Capturing network properties with a functional form for the multi-modal macroscopic fundamental diagram

Working Paper

Author(s):
Loder, Allister; Dakic, Igor; Bressan, Lea; Ambühl, Lukas; Bliemer, Michiel C.J.; Menendez, Monica; Axhausen, Kay W. (iD)

Capturing network properties with a functional form for the multi-modal macroscopic fundamental diagram

Allister Loder

Igor Dakic

Lea Bressan

Lukas Ambühl

Michiel C.J. Bliemer

Monica Menendez

Kay W. Axhausen

ETH

IVT Institut für Verkehrsplanung und Transportsysteme
Institute for Transport Planning and Systems

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Capturing network properties with a functional form for the multi-modal macroscopic fundamental diagram

Allister Loder
IVT, ETH Zurich
aloder@ethz.ch

Igor Dakic
IVT, ETH Zurich

Lea Bressan
IVT, ETH Zurich

Lukas Ambühl
IVT, ETH Zurich

Michiel C.J. Bliemer
Institute of Transport and Logistics Studies, University of Sydney

Monica Menendez
Division of Engineering, New York University Abu Dhabi
Tandon School of Engineering, New York University

Kay W. Axhausen
IVT, ETH Zurich

Abstract

In urban road networks, the interactions between different modes can clearly impact the overall travel production. Although those interactions can be quantified with the multi-modal macroscopic fundamental diagram; so far, no functional form exists for this diagram to explicitly capture operational and network properties. In this paper, we propose a methodology to generate such functional form, and we show its applicability to the specific case of a bi-modal network with buses and cars. The proposed functional form has two components. First, a three dimensional lower envelope limits travel production to the theoretical best-case situation for any given number of vehicles for the different modes. The lower envelopes parameters are derived from topology and operational features of the road network. Second, a smoothing parameter quantifies how interactions between all vehicle types reduce travel production from the theoretical best-case. The smoothing parameter is estimated with network topology and traffic data. In the case no traffic data is available, our functional form is still applicable. The lower envelope can be approximated assuming fundamental parameters of traffic operations. For the smoothing parameter, we show that it always hold similar values even for different networks, making its approximation also possible. This feature of the proposed functional form is an advantage compared to curve fitting, as it provides a reasonable shape for the multi-modal macroscopic fundamental diagram irrespective of traffic data availability. The methodology is illustrated and validated using simulation and empirical data sets from London and Zurich.

1 Introduction

Cities are complex and so are their transportation systems. As all modes of transportation compete for scarce urban resources, such as space and funding, their interactions should determine the allocation of urban resources needed to achieve optimal productivity of the entire system. In terms of road traffic, one of the most recognizable interactions is between cars and buses and/or trams on urban roads. There exist many studies that summarize the trade-off between the cruising speed and space consumption of the two modes, raising the question of the optimal share between them in terms of vehicles or space. Smeed (1968) was among the first to discuss this problem analytically. However, methods beyond a simple bus-car-equivalent for entire cities did not exist before the concept of the three-dimensional macroscopic fundamental diagram (3D-MFD) was introduced (Geroliminis *et al.*, 2014).

The 3D-MFD originates from the unimodal macroscopic fundamental diagram (MFD) initially proposed by Daganzo (2007). The novel thinking abstracts urban traffic with vehicles flowing through reservoirs until they exit the network. In other words, with this macroscopic perspective the goal is to find the travel time until a vehicle leaves the network, for which the only information required is travel distance and the speed inside each reservoir. The speed must be the average space-mean speed including all delays from stopping at an intersection or bus stop, and/or congestion. Inside each reservoir, all traffic states are captured by the MFD, a well-defined relationship between the accumulation of vehicles and flow or speed of vehicles in each reservoir. The existence of such well-defined relationship is based on several assumptions: traffic should be homogenously distributed, demand should be slowly-varying, the network should be redundant, all links should be relatively similar and their fundamental diagram should not be significantly affected by turning maneuvers (Daganzo and Geroliminis, 2008). That being said, and despite the complexity and irregularities in real urban networks, Geroliminis and Daganzo (2008) showed that the MFD can also be obtained empirically. Moreover, insights from the MFDs have real-world consequences, e.g. for traffic control (Haddad and Geroliminis, 2012) and space allocation in cities (Zheng and Geroliminis, 2013). Nevertheless, the issues of measuring unbiased MFDs (Leclercq *et al.*, 2014), identifying network dynamics (Gayah and Daganzo, 2011), and accounting for the heterogeneous distribution of traffic (Ji and Geroliminis, 2012), must be considered. MFDs are clearly not intended for detailed planning of, e.g. intersection design; but they do provide a good compromise between data requirements and network-wide performance indicators.

