On the design of an optimal flexible bus dispatching system with modular bus units: Using the three-dimensional Macroscopic Fundamental Diagram

Igor Dakic[a], Kaidi Yang[b,*], Monica Menendez[c], Joseph Y. J. Chow[d]

[a] *Traffic Engineering Group, Institute for Transport Planning and Systems, ETH Zurich, Switzerland*
[b] *Autonomous Systems Lab, Stanford University, USA*
[c] *Division of Engineering, New York University Abu Dhabi, United Arab Emirates*
[d] *C2SMART University Transportation Center, New York University, USA*

Abstract

This study proposes a novel *flexible bus dispatching system* where a fleet of fully automated modular bus units, together with conventional buses, serves the passenger demand. These modular bus units can either operate individually or combined (forming larger modular buses with a higher passenger capacity). This provides enormous flexibility to manage the service frequencies and vehicle allocation, reducing thereby the operating cost and improving passenger mobility.

We develop an optimization model used to determine the optimal composition of bus units and the optimal service frequency at which the buses (both conventional and modular) are dispatched across each bus line. We explicitly account for the dynamics of traffic congestion and complex interactions between the modes at the network level based on the recently proposed three-dimensional macroscopic fundamental diagram (3D-MFD). To the best of our knowledge, this is the first application of the 3D-MFD and modular bus units for the frequency setting problem in the domain of bus operations.

Numerical results show the improvements in the total system cost made by adapting the number of combined modular bus units and their dispatched frequencies to the evolution of both, the car and the passenger demand. A comparison with the commonly used approach that considers only the bus system (neglecting the complex multimodal interactions and congestion propagation) reveals the value of the proposed modeling framework.

Keywords: Optimal bus dispatching system; Modular bus units; Frequency setting problem; 3D-MFD

1. Introduction

In the past few decades, urbanization has become a prevalent trend for many places around the globe. As cities continue to grow, they require substantial infrastructure resources, placing various burdens to society and economic development. To address the social, economic, and environmental challenges associated with rapid urbanization, it is crucial to provide a sustainable transportation system. This is because transportation connects people and goods for social and economic interactions (Krugman, 1991).

Public transport is considered to be a crucial aspect of a sustainable urban development, as it allows more passengers to efficiently travel across an urban area (Roca-Riu et al., 2020). Nevertheless, to be attractive and a suitable alternative for the car owners, public modes of transport need to provide a good level of service. This can be achieved in many different ways, e.g. through a proper network design, frequency setting, timetable design, fleet assignment, and crew assignment. Due to their high complexity, the optimization of each of these stages is usually performed sequentially and is considered as a separate problem, influencing the

*Corresponding author. Tel.: +1 650 885 1186
Email addresses: i.dakic@ivt.baug.ethz.ch (Igor Dakic), kaidi.yang@stanford.edu (Kaidi Yang), monica.menendez@nyu.edu (Monica Menendez), joseph.chow@nyu.edu (Joseph Y. J. Chow)

decisions taken at the subsequent stage (Ceder, 2007; Desaulniers and Hickman, 2007). The decisions can be made for different planning horizons, depending on whether they are strategical, tactical, or operational (Martínez et al., 2014).

One of the prominent research topics in the public transport community, more specifically in the domain of bus operations, is the frequency setting problem. In this problem, the time intervals between consecutive buses need to be determined in a way that they can satisfy the total passenger demand generated during the planning horizon without a disproportionate increase in operator costs. As such, the problem belongs to the group of tactical problems, related to improving the level of service and reducing the operating cost of buses. Methods used to find the optimal bus frequency are usually based on minimizing the passenger waiting time, transfer time, total passenger cost, total operator cost, or a combination of them. Existing studies on this topic, up to now, rarely consider the impact of traffic conditions on the optimal bus frequency or the effects that the optimal bus frequency might have on traffic (Dakic et al., 2020).

To ensure that the determined bus frequency actually leads to an efficient bus service, it is crucial to account for the impact of traffic conditions, especially in case of mixed traffic (Loder et al., 2019a). With the emerging concepts related to the macroscopic modeling of large-scale networks, we can now study urban traffic dynamics in a parsimonious way. These concepts are mostly based on the Macroscopic Fundamental Diagram (MFD) (Daganzo, 2007; Geroliminis and Daganzo, 2008), also known as the Network Fundamental Diagram (Mahmassani et al., 2013). The MFD has been shown to be a useful and elegant tool to determine the current state of traffic in an urban transportation network. It links the vehicle accumulation and the travel production with a well-defined and reproducible curve. Although initially the MFDs were studied within the context of uni-modal (i.e. car) traffic, recent research efforts have resulted in the development of a three-dimensional MFD (3D-MFD), applicable for bi-modal traffic (Geroliminis et al., 2014; Loder et al., 2017, 2019b; Dakic et al., 2020). The 3D-MFD offers new ways to analyze complex interactions between different transportation modes, i.e. cars and buses (Dakic and Menendez, 2018), and has been mainly used to develop efficient perimeter control schemes for bi-modal urban networks (Ampountolas et al., 2017; Chiabaut et al., 2018; He et al., 2019; Dakic et al., 2019; Yang et al., 2019). The potential to apply the 3D-MFD for the purpose of capturing traffic conditions for the frequency setting problem has not been explored whatsoever.

When determining the optimal bus frequency, one of the most constraining parameters is the available vehicle fleet. The vehicle fleet can consist not only of conventional buses, but also of modular bus units that can either operate individually or combined together (forming thereby a single modular bus of a higher passenger capacity). In railway systems, for example, the vehicle fleet corresponds to a stock of cars and locomotives (commonly referred to as train units) that, when combined, form a single train. The number of train units assigned to a train is, in most cases, determined according to the predefined dispatched frequency of trains and the level of passenger demand (Cordeau et al., 2000, 2001; Lingaya et al., 2002; Fioole et al., 2006; Cacchiani et al., 2010). Only recently did researchers investigate the potential for combining the allocation of train units and the optimization of the dispatched frequency (Cadarso et al., 2013; Yue et al., 2017; Wang et al., 2018; Chen et al., 2019).

Recent advances in vehicle technology have opened new opportunities to apply similar concepts (i.e. combining and splitting of vehicle units) also for the bus systems (Guo et al., 2018; Chen et al., 2019, 2020). Such concept, herein called *flexible bus dispatching system*, can be especially useful for public transport operators, as it offers new perspectives and enormous flexibility to better manage the allocation of the vehicle resources and reduce the operating cost. For example, rather than dispatching a bus with a high passenger capacity in case of low passenger demand, an operator can send a group of few combined modular and fully automated bus units that, overall, has lower passenger capacity (increasing thereby the passenger occupancy) and lower operating cost. Furthermore, as the bus units are fully automated, there is no cost for assigning bus drivers to them.

Building on the knowledge of the 3D-MFD and the advanced automated vehicle technology, in this study we propose such a flexible bus dispatching system and provide a framework on how to model it. In particular, we develop an optimization model to determine the optimal number of combined automated bus units and the optimal frequency at which the units are dispatched across different bus lines, while accounting for the traffic dynamics at the network level. Our goal is to investigate whether such an optimization model can maintain (or even reduce) the operating cost, while increasing the level of service, and vice-versa. To the best of our knowledge, this is the first application of the 3D-MFD and modular bus units for the frequency setting problem in the domain of bus operations.

Overall, the contributions of this research are twofold. First, we propose a flexible bus dispatching system that allows to dynamically combine and split modular and fully automated bus units for the purpose of improving the operation of the bus system. Second, we apply the 3D-MFD concept for the frequency setting problem to capture the complex interactions between buses and cars in terms of traffic congestion, accounting for both vehicle and passenger dynamics.

The remainder of this paper is organized as follows. Section 2 provides an overview of the current state of the art. Section 3 describes in detail the proposed methodological framework used to determine the optimal number of combined automated bus units and the optimal frequency at which the units are dispatched. In Section 4, we quantify the improvements acquired by the proposed system in comparison to that with only conventional buses and show the value of the proposed modeling approach. Concluding remarks and future research directions are given in Section 5.

2. Literature review

While a comprehensive literature review on the frequency setting problem can be found in Ibarra-Rojas et al. (2015), for the reader's convenience, here we review some of the most relevant studies, along with the recent works that have been published since then.

Early studies on the optimal bus frequency were based on analytical models (Newell, 1971; Salzborn, 1972; Schéele, 1980; Han and Wilson, 1982) or heuristic methods (Furth and Wilson, 1981), and were usually formulated with the aid of graph theory, with nodes and arcs representing bus stops and street segments, respectively.

Vuchic et al. (1978) introduced the maximum load section-based method to determine the service frequency that can meet the passenger demand. Following a similar approach, Ceder and Wilson (1986) established four alternative methods based on the passenger count data. The objectives were twofold: to minimize the fleet size and to meet the passenger demand. The minimum fleet size was also used as the objective function by Salzborn (1972), who derived a continuous model for the optimal bus frequency in case of both, single-bus routes (Salzborn, 1972) and multiple-bus routes (Salzborn, 1980). Hadas and Shnaiderman (2012) and Li et al. (2013) studied the frequency setting problem with stochasticity in the demand, arrival times, boarding/alighting times, and travel times. Using a non-linear program to formulate the problem, Verbas and Mahmassani (2013) and Verbas et al. (2015) provided an optimal allocation of resources over space and time, while minimizing the weighted sum of ridership and wait time savings. Recent work by Gkiotsalitis and Cats (2018) determined the optimal bus frequency based on a travel time variability parameter.

