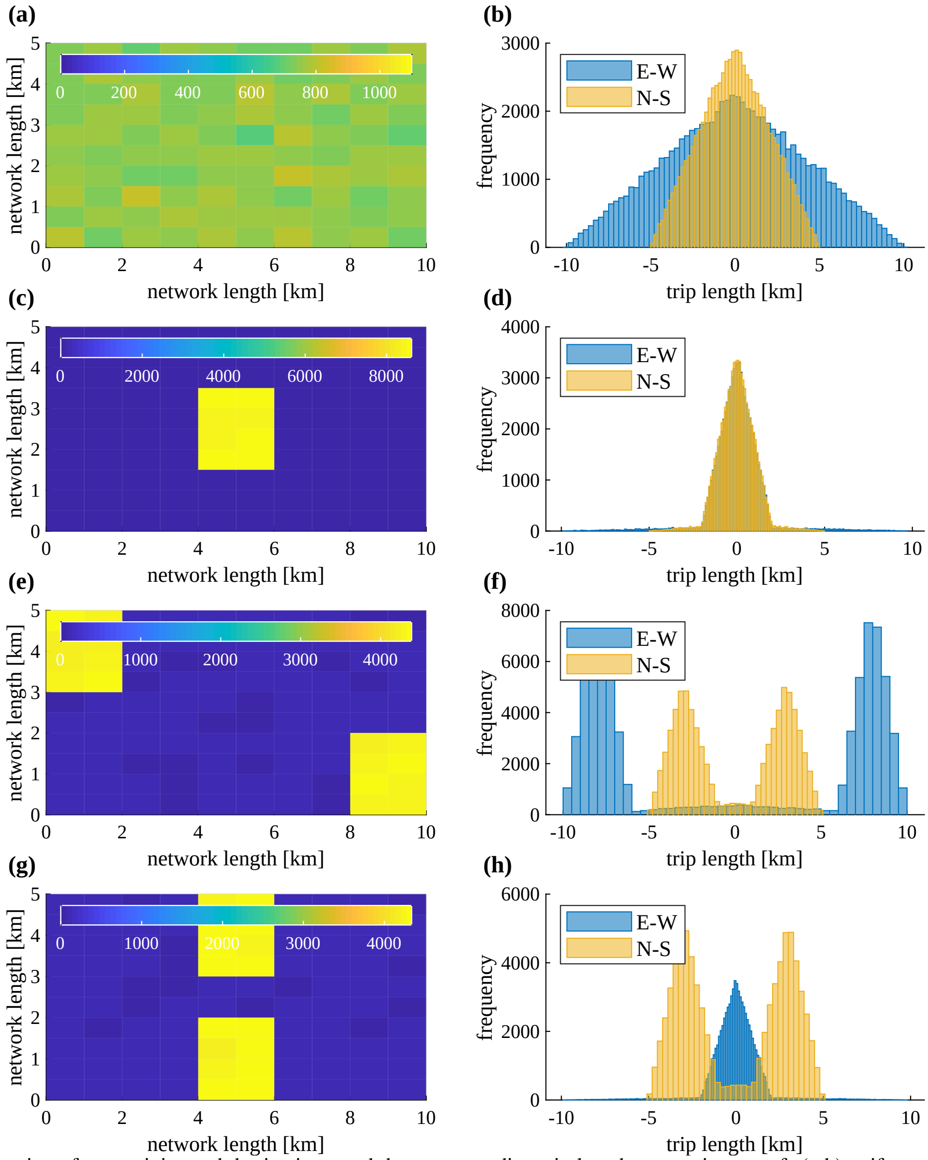

Figure 7: Concentration of user origins and destinations, and the corresponding trip length pattern in case of: (a-b) uniform city; (c-d) a mono-centric city; (e-f) a commuter city; and (g-h) a twin city.

Table 3 shows the results, including the optimal design parameters, operator and user costs, as well as the corresponding mode choice, for each type of user behavior and trip length pattern. The headways are given in minutes, whereas the costs are expressed in hours. First thing to notice is that both, the user behavior and the trip length patterns, affect the optimal bus network configuration and the passenger mode choice. The results suggest that for longer trip length patterns more users tend to choose the bus system. This is due to the fact that, besides the in-vehicle time traveled, the mileage cost component in the car user cost function (defined as in Daganzo, 2010) also linearly increases with an increase in the trip length (see Eq. 23). This is not the case for the bus mode, as all the cost components, except for the in-vehicle time traveled, remain the same, i.e. they are independent of the user trip length.

We also observe that for different trip length patterns the optimal bus network configuration does not always provide a fully dedicated or mixed (with car traffic) bus system across the entire network (i.e. all cardinal directions), as it was the case in the previous studies (see e.g. Daganzo, 2010; Estrada et al., 2011; Amirgholy et al., 2017). In fact, most of the decision variables vary along E-W and N-S, indicating the importance of considering a heterogeneous bus network design across cardinal directions. This holds true for both types of user behavior.

Table 3: Optimal design parameters, operator cost, user cost, and the corresponding mode choice.