In general, multi-modal MFDs (with three dimensions or more) capture the joint effects of multiple modes on the total travel production of the network. In particular, the 3D-MFD captures the joint effects of buses and cars on the total vehicle travel production. On the x and y axes of the 3D-MFD are the accumulations of cars and buses, respectively, while the z-axis describes the total travel production (in vehicle-kilometers per unit time) of all vehicles in the network. With additional information on (average) vehicle occupancy levels, the 3D-MFD can then be expressed in terms of passenger travel production (in passenger-kilometers per unit time) (Geroliminis *et al.*, 2014; Loder *et al.*, 2017; Chiabaut, 2015). Multi-modal MFDs are a powerful tool to investigate and understand the multi-modal performance of entire urban road networks (Ampountolas *et al.*, 2017; Zheng *et al.*, 2017; Amirgholy *et al.*, 2017). They can be estimated using both simulation and empirical observations (Geroliminis *et al.*, 2014; Loder *et al.*, 2017; Castrillon and Laval, 2018; Dakic and Menendez, 2018) or derived numerically (Boyaci and Geroliminis, 2011; Chiabaut, 2015; Dakic *et al.*, 2019). So far, however, no particular functional form for multi-modal MFD exists, and no proposal has been made to link the physical properties of the road and bus network topology as well as the traffic operations to the shape of the 3D-MFD. Such a perspective would facilitate the analysis of the productivity of entire multi-modal urban networks, similarly to the MFD applied to car traffic only (Daganzo *et al.*, 2011). Therefore, in this paper we propose a geometric approach to a functional form for multi-modal MFDs with a particular application to the 3D-MFD.

The proposed functional form has two components. First, a lower multi-dimensional envelope limits travel production to the theoretical best-case situation for any given combination of vehicles in the network (Daganzo and Geroliminis, 2008; Daganzo and Knoop, 2016). Second, a smoothing parameter quantifies how interactions between all vehicle types reduce travel production from the theoretical best-case (Ambühl *et al.*, 2018). The proposed method is inspired by the method of cuts for the MFD estimation (Daganzo and Geroliminis, 2008; Leclercq *et al.*, 2014) and is flexible to accommodate any physical boundaries in urban traffic systems. The multi-dimensional envelope reflects physical network properties and multi-modal traffic dynamics, e.g. traffic signals (Daganzo *et al.*, 2017; Daganzo and Geroliminis, 2008) and bus operations (Boyaci and Geroliminis, 2011; He *et al.*, 2018). The envelopes parameters can be obtained from the topology of the network and traffic data related to the operations of the different modes. In the absence of traffic data, fundamental parameters describing the overall traffic operations can be used instead. The smoothing parameter can be estimated from observed traffic data. If such data is not available, the smoothing parameter can be approximated given that it shows simila values across networks.

The remainder of the paper is organized as follows. Section 2 introduces the general functional form for multi-modal MFDs and discusses its relevant mathematical constraints. Then Section 3 proposes a functional form for the 3D-MFD, discusses its relevant parameters, and shows how the lower envelope of the 3D-MFD is built based on the network and operational properties. In Section 4, we propose how to derive the speed for buses and cars based on the 3D-MFD. Then, Section 5 demonstrates how the passenger 3D-MFD is generated using the proposed methodology. Finally, Section 6 discusses the estimation of the lower envelope for the 3D-MFD in general, and then validates the functional form for the 3D-MFD as well as the proposed speed functions using 3D-MFDs observed from simulation data of an abstract grid network and two empirical data sets, one from London (United Kingdom) and one from Zurich (Switzerland).

2 A new functional form for the multi-modal macroscopic fundamental diagram

The proposed geometric approach to derive a functional form for multi-modal MFDs considers an urban network of total length L with M different modes circulating on it. Each mode $m \in M$ can have its dedicated infrastructure, using a η_m share of L. The share of mixed-used infrastructure used by all modes is then $\eta_{mixed} = 1 - \sum_{m \in M} \eta_m$. Each modes vehicle accumulation is N_m, and we define that $\mathbf{N}$ in $\mathbb{R}_+^M$ is the vector of vehicle accumulations with elements $(N_1; N_2; \ldots; N_M)$. Each mode has a travel production of $\pi_m \in \mathbb{R}_+$ and all modes together have a joint travel production of $\Pi \in \mathbb{R}_+$.

Physical constraints limit the number of vehicles an urban network can accommodate, as well as the travel production for any accumulation $\mathbf{N}$. Thus, we are interested in identifying the $M + 1$ dimensional boundary between the physically possible and impossible traffic states (i.e. the maximum travel production that can be obtained for different combinations of N_m, $\forall m \in M$). That multi-dimensional space includes M dimensions (one for the accumulation of each mode), plus one dimension for the total travel production Π across all modes. The boundary then corresponds to a theoretical best-case situation, which for a variety of intuitive reasons will never be achieved in reality, e.g. heterogeneity of traffic, dynamics of the different modes, vehicle interactions.