Overall, the previous studies have rarely addressed the frequency setting problem under varying traffic conditions, especially in case of mixed traffic. To the best of our knowledge, the only attempt made in this direction can be found in Zhang and Liu (2019), who proposed a responsive bus dispatching strategy based on a doubly dynamical approach (as in Liu and Geroliminis, 2017). Although that study investigated a time-dependent bus dispatching problem in a bi-modal network, it only considered a simple network with two main directions, without analyzing possible effects of the dispatched frequency on the passenger dynamics across different bus stops, and how the available vehicle fleet should be optimally distributed across different bus lines during the planning horizon. Moreover, no research has been conducted regarding the concept of combining and splitting modular vehicle units along fixed bus lines. While some studies can be found in the scientific literature for railway systems (Cats and Haverkamp, 2018; Cordeau et al., 2000, 2001; Lingaya et al., 2002; Fioole et al., 2006; Cacchiani et al., 2010; Cadarso et al., 2013; Yue et al., 2017; Wang et al., 2018) and more recently for the metro system (Chen et al., 2019, 2020), strategies developed for railway systems are not readily applicable to bus systems. This is because the railway system represents a closed environment, where trains do not interact with other modes. For the bus system, however, we need to consider the complex interactions between buses and cars.

This study aims to fill these research gaps by: (i) proposing a novel, flexible bus dispatching system, in which the bus fleet consists of modular and fully automated bus units that can either be combined together or operate individually; and (ii) jointly optimizing the allocation of the modular bus units and their dispatched frequency across different bus lines. To this end, we develop a macroscopic framework to account for the traffic dynamics and the interactions between buses and cars, based on the 3D-MFD.

3. Methodological framework

3.1. System description

We consider here a bi-modal urban network consisting of buses and cars denoted as b and c, respectively. Without loss of generality, we assume that the congestion in the network is homogeneous, i.e. it exhibits a well-defined 3D-MFD, as otherwise the network can be partitioned into several homogeneous regions using well-established partitioning algorithms (see e.g. Saeedmanesh and Geroliminis, 2017; Ambühl et al., 2019).

For a given network, we index bus lines by $l \in \mathcal{L} = \{1, ..., |\mathcal{L}|\}$, where $|\mathcal{L}|$ is the total number of bus lines. In the proposed flexible bus dispatching system the bus fleet consists of two types of buses: (i) conventional buses denoted as r; and (ii) modular buses denoted as m. The bus type is indexed by $i \in \mathcal{I} = \{r, m\}$. Each modular bus can include $u \in \mathcal{U}_{m,l} = \{1, ..., |\mathcal{U}_{m,l}|\}$ modular bus units as shown in Fig. 1, where $|\mathcal{U}_{m,l}|$ is the maximum number of modular bus units that can be combined for line l. To simplify the notation, each conventional bus is considered as an individual conventional bus unit, i.e. $|\mathcal{U}_{r,l}| = 1, \ \forall l \in \mathcal{L}$ (see Fig. 1). In addition, all units of the same type i are considered to be equivalent in terms of operational characteristics and passenger capacity C_i. Finally, we assume that each conventional bus is equivalent to ζ number of modular bus units (in terms of capacity). The size of the bus fleet B can therefore be expressed either in terms of the equivalent number of conventional ($B = B_r + B_m/\zeta$) or modular ($B = B_r\zeta + B_m$) bus units. This gives the following formulation for the penetration rate of each type of bus units, p_i, defined as the share of the number of bus units of a given type i within the total number of equivalent bus units, i.e.:

$$p_i = \begin{cases} [B_r\zeta]/[B_r\zeta + B_m], & \text{if } i = r, \\ B_m/[B_r\zeta + B_m], & \text{if } i = m. \end{cases} \tag{1}$$

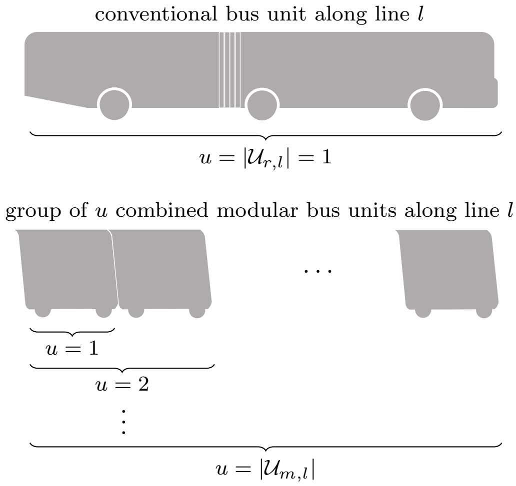

Figure 1: Schematic representation of the types of bus units that form the bus fleet in the proposed flexible bus dispatching system.

Before we formulate the objective function for determining the optimal number of combined modular bus units and the optimal frequency at which the units are dispatched, let us first introduce the modeling variables from both vehicular and passenger perspective. For the readers convenience, Table 1 provides the list of the most important notation used in this paper.

Table 1: Nomenclature.

$\mathcal{L}$	set of bus lines indexed by l
$\mathcal{S}_l$	set of segments along line l indexed by s
$\mathcal{I}$	set of types of bus units indexed by i: conventional unit ($i = r$); modular unit ($i = m$)
$\mathcal{T}$	set of time intervals indexed by t
τ	length of time interval
B_i	total number of available bus units of type i
C_i	passenger capacity of a bus unit of type i
π_i	operating cost of a bus unit of type i
η_i	bus unit-car equivalent for a bus unit of type i
$\lvert U_{i,l}\rvert$	maximum number of bus units of type i that can be combined along line l
Ω	average value of time
L_c	length of the car network
L_b	length of the bus network
ℓ_c	average vehicular trip length for the car mode
ℓ_b	average vehicular trip length for the bus mode
$\ell_{b,\text{pax}}$	average passenger trip length for the bus mode
$v_{c,\max}$	free-flow speed of cars
$v_{b,\text{mfd}}(t)$	average bus speed given by the 3D-MFD during interval t
$v_{b,\text{mix}}(t)$	average bus speed in the subnetwork with mixed segments during interval t
$v_{b,\text{ded}}$	average bus speed in the subnetwork with dedicated segments
$Q_c(t)$	circulating flow for the car mode given by the 3D-MFD during interval t
$Q_b(t)$	circulating flow for the bus mode given by the 3D-MFD during interval t
$G_c(t)$	trip completion flow for the car mode given by the 3D-MFD during interval t
$G_b(t)$	trip completion flow for the bus mode given by the 3D-MFD during interval t
$D_{c,\text{int}}(t)$	internal (originating from within the network) car demand during interval t
$D_{c,\text{ext}}(t)$	external (going into the network) car demand during interval t
$\overline{N}_c$	maximum car accumulation in the subnetwork with car only and mixed segments
$N_c(t)$	car accumulation in the entire network at the beginning of interval t
$N_b(t)$	bus accumulation in the entire network at the beginning of interval t
$N_{b,\text{mix}}(t)$	bus accumulation in the subnetwork with mixed segments at the beginning of interval t
$N_{b,\text{ded}}(t)$	bus accumulation in the subnetwork with dedicated segments at the beginning of interval t
$\ell_{b,l,s}$	length of bus segment s of line l
$v_{b,l,s}(t)$	bus speed on segment s of line l during interval t
$\varphi_{c,l,s}(t)$	receiving flow of cars to segment s of line l during interval t
$\psi_{b,i,l}(t)$	dispatched flow of bus units of type i along line l during interval t
$\Psi_{b,i,l}(t)$	dispatched flow of buses of type i along line l during interval t
$\phi_{b,i,l,s}(t)$	transferring flow of bus units of type i between segments s and $s+1$ of line l during interval t
$\Phi_{b,i,l,s}(t)$	transferring flow of buses of type i between segments s and $s+1$ of line l during interval t
$\overline{N}_{c,l,s}(t)$	maximum car accumulation that can be accommodated on segment s of line l at the beginning of time interval t
$N_{c,l,s}(t)$	accumulation of cars on segment s of line l at the beginning of time interval t
$N_{b,i,l,s}(t)$	accumulation of buses of type i on segment s of line l at the beginning of interval t
$n_{b,i,l,s}(t)$	accumulation of bus units of type i on segment s of line l at the beginning of interval t
$u_{i,l,s}(t)$	average number of combined units of type i on segment s of line l at the beginning of interval t
$P_{l,s,s'}(t)$	number of on-board passengers on segment s with destination on segment s' of line l at the beginning of interval t
$\Lambda_{l,s,s'}(t)$	number of boarding passengers on segment s with destination on segment s' of line l during interval t
$\Gamma_{l,s,s'}(t)$	outflow of passengers between segments s and $s+1$ with destination on segment s' of line l during interval t
$\omega_{l,s,s'}(t)$	number of unserved passengers on segment s with destination on segment s' of line l at the beginning of interval t
$\lambda_{l,s,s'}(t)$	average arrival rate of bus passengers on segment s with destination on segment s' of line l during interval t
$\delta_{l,s,s'}(t)$	share of passengers on segment s with destination on segment s' of line l during interval t

3.2. Modeling vehicular dynamics

For the purpose of modeling vehicular dynamics at the network level, we propose a framework based on the 3D-MFD. Let $N_b(t)$ and $N_c(t)$ be the total bus and car accumulations[1], respectively, at the beginning