User behavior	Trip pattern	$z_{s,x}$	$z_{s,y}$	$z_{l,x}$	$z_{l,y}$	H_x	H_y	$\zeta_{x,b,db}$	$\zeta_{y,b,db}$	Z_O	Z_U	Z	Pr(b)
Type 1	(i)	2	3	2	2	6	5	0	0.5	0.065	0.538	0.603	0.36
	(ii)	1	2	3	3	8	6	0	0	0.061	0.357	0.419	0.36
	(iii)	3	3	1	2	3	3	1	1	0.112	0.826	0.939	0.46
	(iv)	1	3	3	2	7	4	0	0.9	0.081	0.479	0.560	0.40
Type 2	(i)	2	4	4	1	5	3	0	1	0.068	0.491	0.559	0.42
	(ii)	1	4	7	1	6	4	0	0.7	0.061	0.311	0.372	0.42
	(iii)	3	3	1	2	3	4	1	1	0.099	0.792	0.891	0.48
	(iv)	1	4	7	1	6	3	0	1	0.064	0.437	0.501	0.44

A closer look into the mode choice distribution reveals that the probability of choosing a given mode m is not constant across the user trip lengths, but follows certain distribution (Fig. 8). This distribution is not unique, but varies across the trip length patterns, indicating the importance of modeling the mode choice at the trip length level. Interestingly, the distribution for Type 2 behavior shows higher deviation and is shifted towards right (i.e. higher number of users choose to use the bus system) compared to Type 1. The reason for this is that the bus system designed according to Type 2 behavior provides higher level of service, as elaborated further below.

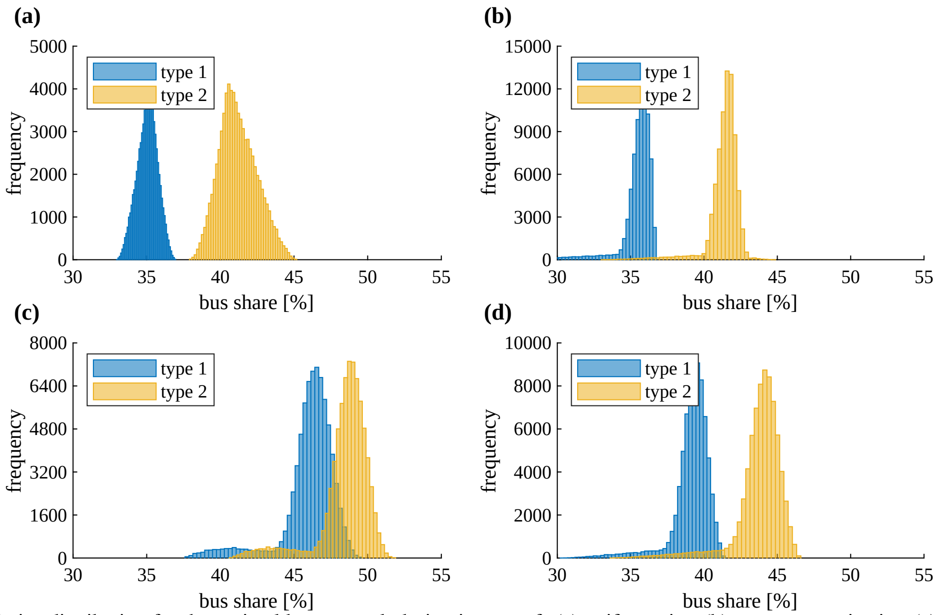

Figure 8: Mode choice distribution for the optimal bus network design in case of: (a) uniform city; (b) a mono-centric city; (c) a commuter city; and (d) a twin city.

For quantifying the effects of user behavior, we compare the objective functions of Type 1 and Type 2, such that $\Delta Z_O = [Z_{2,O} - Z_{1,O}]/Z_{1,O}$, $\Delta Z_U = [Z_{2,U} - Z_{1,U}]/Z_{1,U}$, and $\Delta Z = [Z_2 - Z_1]/Z_1$ (Table 4). We observe that for all trip length patterns users can benefit (according to the cost functions used here) if they are willing to adjust the number of transfers to minimize the walking distance at the origin and the destination. This is consistent with empirical results from the city of Barcelona (Badia et al., 2017). Potential improvements can range from 4.2% to 12.9%, depending on the prevailing trip length pattern.

To achieve the user willingness to adjust the number of transfers, i.e. provide a well-designed transfer-based network, the user waiting time needs to be minimized. This is indicated by the optimal decision variables in Table 3, where we can see that the optimal bus network configuration for Type 2 behavior provides higher bus frequency compared to Type 1 (i.e. lower headways in the two directions for almost all types of cities). As a result, the operator cost might increase for some trip length patterns (e.g. uniform city), as shown in Table 4. However, since the operator costs are significantly lower than the user costs, the total cost function still gets reduced if the network is designed according to Type 2 user behavior. The improvements can range from 5.1% to 11.1% (see Table 4).

Table 4: Effects of user behavior on the cost functions.

Trip pattern	ΔZ_O	ΔZ_U	ΔZ
(i)	+4.46%	-8.85%	-7.42%
(ii)	-1.02%	-12.87%	-11.13%
(iii)	-11.62%	-4.22%	-5.1%
(iv)	-21.24%	-8.83%	-10.63%

Similarly, the effects of trip length patterns are quantified by comparing the costs obtained when the bus network parameters are optimally determined for a given trip length pattern, with those obtained if we were to assume uniformly distributed origins and destinations (Table 5). A negative value means that accounting for the actual trip length distribution does lead to lower costs. As expected, such lower total costs are achieved for both types of user behavior, with a similar magnitude of improvements. The highest improvement (3.2% for Type 1; 3.5% for Type 2) is acquired for the longest trip length pattern (a commuter city), whereas a twin city exhibits the lowest improvement in the total cost function (0.1% for Type 1; 1.8% for Type 2). Notice that in some cases the operator costs increase when we use the actual trip length distribution (e.g. for a commuter city). Nevertheless, such increases are compensated with a much more efficient level of service from the user perspective, leading to overall lower costs.

Table 5: Effects of trip length patterns on the cost functions.