We expect the boundary to look differently in each network as infrastructure and vehicle technology parameters can substantially influence the theoretical best-case situation and

thus the resulting boundary. This makes a general mathematical formulation difficult. Suppose that we can define a set of J functions $\Pi_{j \in J} \colon \mathbf{N} \to \Pi$, where each Π_j provides a physically meaningful constraint of one feature of multi-modal traffic (e.g. gridlock, saturation) relating $\mathbf{N}$ and the theoretical best-case travel production Π. Then the network overall theoretical best-case situation across all modes and traffic states is the lower envelope (i.e. minimum) of all Π_j functions as formulated by Eqn. 1. Each Π_j function can be seen as a hyperplane of $M + 1$ dimensions.

$$\Pi\left(\mathbf{N}\right) = \min\left(\Pi_1\left(\mathbf{N}\right); \Pi_2\left(\mathbf{N}\right); \ldots; \Pi_J\left(\mathbf{N}\right)\right) \tag{1}$$

The definition of functions $\Pi_j\left(\mathbf{N}\right)$ is context specific (e.g. in our application to the 3D-MFD in Section 3, we use seven linear functions, i.e. planes, to create the lower envelope). Nevertheless, the function resulting from Eqn. 1 must satisfy the following physical properties.

First, we denote κ_m as the maximum accumulation of vehicles of mode m in the network, and we assume that each possible maximum accumulation leads to gridlock. κ_m is clearly a function of the number of vehicles $\mathbf{N}$ across all modes, because different modes might share a portion of the infrastructure, η_{mixed}. Eqn. 2 shows this constraint for κ_m. Assuming l_m is the jam spacing for mode m, then $\kappa_m\left(\mathbf{N}\right)$ is equal to the gridlock accumulation of mode m on its dedicated and mixed-used infrastructure, minus the number of vehicles from other modes $m' \in M \setminus \{m\}$ present on the mixed-use infrastructure. For simplicity, we assume that each mode m is fully using its dedicated infrastructure. Thus, κ_m has lower and upper bounds given by $\dfrac{L}{l_m}\eta_m \leq \kappa_m \leq \dfrac{L}{l_m}\left(\eta_m + \eta_{mixed}\right), \forall m \in M$.

$$\kappa_m\left(\mathbf{N}\right) = \frac{L}{l_m}\left(\eta_m + \eta_{mixed}\right) - \sum_{m' \in M \setminus \{m\}} \frac{l_{m'}}{l_m}\left(\kappa_{m'} - \frac{L}{l_{m'}}\eta_{m'}\right), \quad \forall m \in M \tag{2}$$

Second, we require that travel production is always positive as long as the accumulation

of any mode is non-negative and not all modes are gridlocked. See Eqn. 3.

$$\Pi\left(\mathbf{N}\right) > 0 \quad \text{for } 0 < N_m < \kappa_m, \quad \forall m \in M \tag{3}$$

Third, we define the average network speed as $V \equiv \Pi / \sum_m N_m$, so that it must always decrease with any accumulation of vehicles, as given by Eqn. 4.

$$\frac{\partial V}{\partial N_m} < 0, \quad \forall m \in M \tag{4}$$

Fourth, we require that the production is zero, i.e. $\Pi\left(\mathbf{N}\right) = 0$, if either no vehicles circulate on the network, or all modes reach κ_m, or any combination of these two cases, as formulated in Eqn. 5.

$$\Pi\left(\mathbf{N}\right) = 0, \text{if } N_m = \kappa_m \vee N_m = 0, \quad \forall m \in M \tag{5}$$

Fifth, we require that that the total production Π (in Eqn. 6) equals the sum of all modes productions π_m. As Eqn. 1 provides a lower envelope for the theoretical best-case situation, we are particularly interested in the maximum possible production for any accumulation of vehicles. Thus, we have to find the distribution of vehicles among η_m and η_{mixed} that maximizes Π.

$$\Pi\left(\mathbf{N}\right) = \max \sum_{m \in M} \pi_m\left(N_m\right) \tag{6}$$

Eqns. 1-6 provide a meaningful lower envelope for a theoretical best-case situation. As it is expected that this best-case situation will never be achieved, we propose to reduce the travel production by $\lambda\left(\mathbf{N}\right)$, as formulated in Eqn. 7 with the smooth approximation

of the minimum operator (Cook, 2011; Bliemer *et al.*, 2017; Ambühl *et al.*, 2018).

$$\Pi\left(\mathbf{N}\right) = -\lambda\left(\mathbf{N}\right)\log\left(\exp\left(-\frac{\Pi_1\left(\mathbf{N}\right)}{\lambda\left(\mathbf{N}\right)}\right) + \exp\left(-\frac{\Pi_2\left(\mathbf{N}\right)}{\lambda\left(\mathbf{N}\right)}\right) + \cdots + \exp\left(-\frac{\Pi_J\left(\mathbf{N}\right)}{\lambda\left(\mathbf{N}\right)}\right)\right) \tag{7}$$

The smoothing parameter $\lambda\left(\mathbf{N}\right)$ has an interesting interpretation in this context as it quantifies with a single metric the losses in the travel production due to vehicle interactions not captured by the functions Π_j. It is clear that this parameter is non-linear, but provides an opportunity to reduce the complexity of the problem. The larger the value for $\lambda\left(\mathbf{N}\right)$ is, the stronger the vehicle interactions become, resulting in higher reduction of the total production. In case $\lambda\left(\mathbf{N}\right) = 0$, no vehicle interactions in addition to those captured by functions Π_j take place.