[1] Note that in case of modular bus units the bus accumulation denotes the number of compositions of modular bus units in the network (e.g. a composition of $u = 3$ modular bus units is considered as a single modular bus with the passenger capacity of $3C_m$). In other words, we assume that buses affect the car traffic only due to their stops, neglecting the differences in the vehicle size (similarly to e.g. Zheng and Geroliminis (2013) and Dakic and Menendez (2018)).

of time interval $t \in \mathcal{T} = \{1, ..., |\mathcal{T}|\}$, where $|\mathcal{T}|$ is the total number of intervals included in the planning horizon. Considering that the bus system can operate in either dedicated or mixed lane fashion, we can further distinguish between the bus accumulation in the subnetwork with mixed ($N_{b,\text{mix}}(t)$) and dedicated ($N_{b,\text{ded}}(t)$) bus lanes, i.e.:

$$N_b(t) = N_{b,\text{mix}}(t) + N_{b,\text{ded}}(t). \tag{2}$$

To capture the effects of car traffic on bus operations and vice versa (in other words, the interactions between the two modes), we follow the approach by Geroliminis et al. (2014) and Loder et al. (2017), and assume a first order approximation, i.e. a linear model between the bus and the car speed. This leads to the following formulation of the average bus speed given by the 3D-MFD during time interval t:

$$v_{b,\text{mfd}}(t) = \alpha_{b,\text{mix}} N_{b,\text{mix}}(t) + \alpha_{b,\text{ded}} N_{b,\text{ded}}(t) + \alpha_c N_c(t) + \beta, \tag{3}$$

where α_b, α_c, and β are parameters that capture the marginal effect of each mode on the average bus speed, as well as the effects of the bus network topology. They can be estimated with real data (see Loder et al. (2017) for more details). We assume that these parameters are exogenously given.

The average bus speed given by the 3D-MFD can further be expressed as an average of the bus space-mean speeds in subnetworks with mixed ($v_{b,\text{mix}}(t)$) and dedicated ($v_{b,\text{ded}}(t)$) bus lanes, weighted by the total bus accumulations in the respective subnetworks:

$$v_{b,\text{mfd}}(t) = \frac{1}{N_b(t)} [v_{b,\text{mix}}(t) N_{b,\text{mix}}(t) + v_{b,\text{ded}} N_{b,\text{ded}}(t)]. \tag{4}$$

Notice from Eq. 4 that $v_{b,\text{ded}}$ is not a function of the time interval, i.e. $v_{b,\text{ded}}(t) = v_{b,\text{ded}}, \ \forall t \in \mathcal{T}$. This is because the traffic conditions do not affect bus operations along a dedicated bus segment (assuming that a dedicated bus lane operates with a transit signal priority). As a result, $v_{b,\text{ded}}$ can simply be modeled as the free-flow speed of buses (including the dwell time and acceleration/deceleration time while departing/approaching from/to a bus stop), i.e. $v_{b,\text{ded}} = \beta$ (see Eq. 3). Therefore, the bus space-mean speed in the subnetwork with mixed lanes can be computed by solving Eq. 4 for $v_{b,\text{mix}}(t)$, giving the following formulation for $v_{b,\text{mix}}(t)$:

$$\begin{aligned} v_{b,\text{mix}}(t) &= v_{b,\text{mfd}}(t) - \frac{N_{b,\text{ded}}(t)}{N_{b,\text{mix}}(t)} [v_{b,\text{ded}} - v_{b,\text{mfd}}(t)] \\ &\overset{(\text{Eq.}3)}{=} \beta + \frac{N_b(t)}{N_{b,\text{mix}}(t)} [\alpha_{b,\text{mix}} N_{b,\text{mix}}(t) + \alpha_{b,\text{ded}} N_{b,\text{ded}}(t) + \alpha_c N_c(t)]. \end{aligned} \tag{5}$$

Note that, for some of the bus lines, the lane allocation might not remain constant along the whole line. This, in turn, affects the operating regime, i.e. the speed of the bus system along a given line. To account for such effects, we split each bus line l into multiple segments, indexed by $s \in \mathcal{S}_l = \{1, ..., |\mathcal{S}_l|\}$, with similar length $\ell_{b,l,s}$, where $|\mathcal{S}_l|$ is the total number of segments along line l. These segments are grouped into two categories[2], i.e. per lane allocation type: $\mathcal{S}_l = \mathcal{S}_{l,\text{mix}} \cup \mathcal{S}_{l,\text{ded}}$. Following the rationale of the cell transmission model (CTM) (Daganzo, 1994), we set the length of the smallest segment across all bus lines, $\min_{l \in \mathcal{L}, s \in \mathcal{S}_l} \ell_{b,l,s}$, to be such that it satisfies the following condition:

$$\tau v_{c,\max} \leq \min_{l \in \mathcal{L}, s \in \mathcal{S}_l} \ell_{b,l,s}, \tag{6}$$

where τ and $v_{c,\max}$ stand for the duration of the time interval and the free-flow speed of cars, respectively. Now, for each segment s of a given bus line l, we can define the operating bus space-mean speed using Eq. 7. Recall that we assume a homogeneous network, thus $v_{b,\text{mix}}(t)$ is the same across the network for any given interval t.

$$v_{b,\ell,s}(t) = \begin{cases} v_{b,\text{mix}}(t), & \text{if } s \in \mathcal{S}_{l,\text{mix}}, \\ v_{b,\text{ded}}, & \text{if } s \in \mathcal{S}_{l,\text{ded}}. \end{cases} \tag{7}$$

[2]For a given segment we assign the lane allocation that is more prevalent along its length.

We can also relate the total bus accumulation in each subnetwork (with mixed and dedicated bus segments) to the bus accumulation across all segments of the same type and across both types of buses (Eq. 8).

$$N_{b,\text{mix}}(t) = \sum_{i\in\mathcal{I}}\sum_{l\in\mathcal{L}}\sum_{s\in\mathcal{S}_{l,\text{mix}}} N_{b,i,l,s}(t), \tag{8a}$$

$$N_{b,\text{ded}}(t) = \sum_{i\in\mathcal{I}}\sum_{l\in\mathcal{L}}\sum_{s\in\mathcal{S}_{l,\text{ded}}} N_{b,i,l,s}(t). \tag{8b}$$

Now that we have introduced all the notation and defined all modeling variables, we can formulate the mass conservation equations used to model the evolution of the vehicle accumulation for each mode. Given the assumption of the homogeneous traffic conditions across the network, which assumes that the average car speed in mixed lanes is similar to the average car speed in car only lanes (as in e.g. Zheng and Geroliminis (2013), Dakic and Menendez (2018), and Zhang and Liu (2019)), the system dynamics for the car mode can be modeled at the network level (Eq. 9).

$$N_c(t+1) = N_c(t) + D_{c,\text{int}}(t)\tau + D_{c,\text{ext}}(t)\tau - G_c(N_b(t), N_c(t))\tau, \tag{9}$$

where $D_{c,\text{int}}(t)$ and $D_{c,\text{ext}}(t)$ stand for the internal (originating from within the network) and external (incoming to the network) car demand, respectively, during time interval t; and $G_c(N_b(t), N_c(t))$ is the trip completion flow for the car mode given by the 3D-MFD during time interval t.

On the other hand, due to different operational regimes of buses across different lane allocations along a bus line, as stated before, the system dynamics for the bus mode should be modeled on the segment level (Eq. 10). Notice from Eq. 10 that in addition to tracking the number of buses of each type, $N_{b,i,l,s}(t)$, we also track the number of bus units of a given type i, $n_{b,i,l,s}(t)$, for any time interval t. This way, we can obtain an average number of modular bus units contained in a modular bus on segment s of line l during time interval t as $u_{m,l,s}(t) = n_{b,m,l,s}(t)/N_{b,m,l,s}(t)$. Recall that for conventional bus units $u_{r,l,s}(t) = 1$, i.e. $n_{b,r,l,s}(t) = N_{b,r,l,s}(t)$. Fig. 2 illustrates the proposed framework for modeling the vehicular dynamics for the bi-modal system.

$$N_{b,i,l,s}(t+1) = \begin{cases} N_{b,i,l,s}(t) + \Psi_{b,i,l}(t)\tau - \Phi_{b,i,l,s}(t)\tau, & \text{if } s = 1, \\ N_{b,i,l,s}(t) + \Phi_{b,i,l,s-1}(t)\tau - \Phi_{b,i,l,s}(t)\tau, & \text{if } 1 < s < |\mathcal{S}_l|, \\ N_{b,i,l,s}(t) + \Phi_{b,i,l,s-1}(t)\tau - [N_{b,i,l,s}(t)/N_b(t)]G_b(N_b(t), N_c(t))\tau, & \text{if } s = |\mathcal{S}_l|, \end{cases} \tag{10a}$$

$$n_{b,i,l,s}(t+1) = \begin{cases} n_{b,i,l,s}(t) + \psi_{b,i,l}(t)\tau - \phi_{b,i,l,s}(t)\tau, & \text{if } s = 1, \\ n_{b,i,l,s}(t) + \phi_{b,i,l,s-1}(t)\tau - \phi_{b,i,l,s}(t)\tau, & \text{if } 1 < s < |\mathcal{S}_l|, \\ n_{b,i,l,s}(t) + \phi_{b,i,l,s-1}(t)\tau - [n_{b,i,l,s}(t)/N_b(t)]G_b(N_b(t), N_c(t))\tau, & \text{if } s = |\mathcal{S}_l|, \end{cases} \tag{10b}$$

where $\Psi_{b,i,l}(t)$ is the dispatched flow of buses and $\psi_{b,i,l}(t)$ is the dispatched flow of bus units of type i along line l during time interval t; $\Phi_{b,i,l,s}(t)$ and $\phi_{b,i,l,s}(t)$ denote the transferring flow of buses and the transferring flow of bus units, respectively, of type i between segments s and $s+1$ of line l during time interval t; and $G_b(N_b(t), N_c(t))$ stands for the trip completion flow for the bus mode given by the 3D-MFD during time interval t.