User behavior	Trip pattern	ΔZ_O	ΔZ_U	ΔZ
Type 1	(ii)	-5.48%	-0.24%	-1.04%
	(iii)	+109.96%	-9.83%	-3.24%
	(iv)	+38.46%	-4.6%	-0.09%
Type 2	(ii)	-10.44%	-1.06%	-2.73%
	(iii)	+57.35%	-8.01%	-3.55%
	(iv)	-1.36%	-1.81%	-1.75%

It is also worth mentioning that the optimal bus network design for a uniform city that does not always satisfy the constraint for the maximum bus occupancy (Eq. 12) when applied to other trip length patterns, thus cannot be considered to be optimal. This is the case for a commuter city, where an increase in the operator cost (imposed by considering the actual trip length distribution) is justified by ensuring that the bus capacity is enough to serve the maximum bus occupancy during the peak hour.

4. Conclusions

In this study we investigate the effects of user behavior and trip length patterns on the optimal bus network design, as well as the effects that the bus network structure might have on the traffic performance and passenger mode choice. For this purpose, we introduce several major extensions to the classical problem formulation. First, we consider two types of user behavior when selecting bus routes: (i) minimization of the number of transfers; and (ii) minimization of the walking distance. Second, we propose a new concept for considering trip length heterogeneity related to different origin-destination trips in the network. Third, we explicitly account for mixed traffic (i.e. mutual influence of cars and buses in the network) and mode choice at the trip level. Forth, we use a dynamic description of the peak hour demand by distinguishing between the loading and unloading phase. This way, we are able to achieve a more realistic modeling of network congestion, hence determine a mean speed that depends on the demand profile and mode choice.

Overall, the contributions of this research are sixfold: (i) we account for spatially non-uniform network topology and distribution of the passenger demand across cardinal directions; (ii) we use the distribution of the user trip lengths as an intermediate level of abstraction to determine the optimal design parameters. Such a level of abstraction allows not only to account for different trip length patterns (per and across cardinal directions) for the same level of passenger demand, but also to capture more accurately the network loading for all modes; (iii) we consider the passenger mode choice, thus include the travel costs for both, the bus and the car mode, into the objective function. This way we optimize the performance of the whole network, while taking into account all transport modes; (iv) we consider the influence of the passenger demand and the network topology on the traffic performance, which, in turn, affects the passenger mode choice (i.e. the demand). In other words, we use a full feedback loop to model all aforementioned dependencies; (v) we incorporate different lane allocations into the bus network design problem, and account for their different influences on car traffic; and (vi) we investigate the effects of user behavior and trip length patterns on the optimal bus network design; all in an analytical manner.

Numerical experiments performed for different trip length patterns, reflecting different spatial demand distributions, show that both, the user behavior and the trip length patterns, affect the optimal bus network configuration. Results reveal that the probability of choosing any given mode is not constant across the user trip lengths, but follows certain distribution. This distribution is not unique, but varies across the trip length patterns, indicating the importance of modeling the mode choice at the trip length level. Finally, we demonstrate the significance of addressing simplifications made in previous studies. From that perspective, we show that the bus network parameters optimally determined for a given trip length pattern lead to lower costs compared to those obtained if we were to assume uniformly distributed origins and destinations. In other words, accounting for the actual trip length distribution allows the operator to design the bus network in accordance to the prevailing demand pattern.

5. Acknowledgments

This work was supported by the Swiss National Science Foundation (SNSF) under the project name DIPLOMAT, contract 205121_165644. Ludovic Leclercq acknowledges funding support by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 646592 – MAGnUM project). Monica Menendez acknowledges funding support by the NYUAD Center for Interacting Urban Networks (CITIES), funded by Tamkeen under the NYUAD Research Institute Award CG001 and by the Swiss Re Institute under the Quantum Cities™ initiative.

Appendix: Proofs

Result 1. *The expected walking time at the origin and the destination for Type 2 user behavior is given by Eq. 28:*

$$A_{2,b}(\mathcal{D}) = \begin{cases} [6l_x l_y + 6l_x s_x - 2l_y^2 - 3l_y s_x + 3l_y s_y]/[12l_x w], & \text{if } l_x \geq l_y, \\ [6l_x l_y + 6l_y s_y - 2l_x^2 - 3l_x s_y + 3l_x s_x]/[12l_y w], & \text{if } l_x < l_y. \end{cases}$$

Proof. Recall that for this type of behavior we assume users to choose the closest origin and destination stops such that they minimize the walking distance. Consequently, how the users access/exit the bus system (using a vertical or a horizontal bus line) depends on where their origin/destination is located within a rectangle defined by the horizontal and vertical line spacings (see Fig. 9). This, in turn, defines the walking distance (thus the access time), as well as the number of transfers (thus the waiting time). From that perspective, we can distinguish four regions within the rectangle, as shown in Fig. 9. Note that if the origin/destination falls within region (1) or (3), the users access/exit the bus system at the stop closest to their origin/destination along the nearest horizontal bus line. Otherwise, they access/exit the bus system at the closest stop along the nearest vertical bus line. As the shape of the rectangle and the corresponding regions depends on the ratio between l_x and l_y, we consider the following two cases:

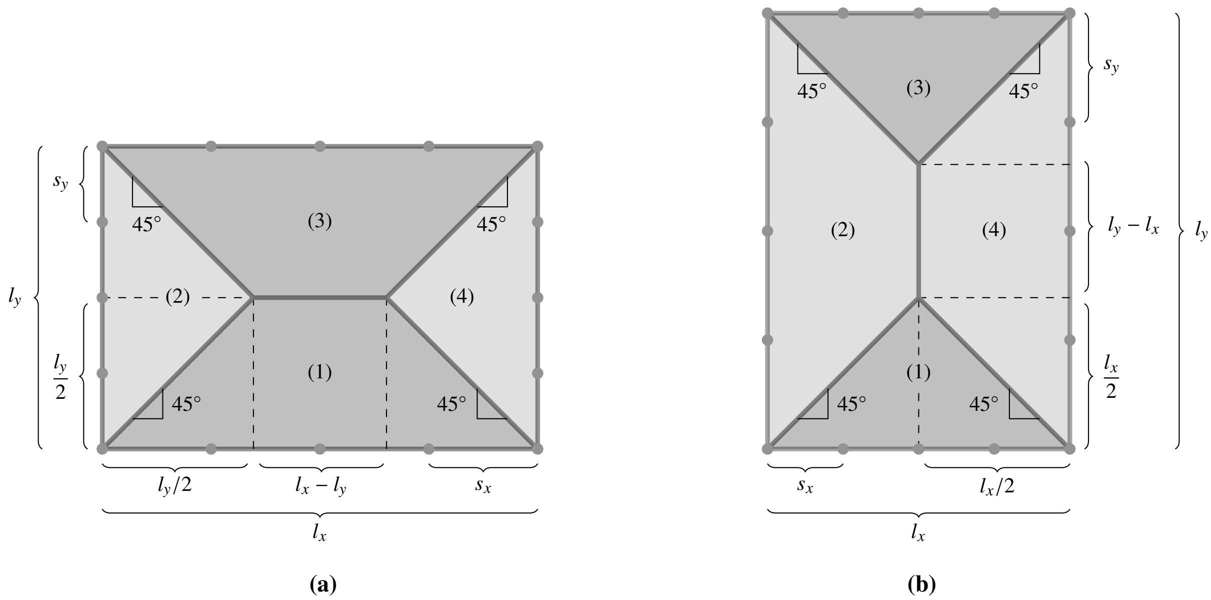

Figure 9: Regions defining potential locations of an origin/destination within a rectangle defined by sides l_x and l_y in case: (a) $l_x \geq l_y$; and (b) $l_x < l_y$.

Case 1. ($l_x \geq l_y$): regions (1) and (3) consist of two subregions: a rectangle with sides $l_x - l_y$ and $l_y/2$; and two right triangles with legs $l_y/2$ and $l_y/2$. The first subregion has the area of $P_{\square}^{(1)} = (l_x - l_y)l_y/2$ and the associated average walking distance of $(1/2)(l_y/2) = l_y/4$ along N-S and $(1/2)(s_x/2) = s_x/4$ along E-W directions, i.e. the average walking distance is $A_{\square}^{(1)} = (l_y/4 + s_x/4)/w$. The area of the second subregion (combining the two triangles) is $P_{\triangle}^{(1)} = l_y^2/4$, with the associated average walking distance of $(1/3)(l_y/2) = l_y/6$ along N-S and $(1/2)(s_x/2) = s_x/4$ along E-W directions, i.e. $A_{\triangle}^{(1)} = (l_y/6 + s_x/4)/w$. Similarly, regions (2) and (4) have each the area of $P_{\triangle}^{(2)} = l_y^2/4$, with the associated average walking distance of $(1/3)(l_y/2) = l_y/6$ along E-W and $(1/2)(s_y/2) = s_y/4$ along N-S directions, i.e. $A_{\triangle}^{(2)} = (l_y/6 + s_y/4)/w$. Then, assuming that the expected access and egress distances traveled are the same (similarly to Daganzo, 2010; Estrada et al., 2011), the expected walking time can be computed by considering all possible combinations of the subregions in which the origin/destination might be located:

$$A_{2,b}(\mathcal{D}) = 2\left[A_{\square}^{(1)} \cdot \frac{P_{\square}^{(1)}}{P_{\square}^{(1)} + P_{\triangle}^{(1)} + P_{\triangle}^{(2)}} + A_{\triangle}^{(1)} \cdot \frac{P_{\triangle}^{(1)}}{P_{\square}^{(1)} + P_{\triangle}^{(1)} + P_{\triangle}^{(2)}} + A_{\triangle}^{(2)} \cdot \frac{P_{\triangle}^{(2)}}{P_{\square}^{(1)} + P_{\triangle}^{(1)} + P_{\triangle}^{(2)}}\right]$$
$$= [6l_x l_y + 6l_x s_x - 2l_y^2 - 3l_y s_x + 3l_y s_y]/[12l_x w].$$

Case 2. $(l_x < l_y)$: regions (1) and (3) have the area of $P_{\triangle}^{(1)} = l_x^2/4$ and the associated average walking distance of $(1/3)(l_x/2) = l_x/6$ along N-S and $(1/2)(s_x/2) = s_x/4$ along E-W directions, i.e. $A_{\triangle}^{(1)} = (l_x/6 + s_x/4)/w$. On the other hand, regions (1) and (3) consist of two subregions (similarly to regions (1) and (3) in Case 1): a rectangle with sides $l_y - l_x$ and $l_x/2$; and two right triangles with legs $l_x/2$ and $l_x/2$. The first subregion has the area of $P_{\square}^{(2)} = (l_y - l_x)l_x/2$ and the associated average walking distance of $(1/2)(l_x/2) = l_y/4$ along E-W and $(1/2)(s_y/2) = s_y/4$ along N-S directions, i.e. $A_{\square}^{(2)} = (l_x/4 + s_y/4)/w$. The area of the second subregion (combining the two triangles) is $P_{\triangle}^{(2)} = l_x^2/4$, with the associated average walking distance of $(1/3)(l_x/2) = l_x/6$ along E-W and $(1/2)(s_y/2) = s_y/4$ along N-S directions, i.e. $A_{\triangle}^{(2)} = (l_x/6 + s_y/4)/w$. Similarly to the previous case, the expected walking time is given as:

$$A_{2,b}(\mathcal{D}) = 2\left[A_{\triangle}^{(1)} \cdot \frac{P_{\triangle}^{(1)}}{P_{\triangle}^{(1)} + P_{\square}^{(2)} + P_{\triangle}^{(2)}} + A_{\square}^{(2)} \cdot \frac{P_{\square}^{(2)}}{P_{\triangle}^{(1)} + P_{\square}^{(2)} + P_{\triangle}^{(2)}} + A_{\triangle}^{(2)} \cdot \frac{P_{\triangle}^{(2)}}{P_{\triangle}^{(1)} + P_{\square}^{(2)} + P_{\triangle}^{(2)}}\right]$$
$$= [6l_x l_y + 6l_y s_y - 2l_x^2 - 3l_x s_y + 3l_x s_x]/[12l_y w].$$

■

Result 2. *The expected number of transfers per trip for Type 2 user behavior is given by Eq. 29:*

$$\varphi_2(\mathcal{D}) = \begin{cases} \Pr(\varphi > 0)[4l_x^2 l_y^2 - 2l_x l_y^3 + l_y^4]/[2l_x^2 l_y^2], & \text{if } l_x \geq l_y, \\ \Pr(\varphi > 0)[4l_x^2 l_y^2 - 2l_x^3 l_y + l_x^4]/[2l_x^2 l_y^2], & \text{if } l_x < l_y. \end{cases}$$

Proof. As mentioned before, the expected number of transfers depends on the location of the user origin and the destination within a rectangle defined by sides l_x and l_y. From that perspective, if the origin is in region (1) or (3) and the destination is in region (1) or (3), but not along the same corridor as the origin, users need to make two transfers (from a horizontal bus line along which they access the system to a vertical one; and from a vertical to another horizontal bus line from which they reach the destination). Likewise, if the origin is in region (2) or (4) and the destination is in region (2) or (4), but not along the same corridor as the origin, users also need to make two transfers (from a vertical bus line along which they access the system to a horizontal one; and from a horizontal to another vertical bus line from which they reach the destination). If the origin is in region (1) or (3) and the destination is in region (2) or (4), or vice versa, users make one transfer. If both, the origin and the destination, are along the same corridor (a horizontal corridor for regions regions (1) and (3), and a vertical corridor for regions (2) and (4)), users do not make any transfer. This leads to the following formulations for the expected number of transfers for each rectangular shape shown in Fig. 9. Recall that the probability that both, the origin and the destination, are along the same corridor is equivalent to the probability of making zero transfer, $\Pr(\varphi = 0) = [l_y\phi_x + l_x\phi_y - l_x l_y]/[\phi_x\phi_y]$.

Case 1. $(l_x \geq l_y)$:

$$\varphi_2(\mathcal{D}) = 0 \cdot \Pr(\varphi = 0) + \Pr(\varphi > 0)[1 \cdot \Pr(\varphi = 1) + 2 \cdot \Pr(\varphi = 2)]$$
$$= \Pr(\varphi > 0)\left[1 \cdot \frac{2(P_{\square}^{(1)} + P_{\triangle}^{(1)})P_{\triangle}^{(2)}}{(P_{\square}^{(1)} + P_{\triangle}^{(1)} + P_{\triangle}^{(2)})^2} + 2 \cdot \frac{(P_{\square}^{(1)} + P_{\triangle}^{(1)})^2 + (P_{\triangle}^{(2)})^2}{(P_{\square}^{(1)} + P_{\triangle}^{(1)} + P_{\triangle}^{(2)})^2}\right]$$
$$= \Pr(\varphi > 0)[4l_x^2 l_y^2 - 2l_x l_y^3 + l_y^4]/[2l_x^2 l_y^2]$$

Case 2. $(l_x < l_y)$:

$$\begin{aligned}\varphi_2(\mathcal{D}) &= 0 \cdot \Pr(\varphi = 0) + \Pr(\varphi > 0)[1 \cdot \Pr(\varphi = 1) + 2 \cdot \Pr(\varphi = 2)] \\ &= \Pr(\varphi > 0)\left[1 \cdot \frac{2P_{\triangle}^{(1)}(P_{\square}^{(2)} + P_{\triangle}^{(2)})}{(P_{\square}^{(1)} + P_{\triangle}^{(1)} + P_{\triangle}^{(2)})^2} + 2 \cdot \frac{(P_{\triangle}^{(1)})^2 + (P_{\square}^{(2)} + P_{\triangle}^{(2)})^2}{(P_{\square}^{(1)} + P_{\triangle}^{(1)} + P_{\triangle}^{(2)})^2}\right] \\ &= \Pr(\varphi > 0)[4l_x^2 l_y^2 - 2l_x^3 l_y + l_x^4]/[2l_x^2 l_y^2]\end{aligned}$$

■

Result 3. *The expected waiting time per user including that at the origin and all transfer stops for Type 2 user behavior is given by Eq. 30:*

$$W_{2,b}(\mathcal{D}) = \begin{cases} \Pr(\varphi = 0)\bar{H}/2 + \Pr(\varphi > 0)[H_x(8l_x^2 l_y^2 - 4l_x l_y^3 + l_y^4) + H_y(4l_x^2 l_y^2 + l_y^4)]/[8l_x^2 l_y^2], & \text{if } l_x \geq l_y, \\ \Pr(\varphi = 0)\bar{H}/2 + \Pr(\varphi > 0)[H_x(4l_x^2 l_y^2 + l_x^4) + H_y(8l_x^2 l_y^2 - 4l_x^3 l_y + l_x^4)]/[8l_x^2 l_y^2], & \text{if } l_x < l_y. \end{cases}$$