To be precise, as Ambühl *et al.* (2018) show, $\lambda\left(\mathbf{N}\right)$ is also a measure of heterogeneity in the network. Thus, $\lambda\left(\mathbf{N}\right)$ not only describes the flow reduction from the theoretical best-case due to between-vehicle interactions, but also other unobserved factors that lead to an observed production that is normally below the theoretical best-case, e.g. spatial and temporal heterogeneity of flow, irregularities in traffic operations, etc. Unfortunately, so far, no approach exists to determine the precise contributions of all these sources to $\lambda\left(\mathbf{N}\right)$.

A potential procedure to estimate $\lambda\left(\mathbf{N}\right)$ is to use a non-linear regression on multi-modal MFD observations, e.g. empirical or simulation data (Ambühl *et al.*, 2018). More details are provided later in this paper. However, future research could explore whether $\lambda\left(\mathbf{N}\right)$ can be derived analytically or predicted with a model estimated from empirical observations.

3 A functional form for the 3D-MFD

Here, as a particular application of the multi-modal MFD described above, we create a functional form for the 3D-MFD by defining a three dimensional lower envelope consisting of J planes, which provides an upper limit for the travel production, given different accumulations of cars and buses. The 3D-MFD must be specified for each city context.

The set of planes required for a fully operational functional form for the 3D-MFD must create a geometric shape of a closed cover without holes and other irregularities. Further, the chosen planes should result in a geometric shape that resembles the 3D-MFD very closely. Therefore, planes should describe free-flow, saturated and congested traffic conditions as well as mark the gridlock states for both modes. Only when the planes are carefully selected, the functional form will provide a physically meaningful representation of macroscopic bus and car traffic in cities. Here, we propose to use seven planes. For this, we use the following notation. Eqn. 8 defines plane j in the three-dimensional space given by N_c, N_b, and Π, where $(car_{0,j}, bus_{0,j}, \pi_{0,j})$ are the coordinates of a point in plane j, and $(car_{n,j}, bus_{n,j}, \pi_{n,j})$ is the corresponding normal vector.

$$\left(\begin{pmatrix} N_c \\ N_b \\ \Pi \end{pmatrix} - \begin{pmatrix} car_{0,i} \\ bus_{0,i} \\ \pi_{0,i} \end{pmatrix} \right)^T \cdot \begin{pmatrix} car_{n,i} \\ bus_{n,i} \\ \pi_{n,i} \end{pmatrix} = 0 \tag{8}$$

After solving Eqn. 8 for Π, we obtain Eqn. 9 for plane Π_j as a function of both modes vehicle accumulations. Recall that each plane $\Pi_j(N_c, N_b)$ is a function required for the general functional form as given in Eqn. 7.

$$\Pi_j(N_c, N_b) = \pi_{0,j} - \left(car_{n,j}(N_c - car_{0,j}) + bus_{n,j}(N_b - bus_{0,j}) \right) / \pi_{n,j} \tag{9}$$

The planes for Eqn. 7 will be defined from eleven points that we will introduce below. Figure 5 shows the location of all these points in the 3D-MFD. Here, we focus primarily on planes defining a theoretical best-case situation based on geometric and operational issues. For the operational features, we follow in most cases the basic ideas by Daganzo and Geroliminis (2008) to calculate the points that account for the delays at the network level caused by intersections and signals. Importantly, our proposed set of planes can be further extended and refined with city-specific operational features of multi-modal traffic. All planes are defined by at least three points in the 3D-MFD, which are given in the Cartesian coordinate system in the order of N_c, N_b and Π. All resulting planes are summarized in Table 2.

Name	Key	Unit	Additional comments
Variables			
Accumulation of cars	N_c	[veh]	Normalized to density by dividing by the car network length $(1 - \eta_b)L$.
Accumulation of buses	N_b	[veh]	Normalized to density by dividing by the bus network length $(1 - \eta_c)L$.
Total travel production	Π	[veh-km/h]	
Average speed in the network	v_{MFD}	[km/h]	
Speed of cars in the network	$v_{c,MFD}$	[km/h]	
Speed of buses in the network	$v_{b,MFD}$	[km/h]	
Parameters			
Smoothing parameter	λ	[-]	Value is shape specific.
Total network length	L	[km]	Entire urban road infrastructure.
Fraction of car only roads	η_c	[-]	
Fraction of bus only roads	η_b	[-]	
Jam spacing of cars	l_c	[km]	
Passenger car equivalent of buses	φ	[-]	
Average car delay at intersections	δ_c	[h]	
Free-flow speed of cars in links	$v_{c,0}$	[km/h]	
Backward wave speed of cars in links	$w_{c,0}$	[km/h]	
Free-flow speed of buses in links	$v_{b,0}$	[km/h]	
Backward wave speed of buses in links	$w_{b,0}$	[km/h]	
Car saturation rate at intersections	s_c	[veh/h-lane]	
Bus saturation rate at intersections	s_b	[veh/h-lane]	Computed from the bus fundamental diagram.
Average link length	l	[km]	
Average bus stop spacing	p	[km]	
Average dwell time	Δ	[s]	
Average cycle length	C	[s]	
Average effective green time	G	[s]	
Public transport strategy	ζ	[-]	0,if buses have ultimate priority at signals. 1, if buses are entirely mixed with cars and experience the same delay.