The trip completion flow for the bus and the car mode includes both the internal outflow (reaching destination inside the network) and the external outflow (exiting the network). Geroliminis and Daganzo (2008) showed that $G_b(N_b(t), N_c(t))$ and $G_c(N_b(t), N_c(t))$ are proportional to the total circulating flows, i.e. $Q_b(N_b(t), N_c(t))$ and $Q_c(N_b(t), N_c(t))$, respectively, with a factor that represents the ratio of the network length for a given mode (L_c for cars; L_b for buses) and an average vehicular trip length ($\bar{\ell}_c$ for cars; $\bar{\ell}_b$ for buses):

$$G_c(N_b(t), N_c(t)) = Q_c(N_b(t), N_c(t))\frac{L_c}{\bar{\ell}_c}, \tag{11a}$$

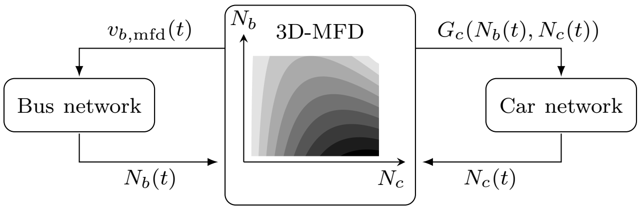

Figure 2: Modeling diagram for the proposed 3D-MFD based framework.

$$G_b(N_b(t), N_c(t)) = Q_b(N_b(t), N_c(t))\frac{L_b}{\bar{\ell}_b}. \tag{11b}$$

Eq. 12 calculates the transferring flow of buses/bus units following the idea of the CTM (Daganzo, 1994), as the minimum of the sending flow from segment s of line l (the first term) and the receiving flow to segment $s+1$ of the same line during a given interval t (the second term).

$$\Phi_{b,i,l,s}(t) = \min\left\{v_{b,l,s}(t)\frac{N_{b,i,l,s}(t)}{\ell_{b,l,s}}, \frac{\varphi_{c,l,s+1}(t)}{\eta_i u_{i,l,s}(t)}\right\}, \tag{12a}$$

$$\phi_{b,i,l,s}(t) = \min\left\{v_{b,l,s}(t)\frac{n_{b,i,l,s}(t)}{\ell_{b,l,s}}, \frac{\varphi_{c,l,s+1}(t)}{\eta_i}\right\}, \tag{12b}$$

where η_i represents a parameter that quantifies by how much one bus unit of type i reduces the car flow (i.e. a bus unit-car equivalent); and $\varphi_{c,l,s}(t)$ denotes the receiving flow of cars to segment s of line l during time interval t. Before we formulate $\varphi_{c,l,s}(t)$, let us first introduce some notation.

Let $\overline{N}_c$ be the maximum car accumulation in the subnetwork where cars are allowed to drive (including car only and mixed segments, i.e. L_c). Recall that in this study we assume homogeneous traffic conditions, i.e. the average car speed in mixed segments is similar to the average car speed in car only segments. This, in turn, implies that we can determine the total car accumulation in the subnetwork with mixed segments at the beginning of time interval t, $N_{c,\text{mix}}(t)$, by multiplying the total car accumulation in the network, $N_c(t)$, by a fraction that represents the ratio of the maximum car accumulation that can be accommodated in the subnetwork with mixed segments (the summation term in Eq. 13) and the maximum car accumulation in the network, $\overline{N}_c$:

$$N_{c,\text{mix}}(t) = \frac{N_c(t)}{\overline{N}_c}\sum_{l\in\mathcal{L}}\sum_{s\in\mathcal{S}_{l,\text{mix}}}\overline{N}_{c,l,s}(t), \tag{13}$$

where $\overline{N}_{c,l,s}(t)$ stands for the maximum car accumulation that can be accommodated on a given segment s of line l at the beginning of time interval t, as a function of the bus accumulation on that same segment (Eq. 14).

$$\overline{N}_{c,l,s}(t) = \overline{N}_c\frac{\ell_{b,l,s}}{L_c} - \sum_{i\in\mathcal{I}} n_{b,i,l,s}(t)\eta_i. \tag{14}$$

Using $\overline{N}_{c,l,s}(t)$, we can further distribute $N_{c,\text{mix}}(t)$ across the mixed segments according to the proportion of the remaining capacity, i.e:

$$N_{c,l,s}(t) = \frac{N_{c,\text{mix}}(t)\overline{N}_{c,l,s}(t)}{\sum_{l\in\mathcal{L}}\sum_{s\in\mathcal{S}_{l,\text{mix}}}\overline{N}_{c,l,s}(t)} = N_c(t)\frac{\overline{N}_{c,l,s}(t)}{\overline{N}_c}. \tag{15}$$

Similarly to the CTM, we can now compute the receiving flow of cars based on the maximum available car accumulation, $\overline{N}_{c,l,s}(t)$, and (only in case of a mixed segment) the number of cars present on a given segment at the beginning of time interval t, as given by Eq. 16, where w is the backward wave speed.

$$\varphi_{c,l,s}(t) = \begin{cases} w[\overline{N}_{c,l,s}(t) - N_{c,l,s}(t)]/\ell_{b,l,s}, & \text{if } s \in \mathcal{S}_{l,\text{mix}}, \\ w\overline{N}_{c,l,s}(t)/\ell_{b,l,s}, & \text{if } s \in \mathcal{S}_{l,\text{ded}}. \end{cases} \tag{16}$$

3.3. Modeling passenger dynamics

In this section, we extend the modeling formulations to account for passenger occupancy dynamics. Taking into account that for the car mode, the number of passengers served can essentially be approximated with the number of served vehicles, i.e. the average car occupancy is $\rho_c \approx 1$, in the following we only describe the evolution of the total bus passengers over time:

$$P_{l,s,s'}(t+1) = \begin{cases} P_{l,s,s'}(t) + \Lambda_{\ell,s,s'}(t) - \Gamma_{\ell,s,s'}(t), & \text{if } s = 1, \\ P_{l,s,s'}(t) + \Lambda_{\ell,s,s'}(t) + \Gamma_{\ell,s-1,s'}(t) - \Gamma_{\ell,s,s'}(t), & \text{if } s > 1, \end{cases} \tag{17}$$

where $P_{l,s,s'}(t)$ is the total number of on-board passengers on segment s with destination on segment s' of line l at the beginning of time interval t; $\Lambda_{l,s,s'}(t)$ represents the total number of boarding passengers on segment s with destination on segment s' of line l during time interval t; and $\Gamma_{l,s,s'}(t)$ denotes the outflow (including the transferring and the trip completion flow, as elaborated further below) of bus passengers between segments s and $s+1$ of line l during time interval t, whose destination is on segment s' of the same line (Eq. 18).

$$\Gamma_{\ell,s,s'}(t) = \begin{cases} \min\left\{P_{l,s,s'}(t), \dfrac{\sum_{i\in\mathcal{I}} P_{l,s,s'}(t)C_i\phi_{b,i,l,s}(t)\tau}{\sum_{i\in\mathcal{I}} n_{b,i,l,s}(t)C_i}\right\}, & \text{if } s \neq s' \wedge s < |\mathcal{S}_l| \\ \min\left\{P_{l,s,s'}(t), \dfrac{P_{l,s,s'}(t)G_b(N_b(t), N_c(t))\tau}{\sum_{i\in\mathcal{I}} N_{b,i,l,s}(t)} \dfrac{\bar{\ell}_b}{\bar{\ell}_{b,\text{pax}}}\right\}, & \text{if } s = s' \vee s = |\mathcal{S}_l| \end{cases} \tag{18}$$

where $\bar{\ell}_{b,\text{pax}}$ stands for the average trip length of bus passengers. Note that in case where passengers do not arrive at their destination (i.e. $s \neq s'$ and $s < |\mathcal{S}_l|$), the outflow $\Gamma_{\ell,s,s'}(t)$ represents the transferring flow from segment s to segment $s+1$ of a given line l during time interval t. Otherwise (i.e. $s = s'$ or $s = |\mathcal{S}_l|$), the outflow represents the trip completion flow of bus passengers given by the 3D-MFD. Further note that the terms $P_{l,s,s'}(t)C_i/\sum_{i\in\mathcal{I}} n_{b,i,l,u,s}(t)C_i$ and $P_{l,s,s'}(t)/\sum_{i\in\mathcal{I}} N_{b,i,l,u,s}(t)$ denote the average occupancy of a bus unit of type i and the average bus occupancy (across both types of buses), respectively, of passengers on segment s of line l during time interval t, whose destination is on segment s' of the same line.