Proof. Following the rationale for the expected number of transfers, we compute the expected waiting time as follows: (i) users with zero transfer wait on average half of the weighted bus headway in the network, i.e. $H(i) = \bar{H}/2 = [N_{b,x}H_x + N_{b,y}H_y]/[2(N_{b,x} + N_{b,y})]$; (ii) users with two transfers, the origin in region (1) or (3), and the destination in region (1) or (3) use two horizontal and one vertical bus line, thus wait on average $H(ii) = H_x + H_y/2$; (iii) users with two transfers, the origin in region (2) or (4), and the destination in region (2) or (4) use two vertical and one horizontal bus line, thus wait on average $H(iii) = H_y + H_x/2$; and (iv) users with one transfer, the origin in region (1) or (3), and the destination in region (2) or (4), or vice versa, use one horizontal and one vertical bus line, thus wait on average $H(iv) = H_x/2 + H_y/2$. As for the previous parameters of the user cost function for the bus mode, the expected waiting time is computed by multiplying these four groups of users by the corresponding probabilities (and in case of users with non-zero transfer by the probability of making a transfer $\Pr(\varphi > 0) = 1 - \Pr(\varphi = 0)$) for each rectangular shape shown in Fig. 9.

Case 1. $(l_x \geq l_y)$:

$$\begin{aligned}W_{2,b}(\mathcal{D}) &= H(i) \cdot \Pr(\varphi = 0) + \Pr(\varphi > 0)[H(ii) \cdot \Pr(ii) + H(iii) \cdot \Pr(iii) + H(iv) \cdot \Pr(iv)] \\ &= H(i) \cdot \Pr(\varphi = 0) + \Pr(\varphi > 0)\left[H(ii) \cdot \frac{(P_{\square}^{(1)} + P_{\triangle}^{(1)})^2}{(P_{\square}^{(1)} + P_{\triangle}^{(1)} + P_{\triangle}^{(2)})^2}\right. \\ &\quad \left. + H(iii) \cdot \frac{(P_{\triangle}^{(2)})^2}{(P_{\square}^{(1)} + P_{\triangle}^{(1)} + P_{\triangle}^{(2)})^2} + H(iv) \cdot \frac{2(P_{\square}^{(1)} + P_{\triangle}^{(1)})P_{\triangle}^{(2)}}{(P_{\square}^{(1)} + P_{\triangle}^{(1)} + P_{\triangle}^{(2)})^2}\right] \\ &= \Pr(\varphi = 0)\bar{H}/2 + \Pr(\varphi > 0)[H_x(8l_x^2 l_y^2 - 4l_x l_y^3 + l_y^4) + H_y(4l_x^2 l_y^2 + l_y^4)]/[8l_x^2 l_y^2]\end{aligned}$$

Case 2. $(l_x < l_y)$:

$$\begin{aligned}W_{2,b}(\mathcal{D}) &= H(i) \cdot \Pr(\varphi = 0) + \Pr(\varphi > 0)[H(ii) \cdot \Pr(ii) + H(iii) \cdot \Pr(iii) + H(iv) \cdot \Pr(iv)] \\ &= H(i) \cdot \Pr(\varphi = 0) + \Pr(\varphi > 0)\left[H(ii) \cdot \frac{(P_{\triangle}^{(1)})^2}{(P_{\square}^{(1)} + P_{\triangle}^{(1)} + P_{\triangle}^{(2)})^2}\right. \\ &\quad \left. + H(iii) \cdot \frac{(P_{\square}^{(2)} + P_{\triangle}^{(2)})^2}{(P_{\square}^{(1)} + P_{\triangle}^{(1)} + P_{\triangle}^{(2)})^2} + H(iv) \cdot \frac{2P_{\triangle}^{(1)}(P_{\square}^{(2)} + P_{\triangle}^{(2)})}{(P_{\square}^{(1)} + P_{\triangle}^{(1)} + P_{\triangle}^{(2)})^2}\right] \\ &= \Pr(\varphi = 0)\bar{H}/2 + \Pr(\varphi > 0)[H_x(4l_x^2 l_y^2 + l_x^4) + H_y(8l_x^2 l_y^2 - 4l_x^3 l_y + l_x^4)]/[8l_x^2 l_y^2]\end{aligned}$$

■

Result 4. *When the trip origins and destinations are uniformly and independently distributed across the network, Eq. 32 becomes equivalent to that of Daganzo (2010) for grid-like structures:*

$$O_p = \begin{cases} (\Lambda/4)H_x/N_{x,b}, & \text{if } p \in \{eb, wb\}, \\ (\Lambda/4)H_y/N_{y,b}, & \text{if } p \in \{nb, sb\}. \end{cases}$$