Table 1: 3D-MFD variables in the upper part and input parameters for the lower envelope for the 3D-MFD. Further parameters calculated from these input parameters are not shown for convenience.

3.1 Definition of points

In this section, we consider that intersections are spaced on average at a distance l. The entire network length is, as before, given by L and expressed in lane-kilometers. We distinguish three different types of lanes: dedicated car lanes, dedicated bus lanes, and mixed lanes. Here, we consider η_b and η_c, i.e. the fraction of the network length where only buses and cars circulate, respectively. Thus, the total network length where only cars circulate is $\eta_c L$, where only buses circulate is $\eta_b L$, and where both modes circulate is $(1 - \eta_b - \eta_c)L$. The bus stops are spaced on average with a distance p. The headway of buses then follows from the design of the network and the number of buses (Daganzo, 2010). Table 1 summarizes all variables and parameters used here.

One trivial point in the system is point $P_0 = (0, 0, 0)^T$ at the origin, where no vehicles in the network circulate and there is no travel production. Consider then the top down view on the car and bus accumulation coordinates in Figure 1. Cars can drive on $(1 - \eta_b) L$ of the network and have a jam spacing l_c. The point P_1 is the maximum accumulation in the network of cars that can be achieved when no public transport operates and the network is gridlocked. The point then equals to:

$$P_1 = \begin{pmatrix} (1 - \eta_b) L / l_c \\ 0 \\ 0 \end{pmatrix}$$

The point P_2 considers the case when no cars circulate and the bus system is gridlocked on $(1 - \eta_c) L$ of the network. For buses, we make the assumption that the bus-car equivalent is φ. Then, the point is defined by:

$$P_2 = \begin{pmatrix} 0 \\ (1 - \eta_c) L / (l_c \varphi) \\ 0 \end{pmatrix}$$

Next, we define the gridlock boundary under mixed traffic conditions with points P_3 and P_4, where the production of traffic is still zero. The point P_3 describes the gridlock case
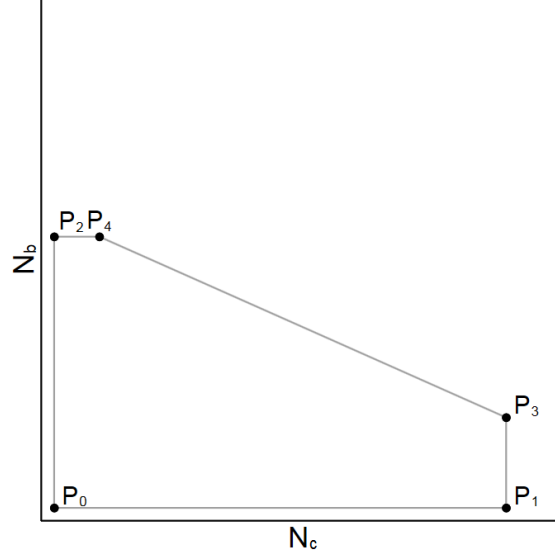
Figure 1: Top-down view on the 3D-MFD which shows the bimodal vehicle accumulation plane. For the defined points, total travel production is zero. The lines connecting the points also represent all combinations of accumulations where travel production is zero.

for all car and mixed lanes, as well as the gridlock scenario on all dedicated bus lanes:

$$
P_3 = \begin{pmatrix} (1 - \eta_b)\, L/l_c \\ \eta_b L/\left(l_c \varphi\right) \\ 0 \end{pmatrix}
$$

On the other hand, the point P_4 mirrors this behavior for buses and is defined with

$$
P_4 = \begin{pmatrix} \eta_c L/l_c \\ (1 - \eta_c)\, L/\left(l_c \varphi\right) \\ 0 \end{pmatrix}
$$

We now consider the situation where no public transport operates, i.e. $N_b = 0$ (Figure 2). At intersections, cars experience on average a delay of δ_c that reduces the road free-flow speed $v_{c,0}$ at the network level to v_c. Here, we approximate the network-wide average free-flow speed v_c as by Eqn. 10, although other more sophisticated approaches could also be used, e.g. by Daganzo and Geroliminis (2008). We use the procedure by Daganzo and Geroliminis (2008) to obtain the network-wide average backward wave speed for cars w_c based on Eqn. 10. In detail, Daganzo and Geroliminis (2008) reverse the direction of the moving observer by replacing $v_{c,0}$ with $w_{c,0}$ and account for the cycle length in the
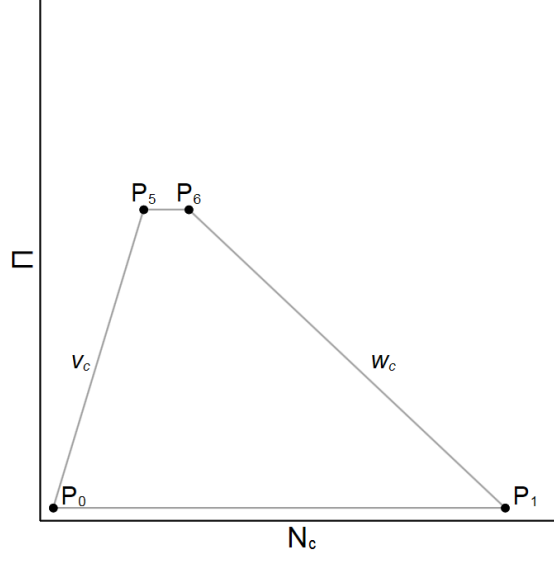
Figure 2: Side view on the 3D-MFD on the total travel production and car accumulation plane. For the defined points, no public transport operates.