The total number of boarding passengers is bounded by the two parameters in Eq. 19: the total number of (accumulated) passengers that want to enter the bus, $\Lambda_{\text{acc},l,s,s'}$, and the total number of passengers that can enter the bus, $\Lambda_{\max,l,s,s'}$:

$$\Lambda_{l,s,s'}(t) = \min\{\Lambda_{\text{acc},l,s,s'}(t), \Lambda_{\max,l,s,s'}(t)\}. \tag{19}$$

Before we derive these two parameters, let us first introduce the evolution of the total number of passengers with destination on segment s' of line l who cannot board the bus on segment s of the same line by the beginning of time interval t:

$$\omega_{l,s,s'}(t+1) = \omega_{l,s,s'}(t) + \lambda_{l,s,s'}(t)\tau - \Lambda_{l,s,s'}(t), \tag{20}$$

where $\lambda_{l,s,s'}(t)$ denotes the average arrival rate of bus passengers on segment s with destination on segment s' of line l during time interval t.

Then, the total number of passengers that want to enter the bus unit on segment s with destination on segment s' of line l during time interval t can be obtained as follows:

$$\Lambda_{\text{acc},l,s,s'}(t) = \omega_{l,s,s'}(t) + \lambda_{l,s,s'}(t)\tau. \tag{21}$$

On the other hand, the total number of passengers with destination on segment s' of line l that can enter the bus on segment s of the same line is determined by the total available capacity across both types of bus units operating along line l during a given time interval:

$$\Lambda_{\max,l,s,s'}(t) = \begin{cases} \delta_{l,s,s'}(t)\Big[\sum_{i\in\mathcal{I}} n_{b,i,l,s}(t+1)C_i - \sum_{s'\in\mathcal{S}_l}[P_{l,s,s'}(t) - \Gamma_{l,s,s'}(t)]\Big], & \text{if } s=1, \\ \delta_{l,s,s'}(t)\Big[\sum_{i\in\mathcal{I}} n_{b,i,l,s}(t+1)C_i - \sum_{s'\in\mathcal{S}_l}[P_{l,s,s'}(t) + \Gamma_{l,s-1,s'}(t) - \Gamma_{l,s,s'}(t)]\Big], & \text{if } s>1, \end{cases} \tag{22}$$

where $n_{b,i,l,s}(t+1)$ is calculated using Eq. 10, and $\delta_{l,s,s'}(t)$ represents the share of passengers on segment s of line l during time interval t whose destination is on segment s' of the same line (Eq. 23).

$$\delta_{l,s,s'}(t) = \frac{\lambda_{l,s,s'}(t)\tau + \omega_{l,s,s'}(t)}{\sum_{s'\in\mathcal{S}_l}[\lambda_{l,s,s'}(t)\tau + \omega_{l,s,s'}(t)]}. \tag{23}$$

3.4. Mathematical formulation of the objective function

Now that we have derived all the parameters for modeling the vehicular and passenger dynamics for both modes, we can formulate the objective function used to determine the optimal number of combined modular bus units and the optimal frequency at which the units are dispatched. Recall that the decision variables include the dispatched flow of buses $\boldsymbol{\Psi} = [\Psi_{b,i,l}(t) : l\in\mathcal{L}, i\in\mathcal{I}, t\in\mathcal{T}]$ and the dispatched flow of bus units $\boldsymbol{\psi} = [\psi_{b,i,l}(t) : l\in\mathcal{L}, i\in\mathcal{I}, t\in\mathcal{T}]$ of both types across all bus lines in the network, for each time interval. The objective function is given by Eq. 24 and, similarly to the previous studies, it consists of two components: operator cost $Z_O(\boldsymbol{\Psi},\boldsymbol{\psi})$; and user cost $Z_U(\boldsymbol{\Psi},\boldsymbol{\psi})$.

$$\min\ Z(\boldsymbol{\Psi},\boldsymbol{\psi}) = Z_O(\boldsymbol{\Psi},\boldsymbol{\psi}) + Z_U(\boldsymbol{\Psi},\boldsymbol{\psi}), \tag{24}$$

s.t. Eqs. 9–23,

$$\Psi_{b,i,l}(t),\ \psi_{b,i,l}(t) \geq 0, \quad \forall l\in\mathcal{L},\ t\in\mathcal{T},\ i\in\mathcal{I}, \tag{25}$$

$$\Psi_{b,i,l}(t) - \psi_{b,i,l}(t) \leq 0, \quad \forall l\in\mathcal{L},\ t\in\mathcal{T},\ i\in\mathcal{I}, \tag{26}$$

$$-|\mathcal{U}_{i,l}|\Psi_{b,i,l}(t) + \psi_{b,i,l}(t) \leq 0, \quad \forall l\in\mathcal{L},\ t\in\mathcal{T},\ i\in\mathcal{I}, \tag{27}$$

$$\sum_{l\in\mathcal{L}}\psi_{b,i,l}(t)\tau + \sum_{l\in\mathcal{L}}\sum_{s\in\mathcal{S}_l} n_{b,i,l,s}(t) - \sum_{l\in\mathcal{L}}\frac{n_{b,i,l,|\mathcal{S}_l|}(t)}{N_b(t)}G_b(N_b(t),N_c(t))\tau \leq B_i, \quad \forall t\in\mathcal{T},\ i\in\mathcal{I}. \tag{28}$$

In terms of the constraints, note that all decision variables need to be positive for physical reasons (Eq. 25). In particular, the number of dispatched buses along any bus line cannot exceed the number of dispatched bus units (Eq. 26), nor the number of combined units can exceed the maximum value for a given line (Eq. 27). Likewise, the total number of bus units required for the operation during the planning horizon cannot exceed the vehicle fleet in case of both conventional and modular bus units (Eq. 28).

The operator cost metrics in Eq. 24 include the cost of assigning bus drivers to conventional bus units, as well as the costs incurred by the energy consumption and maintenance of both conventional and modular bus units (Eq. 29). These costs are computed based on the number of bus units of each type operating during the planning horizon and the corresponding unit costs π_i.

$$Z_O(\boldsymbol{\Psi},\boldsymbol{\psi}) = \sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{I}}\sum_{l\in\mathcal{L}}\sum_{s\in\mathcal{S}_\ell} n_{b,i,l,s}(t)\pi_i\tau. \tag{29}$$

On the other hand, the user cost metrics in Eq. 24 are related to the total time traveled in the system (including the waiting time), computed across both modes and converted into an equivalent monetary cost with a parameter Ω denoting the average value of time (Eq. 30).

$$Z_U(\boldsymbol{\Psi},\boldsymbol{\psi}) = \underbrace{\Omega\sum_{t\in\mathcal{T}}\sum_{l\in\mathcal{L}}\sum_{s\in\mathcal{S}_l}\sum_{s'\in\mathcal{S}_l}[P_{l,s,s'}(t)+\omega_{l,s,s'}(t)]\tau}_{\textit{term 1}} + \underbrace{\Omega\sum_{t\in\mathcal{T}}\sum_{l\in\mathcal{L}}\sum_{s\in\mathcal{S}_l}\sum_{s'\in\mathcal{S}_l}\frac{1}{2}\frac{\lambda_{l,s,s'}(t)\tau|\mathcal{I}|}{\sum_{i\in\mathcal{I}}\Phi_{b,i,l,s}(t)}}_{\textit{term 2}} + \underbrace{\Omega\sum_{t\in\mathcal{T}}\rho_c N_c(t)\tau}_{\textit{term 3}}, \tag{30}$$

where *term 1* calculates the time traveled by bus users and the waiting time due to the carry over of passengers; *term 2* represents a proxy for the passenger waiting time due to the bus headway (defined as half of the average bus headway across all types of bus units, $[\sum_{i\in\mathcal{I}} \Phi_{b,i,l,s}(t)/|\mathcal{I}|]^{-1}$, on a given segment); and *term 3* denotes the time traveled by car users.

4. Numerical experiments and results

4.1. Case study and simulation scenarios

Here we describe the simulation environment for testing the performance of the proposed flexible bus dispatching system. The considered traffic network is inspired by the City of Zurich, Switzerland, and is comprised of five bus lines with varying lane allocation. Dedicated bus segments are placed closer to the center of the network, whereas the mixed lane segments are installed closer to the periphery (Fig. 3). The simulated traffic conditions reflect a typical morning-peak period. The 3D-MFD and the demand profiles (both in terms of cars and passengers) during the 3 hour period (Fig. 4) are designed to mimic the aggregated traffic features of the city center of Zurich (Ambühl et al., 2017), as this one exhibits a well-defined empirical 3D-MFD (see Loder et al. (2017) for more details). To obtain realistic traffic conditions, we impose lower and upper bounds to the conservation equations (both in terms of vehicles and passengers).

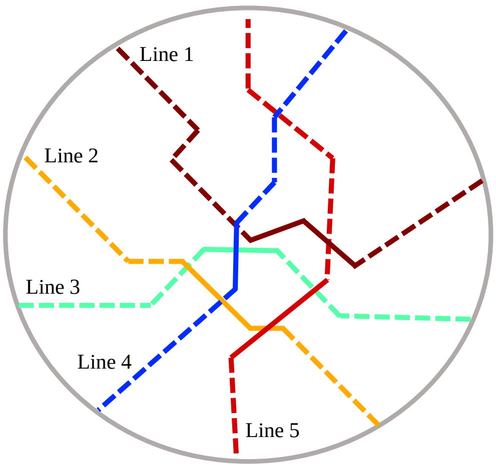

Figure 3: Schematic illustration of the studied network, where solid lines represent dedicated bus segments, whereas dashed lines represent mixed bus segments.