Proof. It suffices to show that the maximum number of on-board passengers along E-W/N-S directions is across the vertical/horizontal cordon that bisects the city, i.e. $\beta = 1/2$ (see Fig. 3). For brevity, we only show this for the *eb* direction. Given the symetry of the network, derivations for other cardinal directions are done in an entirely analogous manner using the corresponding subscripts.
To be able to obtain the same result as Daganzo (2010), we need to use the same assumption that the abscissa and the ordinate of both, the trip origins (x_1, y_1) and the destinations (x_2, y_2), are distributed uniformly and independently on $[0, \phi_x]$ and $[-\phi_y, \phi_y]$, respectively. As abscissae x_1 and x_2 are independent uniforms, their difference $x_1 - x_2$ has a triangular distribution, hence the user trip length in the *eb* direction $\{\ell_x = x_1 - x_2 : \ell_x \geq 0\}$ has the following probability density function: $\{f(\ell_x) = (\phi_x - \ell_x)/\phi_x^2 : \ell_x \in [0, \phi_x]\}$. Furthermore, due to the independence, the joint probability density function becomes $f(\ell_x, \ell_y) = f(\ell_x)f(\ell_y)$. Finally, since Daganzo (2010) does not consider the mode choice, we also assume that $\Pr(b \,|\, \ell_x, \ell_y) = 1, \ \forall \ell_x, \ell_y$. It follows that the maximum number of on-board passengers in the *eb* direction is:

$$\begin{aligned}
Q_{eb} &= \max_{\beta\in[0,1]} \left\{ \int_0^{\phi_x} \int_{-\phi_y}^{\phi_y} \Pr(J_\beta(\ell_x))\Lambda \Pr(b \,|\, \ell_x, \ell_y) f(\ell_x, \ell_y) \mathrm{d}\ell_y \mathrm{d}\ell_x \right\} \\
&= \max_{\beta\in[0,1]} \left\{ \frac{\Lambda}{\phi_x^2} \int_0^{\phi_x} \Pr(J_\beta(\ell_x))(\phi_x - \ell_x)\,\mathrm{d}\ell_x \right\} \\
&= \max_{\beta\in[0,1]} \left\{ \frac{\Lambda}{\phi_x^2} \Bigg[\int_0^{\min\{\beta\phi_x,(1-\beta)\phi_x\}} \frac{\ell_x}{\phi_x - \ell_x}(\phi_x - \ell_x)\,\mathrm{d}\ell_x \right. \\
&\quad + \int_{\min\{\beta\phi_x,(1-\beta)\phi_x\}}^{\max\{\beta\phi_x,(1-\beta)\phi_x\}} \frac{\min\{\beta\phi_x,(1-\beta)\phi_x\}}{\phi_x - \ell_x}(\phi_x - \ell_x)\,\mathrm{d}\ell_x \\
&\quad \left. + \int_{\max\{\beta\phi_x,(1-\beta)\phi_x\}}^{\phi_x} (\phi_x - \ell_x)\,\mathrm{d}\ell_x \Bigg] \right\} \\
&= \max_{\beta\in[0,1]} \left\{ \frac{\Lambda}{\phi_x^2} \Bigg[\frac{1}{2}(\max\{\beta\phi_x, (1-\beta)\phi_x\})^2 \right. \\
&\quad + \min\{\beta\phi_x, (1-\beta)\phi_x\}[\max\{\beta\phi_x, (1-\beta)\phi_x\} - \min\{\beta\phi_x, (1-\beta)\phi_x\}] \\
&\quad \left. + \phi_x[\phi_x - \max\{\beta\phi_x, (1-\beta)\phi_x\}] - \frac{1}{2}[\phi_x^2 - (\max\{\beta\phi_x, (1-\beta)\phi_x\})^2] \Bigg] \right\} \\
&= \max_{\beta\in[0,1]} \left\{ \frac{\Lambda}{\phi_x^2} \Bigg[\min\{\beta\phi_x, (1-\beta)\phi_x\} \max\{\beta\phi_x, (1-\beta)\phi_x\} \right. \\
&\quad - \frac{1}{2}(\min\{\beta\phi_x, (1-\beta)\phi_x\})^2 + \frac{1}{2}\phi_x^2 - \phi_x \max\{\beta\phi_x, (1-\beta)\phi_x\} \\
&\quad \left. + \frac{1}{2}(\max\{\beta\phi_x, (1-\beta)\phi_x\})^2 \Bigg] \right\}
\end{aligned}$$

To complete the proof, let us consider the following two cases:

Case 1. $\beta\phi_x \leq (1-\beta)\phi_x$:

$$\begin{aligned}
Q_{eb}(\beta) &= \frac{\Lambda}{\phi_x^2} \Big[\beta\phi_x(1-\beta)\phi_x - \frac{1}{2}\beta^2\phi_x^2 + \frac{1}{2}\phi_x^2 - \phi_x(1-\beta)\phi_x + \frac{1}{2}(1-\beta)^2\phi_x^2 \Big] \\
&= \Lambda(\beta - \beta^2)
\end{aligned}$$

Case 2. $\beta\phi_x > (1-\beta)\phi_x$:

$$\begin{aligned} Q_{eb}(\beta) &= \frac{\Lambda}{\phi_x^2}\left[(1-\beta)\phi_x\beta\phi_x - \frac{1}{2}(1-\beta)^2\phi_x^2 + \frac{1}{2}\phi_x^2 - \phi_x\beta\phi_x + \frac{1}{2}\beta^2\phi_x^2\right] \\ &= \Lambda(\beta - \beta^2) \end{aligned}$$

Notice that both cases yield the same result. Therefore, we can compute the value of β for which the maximum value of $Q_{eb}(\beta)$ is reached by solving the following equation:

$$\frac{\mathrm{d}Q_{eb}(\beta)}{\mathrm{d}\beta} = 0 \Leftrightarrow \Lambda(1-2\beta) = 0 \Leftrightarrow \beta = \frac{1}{2}$$

It then follows that the maximum number of on-board passengers in the *eb* direction is $Q_{eb} = Q_{eb}(\beta = 1/2) = \Lambda/4$, which is the same result as that of Daganzo (2010). ■

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: Logistics and Transportation Review 103, 261–285.