signal offsets (for simplicity not included here).

$$v_c = \frac{l}{\dfrac{l}{v_{c,0}} + \delta_c} \tag{10}$$

Figure 2 shows the discussed situation with $N_b = 0$ and the network-wide speeds v_c and w_c, which is basically the car MFD. The points P_0 and P_1 are already defined (see Figure 1) and only points P_5 and P_6 remain to be defined. Both points share the common total travel production given by Eqn. 11.

$$\Pi_c = s_c \frac{G}{C} \left(1 - \eta_b\right) L \tag{11}$$

Equation 11 is derived from the stationary cut defined by Daganzo and Geroliminis (2008) with the saturation flow s_c, the average cycle length C, and the average effective green time G. We derive N_c for both points with the fundamental relationship of traffic using the free flow speed v_c and backward wave speed w_c as shown in Figure 2. Then, points P_5 and P_6 equal to:
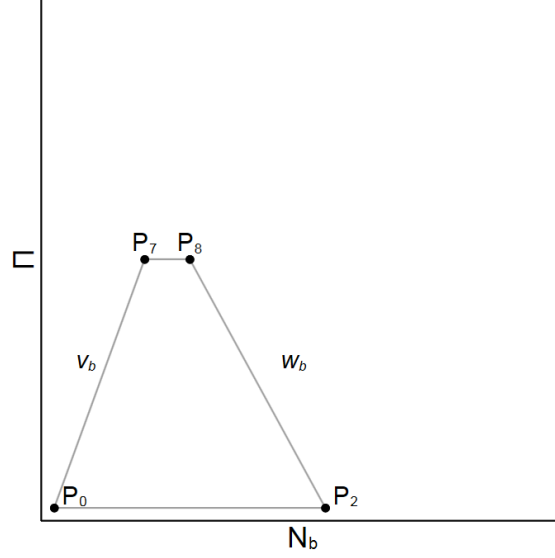
Figure 3: Side view on the 3D-MFD with the total travel production versus the accumulation of buses. For the defined points, no cars are operating.

$$P_5 = \begin{pmatrix} \Pi_c/v_c \\ 0 \\ \Pi_c \end{pmatrix}$$

$$P_6 = \begin{pmatrix} (1 - \eta_b)\dfrac{L}{l_c} - \Pi_c/w_c \\ 0 \\ \Pi_c \end{pmatrix}$$

We now follow the same rationale for the case with only buses operating and no cars circulating, i.e. $N_c = 0$. Figure 3 exhibits this situation where the points P_0 and P_2 are already defined (see Figure 1). Thus, we define points P_7 and P_8 similarly to P_5 and P_6. However, buses have to stop not only at intersections but also at bus stops for boarding and alighting of passengers during the dwell time period Δ. For simplicity, we assume that the dwell time also includes deceleration and acceleration, and that the public transport operator has defined for each stop a scheduled arrival and departure time so that Δ is, under normal circumstances, independent of the demand and human behavior.

Some public transport agencies might include a buffer in the travel time between two stops that allows an operation within a rigid time table even under more congested traffic situations. For that purpose, buses would be equipped with a device that gives drivers advice about the scheduled travel time and that tells them to drive faster or slower, e.g. as implemented in Zurich's bus and tram system. As a consequence, buses would drive at a maximum speed $v_{b,0}$ lower than that of cars $v_{c,0}$ in light traffic conditions. Bus stops are placed at an average distance p. Some cities, e.g. Zurich, have decided to give public transport full priority at each intersection to minimize delay for public transport vehicles, but other cities do not make a difference between cars and public transport vehicles so that public transport vehicles experience the same delay as cars. To capture these different strategies, we define the public transport strategy parameter $\zeta \in [0,1]$. The value zero means full public transport priority and one means buses are fully integrated in traffic, i.e. they experience all car delays. Thus, the maximum commercial speed of buses v_b that includes dwelling, intersections delay, and the time table buffer is given by Eqn. 12 that follows the idea by Daganzo (2010), but accounts for the public transport priority strategy ζ and simplifies the dwelling behavior. The bus backward wave speed w_b is then obtained using the same procedure as for cars in Eqn. 10.

$$v_b = \frac{p}{\dfrac{p}{v_{b,0}} + \delta_c \zeta \dfrac{p}{l} + \Delta} \tag{12}$$