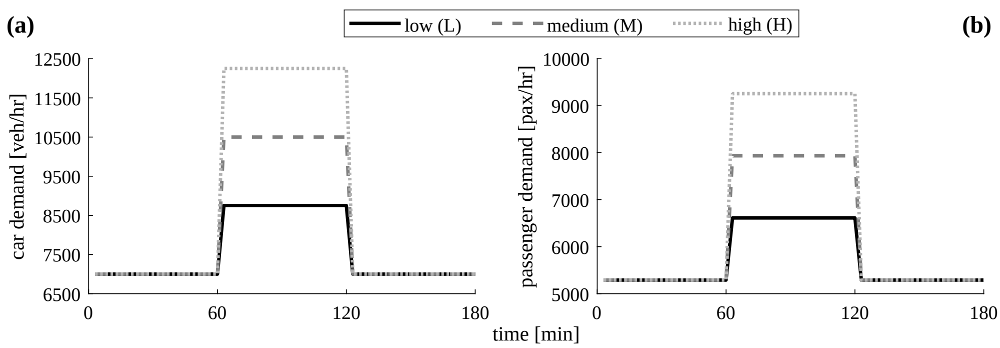

Figure 4: Simulation settings: (a) car demand profile; and (b) public transport passenger demand profile.

Tested traffic scenarios include three types of car demand (low, medium, and high) and three types of public transport passenger demand (low, medium, and high) (see Fig. 4). For each scenario, we vary the penetration rate of modular bus units: $p_m \in \{0\%, 10\%, 20\%, 30\%, 40\%\}$. The penetration rate of $p_m = 0\%$ corresponds to the base case, in which no modular units are incorporated into the bus fleet. In such a base case, the considered bus fleet includes $B_r = 30$ conventional bus units. In other cases with $p_m > 0$, the number of conventional buses depends on the penetration rate of modular bus units, i.e. $B_r = 30(1 - p_m)$. We assume that the bus capacity is $C_r = 120$ pax/veh for conventional buses and $C_m = 20$ pax/veh for modular bus units, which gives $\zeta = 6$. The passenger cost per unit time is assumed to be 20 CHF/hr, comparable to the value of time in the city of Zurich (Axhausen et al., 2006). Finally, we approximate the unit cost π_i for the bus operator by accounting for the driver cost (only in case of conventional bus units), vehicle cost, operational cost, overhead (e.g. administration and planning cost), ticketing cost, etc., as suggested in Sinner (2019), leading to $\pi_r \approx 260$ CHF/hr and $\pi_m \approx 30$ CHF/hr.

The proposed optimization problem is solved using a conventional sequential quadratic programming algorithm (Boggs and Tolle, 1995), which allows to formulate the constraints explicitly, without using the penalty terms in the objective function. To find the optimal solution, we set the number of initial points to 50. Sensitivity analyses show that increasing the number of initial points beyond 50 leads to marginal improvements in the objective function (we omit the results here for brevity) for all scenarios with different demand levels and penetration rates of modular units. In the future, it might be possible to use more in depth sensitivity analysis methods (Ge et al., 2015; Ge and Menendez, 2017) to further improve the efficiency of the solution algorithm. Experiments are run on a 16-core Intel Xeon processor (3.19 GHz) with 256 GB RAM. The computation time for a given scenario and penetration rate of modular units is around 8.5 min.

4.2. Performance of the proposed flexible bus dispatching system

Results are shown in the form of bar plots (Fig. 5) comparing the absolute (in CHF/hr) and relative (in %) improvement in the total system cost (including the operator cost and the system delay) of the proposed flexible bus dispatching system with different penetration rates of modular units to that of the base case. Note that even if the performance is the same, the cost for the base case described here is still not equivalent to the conventional bus cost. This is because we are only dealing with the frequency setting and not vehicle scheduling. The underlying vehicle scheduling problem (deadheading/assigning vehicles to trips or to reposition to merge with other vehicles) for modular bus units has a substantially different cost structure than that for conventional bus (likely much lower cost). Therefore, there are additional cost savings that are not covered here from that perspective. In Fig. 5, traffic scenarios are represented as XY, where X indicates the level of car demand and Y indicates the level of passenger demand for public transport: low (L), medium (M), or high (H). For example, LM stands for the scenario with low car and medium public transport passenger demand.

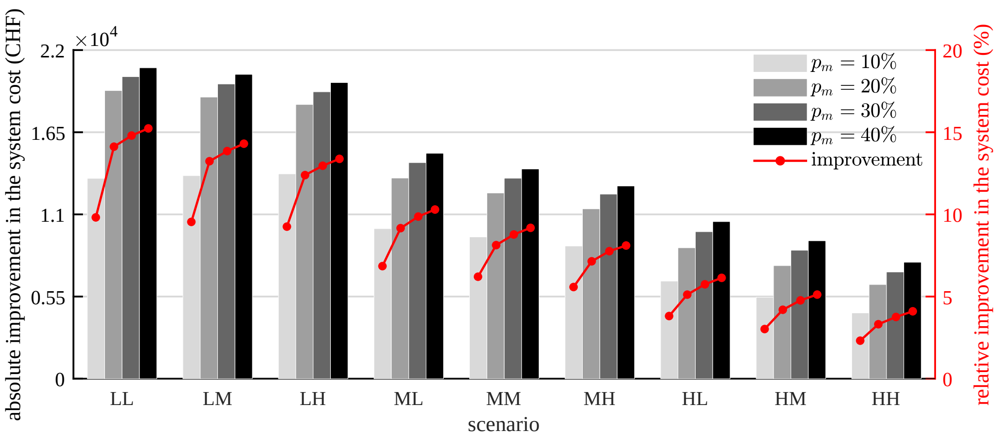

Figure 5: Comparison of the improvement in the total system cost for different penetration rates of modular units.

The first thing to notice is that the proposed system substantially outperforms the bus dispatching system with conventional buses, especially for scenarios with lower car and public transport demand. As the level of demand (in particular for the car mode) increases, the improvement is smaller, but still significant. The reason for this is twofold. First, in case of low demand, the system sends a group of few combined (if not individual) modular units, optimizing thereby the utilization of the vehicle's capacity and reducing the operating cost. Second, due to a lower number of units being contained in a modular bus, the system has more units available in stock, allowing to dispatch more modular buses and reduce the passenger waiting time. This, on the other hand, has minimum impact on car traffic in case of low car demand, due to lower interactions between the modes. That being said, dispatching more modular buses for higher car demand levels does not necessarily have to improve the system, as the impact on car traffic may be significant, resulting in lower overall system performance. This is illustrated in Fig. 6, where we observe that the average (across all bus lines) number of combined modular units increases as the car demand increases, especially during the peak period (Fig. 6b) and for higher penetration rates of modular units.

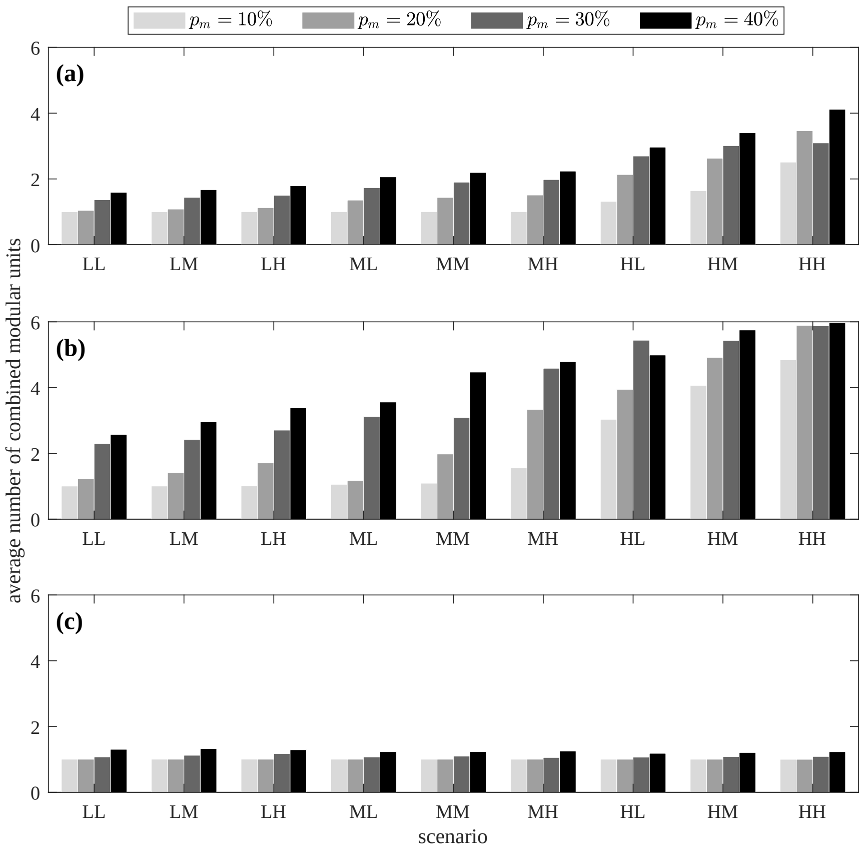

Figure 6: Evolution of the average number of units contained in a modular bus for different penetration rates of modular units: (a) before the peak; (b) during the peak; and (c) after the peak.