Ampountolas, K., Zheng, N., Geroliminis, N., 2017. Macroscopic modelling and robust control of bi-modal multi-region urban road networks. Transportation Research Part B: Methodological 104, 616–637.

Badia, H., Argote-Cabanero, J., Daganzo, C.F., 2017. How network structure can boost and shape the demand for bus transit. Transportation Research Part A: Policy and Practice 103, 83–94.

Badia, H., Estrada, M., Robusté, F., 2014. Competitive transit network design in cities with radial street patterns. Transportation Research Part B: Methodological 59, 161–181.

Badia, H., Estrada, M., Robusté, F., 2016. Bus network structure and mobility pattern: A monocentric analytical approach on a grid street layout. Transportation Research Part B: Methodological 93, 37–56.

Bhat, C.R., Guo, J., 2004. A mixed spatially correlated logit model: formulation and application to residential choice modeling. Transportation Research Part B: Methodological 38, 147–168.

Byrne, B.F., 1975. Public transportation line positions and headways for minimum user and system cost in a radial case. Transportation Research 9, 97–102.

Chen, H., Gu, W., Cassidy, M.J., Daganzo, C.F., 2015. Optimal transit service atop ring-radial and grid street networks: A continuum approximation design method and comparisons. Transportation Research Part B: Methodological 81, 755–774.

Daganzo, C., Menendez, M., 2005. A variational formulation of kinematic waves: Bottleneck properties and examples. Transportation and Traffic Theory. Flow, Dynamics and Human Interaction. 16th International Symposium on Transportation and Traffic Theory , 1–20.

Daganzo, C.F., 2005a. A variational formulation of kinematic waves: basic theory and complex boundary conditions. Transportation Research Part B: Methodological 39, 187–196.

Daganzo, C.F., 2005b. A variational formulation of kinematic waves: Solution methods. Transportation Research Part B: Methodological 39, 934–950.

Daganzo, C.F., 2010. Structure of competitive transit networks. Transportation Research Part B: Methodological 44, 434–446.

Daganzo, C.F., Geroliminis, N., 2008. An analytical approximation for the macroscopic fundamental diagram of urban traffic. Transportation Research Part B: Methodological 42, 771–781.

Dakic, I., Ambühl, L., Schümperlin, O., Menendez, M., 2019a. On the modeling of passenger mobility for stochastic bi-modal urban corridors. Transportation Research Part C: Emerging Technologies (in press). doi:10.1016/j.trc.2019.05.018.

Dakic, I., Menendez, M., 2018. On the use of Lagrangian observations from public transport and probe vehicles to estimate car space-mean speeds in bi-modal urban networks. Transportation Research Part C: Emerging Technologies 91, 317–334.

Dakic, I., Yang, K., Menendez, M., 2019b. Evaluating the effects of passenger occupancy dynamics on a bi-modal perimeter control, in: TRB Annual Meeting Online, Transportation Research Board.

Dakic, I., Yang, K., Menendez, M., Chow, J., 2020. Flexible bus dispatching system with modular and fully automated bus units, in: TRB Annual Meeting Online, Transportation Research Board.

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. Procedia - Social and Behavioral Sciences 17, 113–135.

Fan, W., Mei, Y., Gu, W., 2018. Optimal design of intersecting bimodal transit networks in a grid city. Transportation Research Part B: Methodological 111, 203–226.

Geroliminis, N., Zheng, N., Ampountolas, K., 2014. A three-dimensional macroscopic fundamental diagram for mixed bi-modal urban networks. Transportation Research Part C: Emerging Technologies 42, 168–181.

Haitao, H., Yang, K., Liang, H., Menendez, M., Guler, S.I., 2019. Providing public transport priority in the perimeter of urban networks: A bimodal strategy. Transportation Research Part C: Emerging Technologies 107, 171–192.

Holroyd, E., 1967. The optimum bus service: a theoretical model for a large uniform urban area, in: Proceedings of the Third International Symposium on the Theory of Traffic FlowOperations Research Society of America.

Leclercq, L., Geroliminis, N., 2013. Estimating MFDs in simple networks with route choice. Transportation Research Part B: Methodological 57, 468–484.

Levinson, D., 2012. Network structure and city size. PloS one 7, e29721.

Little, J.D., 1961. A proof for the queuing formula: L= λ w. Operations research 9, 383–387.

Loder, A., Ambühl, L., Menendez, M., Axhausen, K.W., 2017. Empirics of multi-modal traffic networks - Using the 3D macroscopic fundamental diagram. Transportation Research Part C: Emerging Technologies 82, 88–101.

Loder, A., Dakic, I., Bressan, L., Ambühl, L., Bliemer, M.C., Menendez, M., Axhausen, K.W., 2019. Capturing network properties with a functional form for the multi-modal macroscopic fundamental diagram. Transportation Research Part B: Methodological 129, 1–19.

Newell, G., 1979. Some issues relating to the optimal design of bus routes. Transportation Science 13, 20–35.

Nourbakhsh, S.M., Ouyang, Y., 2012. A structured flexible transit system for low demand areas. Transportation Research Part B: Methodological 46, 204–216.

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

Small, K., 2004. Urban Transportation. Concise Encyclopedia of Economics. 2nd ed.

Wirasinghe, S.C., Hurdle, V.F., Newell, G.F., 1977. Optimal parameters for a coordinated rail and bus transit system. Transportation Science 11, 359–374.

Zheng, N., Geroliminis, N., 2013. On the distribution of urban road space for multimodal congested networks. Transportation Research Part B: Methodological 57, 326–341.