We assume that the maximum travel production of buses Π_b is achieved for the case of full public transport priority ($\zeta = 0$) with no delay at intersections. Π_b will be lower in case of similar delays at intersections for buses and cars ($\zeta = 1$). Similar to Π_c in Eqn. 11 for car traffic, we then define Π_b for the bus system in Eqn. 13 with two parts. First, the left part with term $s_b L (1 - \eta_c)$ quantifies the possible maximum production in case of no delays or in case no bus stops are considered. Second, the right part quantifies the fraction of this possible maximum production that can be realized when delays and stopping behavior are accounted for. This idea is inspired by the stationary cut in Daganzo and Geroliminis (2008), where in the G/C ratio G describes moving and C moving and waiting part. Here, this part considers the average fraction of time buses are moving during the journey from one stop to the next, where buses move for time $p/v_{b,0}$, but also experience intersection
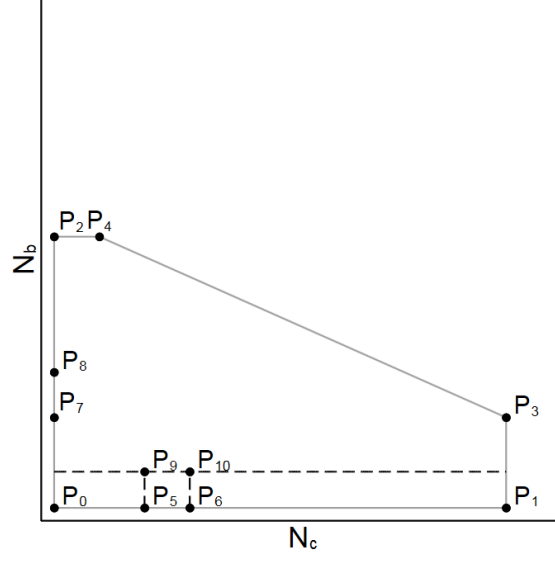
Figure 4: Additional points for the capacity in the 3D-MFD when buses on their dedicated lanes increase travel production from saturated car traffic.

delays $\delta_c \zeta$ at p/l crossing between two stops and delays due to dwelling Δ.

$$\Pi_b = s_b L \left(1 - \eta_c\right) \left(\frac{\dfrac{p}{v_{b,0}}}{\dfrac{p}{v_{b,0}} + \delta_c \zeta \dfrac{p}{l} + \Delta} \right) \tag{13}$$

Consequently, points P_7 and P_8 are defined by:

$$P_7 = \begin{pmatrix} 0 \\ \Pi_b / v_b \\ \Pi_b \end{pmatrix}$$

$$P_8 = \begin{pmatrix} 0 \\ \left(1 - \eta_c\right) L / \left(l_c \varphi\right) - \Pi_b / w_b \\ \Pi_b \end{pmatrix}$$

15

We then define two additional points P_9 and P_{10}. They describe the influence of dedicated bus lanes on the multi-modal capacity of the network. Consider the case with buses running on dedicated lanes. Car traffic is not obstructed by buses and when more buses are added to the system, the travel production is increased at the same level of car accumulation. Given that in this framework we are interested in providing an upper bound for bi-modal traffic, the highest travel production is achieved when all dedicated bus lanes operate at capacity and the remaining part of the network, i.e. all dedicated car and mixed lanes, are saturated with cars.

Thus, the points P_9 and P_{10} are at the same car accumulation as P_5 and P_6, and at a public transport accumulation where capacity is reached on the dedicated bus lanes. The latter sounds rather unrealistic when buses only use their dedicated lanes and not the mixed lanes as they usually run on fixed routes. Nevertheless, as we are interested in finding the highest possible production of bus kilometers points P_9 and P_{10} are then defined as:

$$P_9 = \begin{pmatrix} \Pi_c/v_c \\ \Pi_b \dfrac{\eta_b}{1-\eta_c}/v_b \\ \Pi_b \dfrac{\eta_b}{1-\eta_c} + \Pi_c \end{pmatrix}$$

$$P_{10} = \begin{pmatrix} (1-\eta_b)\dfrac{L}{l_c} - \Pi_c/w_c \\ \Pi_b \dfrac{\eta_b}{1-\eta_c}/v_b \\ \Pi_b \dfrac{\eta_b}{1-\eta_c} + \Pi_c \end{pmatrix}$$

The previously discussed set of points describes the physical limits of the system from a geometric perspective, from which we derive the lower envelope for the 3D-MFD. Points P_0 to P_4 clearly satisfy conditions for the functional form proposed in Eqn. 2 and 5, while points P_5 to P_{10} satisfy Eqn. 6. However, there are many operational features in multi-modal networks that might further limit their productivity. These operational aspects can describe conflicts in mixed traffic, bus bunching, dwelling behavior, effects of the built environment, bus stop design (curb side or bus bay), public transport network design, and routing in the network.

Plane	Points	Description
I	P_0, P_7, P_9	Free flow traffic conditions in both modes. The tilting of the plane describes the trade-off between cars and buses achieving the same travel production.
II	P_1, P_3, P_{10}	Congested traffic states for car traffic.
III	P_5, P_6, P_9, P_{10}	Describes increase in travel production when car network is saturated and buses on their dedicated network add to the travel production.
IV	P_3, P_4, P_9	Wave speed of mixed traffic. The points P_9 is favored over P_{10} as traffic states might deteriorate faster due to mixed traffic conditions.
V	P_7, P_8, P_9	Capacity trade-off between buses and cars when the bus system operates in the saturated state.
VI	P_2, P_8, P_9	Capacity trade-off between buses and cars when the bus system operates in the congested state.
VII	P_2, P_4, P_9	Capacity trade-off between buses and cars when both system operate in the saturated state. We extend this plane also to the congested state of buses to reduce the number of planes.