Consequently, the improvements in the total system delay do not seem to be significant for congested traffic scenarios (Fig. 7), as the system has less flexibility in managing the allocation of the vehicle resources compared to other scenarios. In other words, in such congested cases, the proposed system tends to behave similarly to the conventional bus dispatching system. Nevertheless, the operator cost function still gets significantly improved (Fig. 8). This shows that the proposed system provides more flexibility for dispatching buses to serve the passenger demand while reducing the operator cost compared to the conventional bus dispatching system, even for congested traffic conditions.

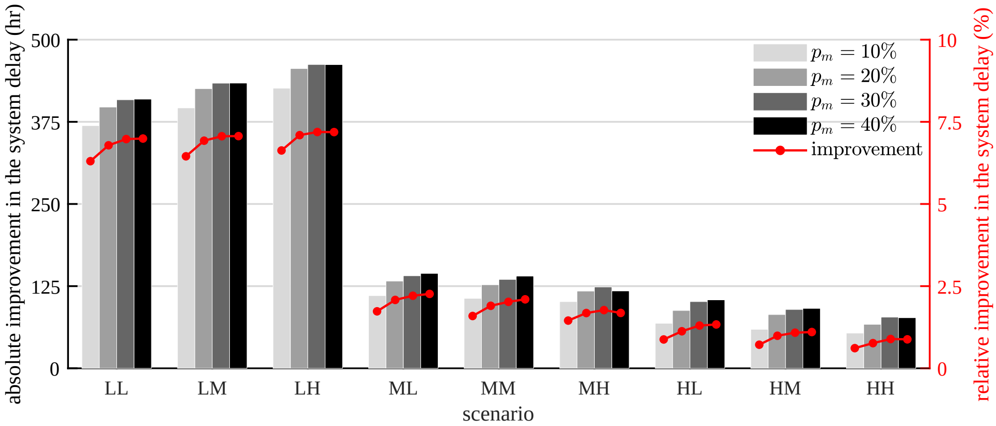

Figure 7: Comparison of the improvement in system delay for different penetration rates of modular units.

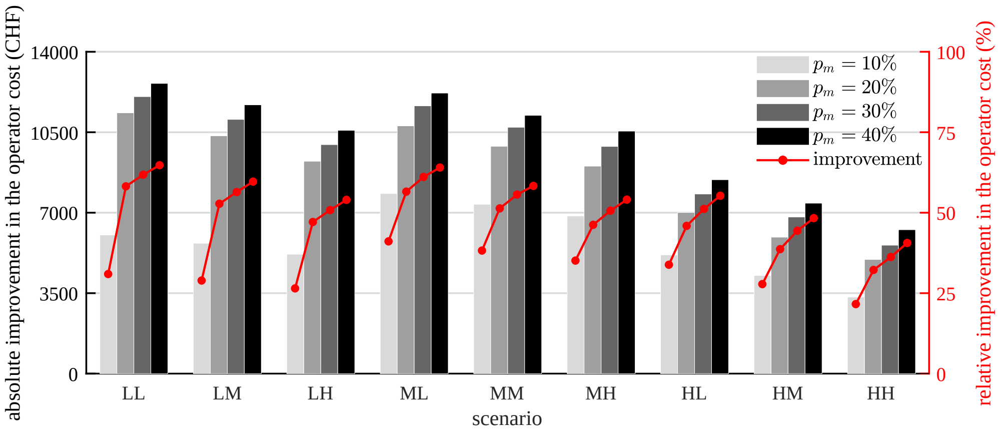

Figure 8: Comparison of the improvement in the operator cost for different penetration rates of modular units.

It is also worth mentioning that with an increase in the penetration rate of modular bus units the improvement in the total system cost also increases. This holds true for all tested scenarios (see Fig. 5). Potential improvements can be up to 16%, depending on the level of car and passenger demand, as well as the penetration rate of modular bus units. As stated before, these improvements mainly come from the significant improvements in the operator cost function, which can be up to 64% for simulated scenarios (Fig. 8).

4.3. Value of considering traffic congestion using the 3D-MFD

So far, we have discussed the performance of the proposed flexible bus dispatching and quantified the improvements compared to the system with only conventional buses. To realistically model both systems, we employed the 3D-MFD to take into account factors such as the complex multimodal interactions and congestion propagation. These factors, however, are ignored in most scientific literature on the frequency setting problem, which typically assumes that the travel times (or equivalently, the bus speeds) are independent of the bus dispatching policy. It is therefore the purpose of this subsection to quantify the value of considering these factors in the proposed modeling framework.

We do this by comparing the results of the proposed approach to those obtained when the optimal solution for the simplified problem (i.e. considering the bus system only) is incorporated into the proposed framework. The results of this comparison are shown in Fig. 9. We observe that significant improvements (up to 54%)

in the total system cost can be achieved with a proper modeling of the complex multimodal interactions and congestion propagation. As expected, the lowest improvements are obtained for the scenarios with uncongested traffic conditions (i.e. low level of car demand) and zero penetration rate of modular units. The reason for this is that, in such scenarios, the interactions between the modes are minimized. As the level of interactions increases, the improvements increase. Notice that the improvements (starting at 10%) also increase once the modular bus units are introduced to the system, even for the uncongested traffic conditions. For congested traffic conditions, the improvements are at least 31% even in the absence of modular units. This indicates the value of the proposed modeling framework in capturing the necessary factors for optimizing the performance of the whole network, while taking into account all transport modes.

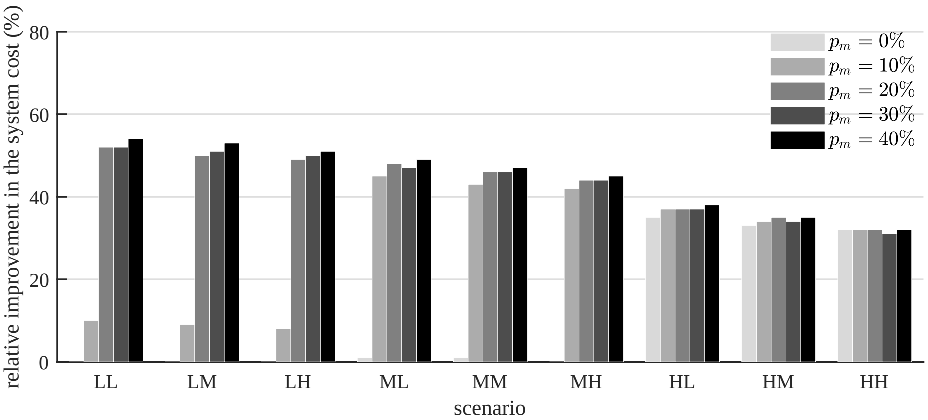

Figure 9: Comparison of the improvement in the system cost made by accounting for the complex multimodal interactions and congestion propagation for different penetration rates of modular units.

5. Conclusion

In this study, we propose a novel concept, called *flexible bus dispatching system*, which offers new perspectives and enormous flexibility to better manage the dispatched frequencies and the allocation of the vehicle resources, reducing thereby the operating cost. In such a flexible bus dispatching system, the bus fleet consists not only of conventional buses, but also of modular and fully automated bus units that can either operate individually or combined together (forming thereby a single modular unit of a higher passenger capacity). To determine the optimal number of combined modular bus units and the optimal frequency at which the units (both conventional and modular) are dispatched across different bus lines, while accounting for the traffic dynamics at the network level, we propose an optimization framework based on the recently proposed three-dimensional macroscopic fundamental diagram (3D-MFD). To the best of our knowledge, this is the first application of the 3D-MFD and modular bus units for the frequency setting problem in the domain of bus operations.

To test the performance and the robustness of the proposed system, we analyze various scenarios, characterizing different levels of car and passenger demand, and penetration rates of modular bus units. Numerical results show that the proposed concept can significantly outperform the existing bus dispatching system with only conventional buses. The improvements in the total system cost are achieved by adapting the number of combined modular bus units and their dispatched frequencies to the evolution of both, the car and the passenger demand. For example, in case of low passenger demand (typically during the off-peak period), the proposed system dispatches compositions of modular bus units that contain only few combined (if not individual) modular units. However, in case of high passenger demand (typically during the peak hour), the dispatched compositions include higher number of combined modular bus units. This way, the proposed system makes a trade-off between the service frequency and the allocation of bus units on the one hand, and the level of service provided to users on the other hand. Finally, a comparison with the approach that considers only the bus system (neglecting the complex multimodal interactions and congestion propagation dynamics) reveals the value of the proposed modeling framework.

A potential future direction is to incorporate the proposed bus dispatching system into a bi-modal perimeter control (Dakic et al., 2019) and/or develop a hierarchical control framework (Yang et al., 2017) that converts the macroscopic level bus dispatching decision to account for more detailed operational requirement (e.g. timetables). Alternatively, we are working on the integration of the proposed system with mobility-on-demand systems, where the public transport operator maintains a larger fleet of automated modular units and some units provide last-mile on-demand services.

6. Acknowledgments

This work was supported by the Swiss National Science Foundation (SNSF) under the project name DIPLOMAT, contract 205121_165644. Kaidi Yang acknowledges funding support by the Swiss National Science Foundation (SNSF) Early Postdoc.Mobility with project number P2EZP2_184244. 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.

References

References

Ambühl, L., Loder, A., Menendez, M., Axhausen, K.W., 2017. Empirical Macroscopic Fundamental Diagrams: New Insights from Loop Detector and Floating Car Data. Presented at 96th Annual Meeting of the Transportation Research Board .

Ambühl, L., Loder, A., Zheng, N., Axhausen, K.W., Menendez, M., 2019. Approximative network partitioning for mfds from stationary sensor data. Transportation Research Record 2673, 94–103.

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.

Axhausen, K.W., König, A., Abay, G., Bates, J.J., Bierlaire, M., 2006. Swiss value of travel time savings. Arbeitsberichte Verkehrs-und Raumplanung 253.

Boggs, P.T., Tolle, J.W., 1995. Sequential quadratic programming. Acta numerica 4, 1–51.

Cacchiani, V., Caprara, A., Toth, P., 2010. Solving a real-world train-unit assignment problem. Mathematical Programming 124, 207–231.

Cadarso, L., Marín, Á., Maróti, G., 2013. Recovery of disruptions in rapid transit networks. Transportation Research Part E: Logistics and Transportation Review 53, 15–33.

Cats, O., Haverkamp, J., 2018. Optimal infrastructure capacity of automated on-demand rail-bound transit systems. Transportation Research Part B: Methodological 117, 378–392.

Ceder, A., 2007. Public transit planning and operation: Modeling, practice and behavior. CRC press.

Ceder, A., Wilson, N.H., 1986. Bus network design. Transportation Research Part B: Methodological 20, 331–344.

Chen, Z., Li, X., Zhou, X., 2019. Operational design for shuttle systems with modular vehicles under oversaturated traffic: Discrete modeling method. Transportation Research Part B: Methodological 122, 1–19.

Chen, Z., Li, X., Zhou, X., 2020. Operational design for shuttle systems with modular vehicles under oversaturated traffic: Continuous modeling method. Transportation Research Part B: Methodological 132, 76–100.

Chiabaut, N., Küng, M., Menendez, M., Leclercq, L., 2018. Perimeter control as an alternative to dedicated bus lanes: A case study. Transportation Research Record 2672, 110–120.

Cordeau, J.F., Soumis, F., Desrosiers, J., 2000. A Benders Decomposition Approach for the Locomotive and Car Assignment Problem. Transportation Science 34, 133–149.

Cordeau, J.F., Soumis, F., Desrosiers, J., 2001. Simultaneous Assignment of Locomotives and Cars to Passenger Trains. Operations Research 49, 531–548.

Daganzo, C.F., 1994. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. Transportation Research Part B: Methodological 28, 269–287.

Daganzo, C.F., 2007. Urban gridlock: Macroscopic modeling and mitigation approaches. Transportation Research Part B: Methodological 41, 49–62.

Dakic, I., Ambühl, L., Schümperlin, O., Menendez, M., 2020. On the modeling of passenger mobility for stochastic bi-modal urban corridors. Transportation Research Part C: Emerging Technologies 113, 146–163.

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., 2019. Evaluating the effects of passenger occupancy dynamics on a bi-modal perimeter control, in: TRB Annual Meeting Online, Transportation Research Board.

Desaulniers, G., Hickman, M.D., 2007. Public transit. Handbooks in operations research and management science 14, 69–127.
Fioole, P.J., Kroon, L., Maróti, G., Schrijver, A., 2006. A rolling stock circulation model for combining and splitting of passenger trains. European Journal of Operational Research 174, 1281–1297.
Furth, P.G., Wilson, N.H.M., 1981. Setting frequencies on bus routes: theory and practice. Transportation Research Record Journal of the Transportation Research Board 818, 1–7.
Ge, Q., Ciuffo, B., Menendez, M., 2015. Combining screening and metamodel-based methods: An efficient sequential approach for the sensitivity analysis of model outputs. Reliability Engineering & System Safety 134, 334–344.
Ge, Q., Menendez, M., 2017. Extending morris method for qualitative global sensitivity analysis of models with dependent inputs. Reliability Engineering & System Safety 162, 28–39.
Geroliminis, N., Daganzo, C.F., 2008. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. Transportation Research Part B: Methodological 42, 759–770.
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.
Gkiotsalitis, K., Cats, O., 2018. Reliable frequency determination: Incorporating information on service uncertainty when setting dispatching headways. Transportation Research Part C: Emerging Technologies 88, 187–207.
Guo, Q.W., Chow, J.Y.J., Schonfeld, P., 2018. Stochastic dynamic switching in fixed and flexible transit services as market entry-exit real options. Transportation Research Part C: Emerging Technologies 94, 288–306.
Hadas, Y., Shnaiderman, M., 2012. Public-transit frequency setting using minimum-cost approach with stochastic demand and travel time. Transportation Research Part B: Methodological 46, 1068–1084.
Han, A.F., Wilson, N.H., 1982. The allocation of buses in heavily utilized networks with overlapping routes. Transportation Research Part B: Methodological 16, 221–232.
He, 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.
Ibarra-Rojas, O., Delgado, F., Giesen, R., Muñoz, J., 2015. Planning, operation, and control of bus transport systems: A literature review. Transportation Research Part B: Methodological 77, 38–75.
Krugman, P., 1991. Geography and trade MIT Press. Cambridge, MA .
Li, Y., Xu, W., He, S., 2013. Expected value model for optimizing the multiple bus headways. Applied Mathematics and Computation 219, 5849–5861.
Lingaya, N., Cordeau, J.F., Desaulniers, G., Desrosiers, J., Soumis, F., 2002. Operational car assignment at VIA Rail Canada. Transportation Research Part B: Methodological 36, 755–778.
Liu, W., Geroliminis, N., 2017. Doubly dynamics for multi-modal networks with park-and-ride and adaptive pricing. Transportation Research Part B: Methodological 102, 162–179.
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., Ambühl, L., Menendez, M., Axhausen, K.W., 2019a. Understanding traffic capacity of urban networks. Scientific reports 9, 1–10.
Loder, A., Dakic, I., Bressan, L., Ambühl, L., Bliemer, M.C., Menendez, M., Axhausen, K.W., 2019b. Capturing network properties with a functional form for the multi-modal macroscopic fundamental diagram. Transportation Research Part B: Methodological 129, 1–19.
Mahmassani, H.S., Saberi, M., Zockaie, A., 2013. Urban network gridlock: Theory, characteristics, and dynamics. Transportation Research Part C: Emerging Technologies 36, 480–497.
Martínez, H., Mauttone, A., Urquhart, M.E., 2014. Discrete Optimization Frequency optimization in public transportation systems: Formulation and metaheuristic approach. European Journal of Operational Research 236, 27–36.
Newell, G.F., 1971. Dispatching Policies for a Transportation Route. Transportation Science 5, 91–105.
Roca-Riu, M., Menendez, M., Dakic, I., Buehler, S., Ortigosa, J., 2020. Urban space consumption of cars and buses: an analytical approach. Transportmetrica B: Transport Dynamics 8, 237–263.
Saeedmanesh, M., Geroliminis, N., 2017. Dynamic clustering and propagation of congestion in heterogeneously congested urban traffic networks. Transportation Research Part B: Methodological 105, 193–211.
Salzborn, F.J.M., 1972. Optimum Bus Scheduling. Transportation Science 6, 137–148.
Salzborn, F.J.M., 1980. Scheduling Bus Systems with Interchanges. Transportation Science 14, 211–231.
Schéele, S., 1980. A supply model for public transit services. Transportation Research Part B: Methodological 14, 133–146.
Sinner, M., 2019. Effects of the Autonomous Bus on the Railway System. Ph.D. thesis. ETH Zurich.
Verbas, Ö., Frei, C., Mahmassani, H.S., Chan, R., 2015. Stretching resources: sensitivity of optimal bus frequency allocation to stop-level demand elasticities. Public Transport 7, 1–20.
Verbas, Ö., Mahmassani, H.S., 2013. Optimal Allocation of Service Frequencies over Transit Network Routes and Time Periods. Transportation Research Record: Journal of the Transportation Research Board 2334, 50–59.
Vuchic, V.R., Day, F.B., Dirshimer, G.N., Kikuchi, S., Rudinger, D.J., 1978. Transit operating manual .
Wang, Y., D'Ariano, A., Yin, J., Meng, L., Tang, T., Ning, B., 2018. Passenger demand oriented train scheduling and rolling stock circulation planning for an urban rail transit line. Transportation Research Part B: Methodological 118, 193–227.
Yang, K., Menendez, M., Zheng, N., 2019. Heterogeneity aware urban traffic control in a connected vehicle environment: a joint framework for congestion pricing and perimeter control. Transportation Research Part C: Emerging Technologies 105, 439–455.
Yang, K., Zheng, N., Menendez, M., 2017. Multi-scale perimeter control approach in a connected-vehicle environment. Transportation Research Part C: Emerging Technologies 94, 32–49.
Yue, Y., Han, J., Wang, S., Liu, X., 2017. Integrated Train Timetabling and Rolling Stock Scheduling Model Based on

Time-Dependent Demand for Urban Rail Transit. Computer-Aided Civil and Infrastructure Engineering 32, 856–873.
Zhang, F., Liu, W., 2019. Responsive bus dispatching strategy in a multi-modal and multi-directional transportation system: A doubly dynamical approach. Transportation Research Part C: Emerging Technologies .
Zheng, N., Geroliminis, N., 2013. On the distribution of urban road space for multimodal congested networks. Transportation Research Part B: Methodological 57, 326–341.