Table 2: Construction of the seven planes from the eleven points.

3.2 Definition of the planes

Based on the eleven points introduced above, we now propose a lower envelope for the 3D-MFD defined by a set of seven planes. Table 2 numbers the planes and shows which combination of points defines each plane. Figure 5 shows the resulting planes, as well as the resulting shape of the 3D-MFD for an artificial network[1]. We consider this set as the minimal amount of planes required to describe the fundamental relationships in the 3D-MFD. The proposed set of planes then satisfies Eqns. 3 and 4. It should be clear, however, that this set of eleven points and the set of seven planes is not the ultimate and complete solution to describe the 3D-MFD. Some other points can also be defined to capture certain traffic characteristics. Also, more planes can be introduced, out of which some can become non-binding. It is also worth mentioning that there might be other combinations of points for the proposed planes that lead to a comparable 3D-MFD shape. Following the rationale for points P_9 and P_{10}, for example, one would argue for two similar points and the related planes to describe the influence of the dedicated car lanes. However, we omit those because the resulting planes would *chop* the lower envelope and thus create an unreasonable 3D-MFD shape for typical common bus and car parameters.
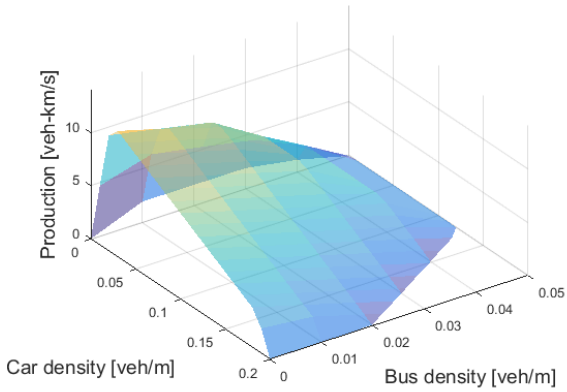
---

[1]In Figure 5 and in Section 6, we use the vehicle density (having the unit of vehicles per lane-meter) instead of accumulation, and express the travel production in vehicle kilometers per second instead of hour to achieve manageable and more comprehensible numbers.
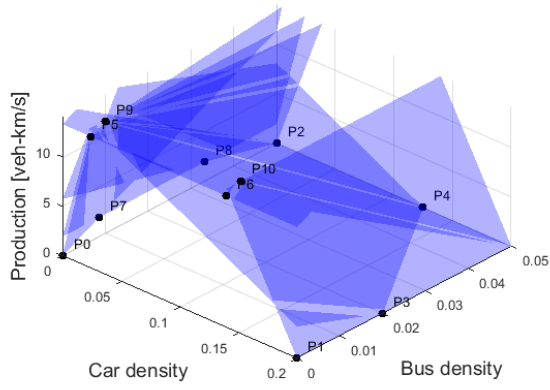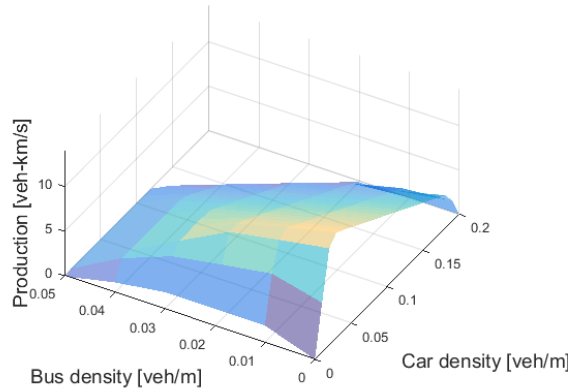
(a) Planes I-VII, view A
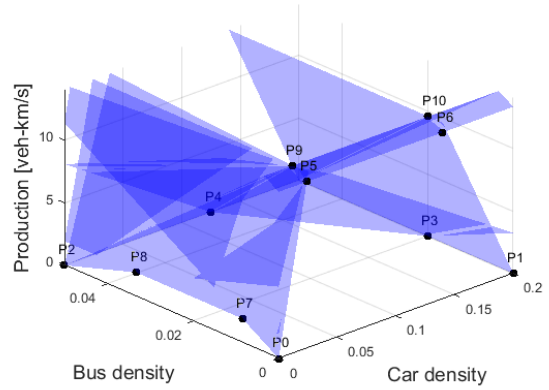
(b) Planes I-VII, view B

(c) 3D-MFD with planes I-VII, view A

(d) 3D-MFD with planes I-VII, view B

Figure 5: Illustrating the planes and the resulting 3D-MFD

In the following, we explain the psychical meaning of each of the defined planes.

Plane *I* characterizes mixed traffic conditions when both modes operate in the free flow conditions. The tilting of this plane represents the number of cars that can be replaced by a bus to maintain the same travel production. Plane *II* captures the traffic states when car traffic is congested and buses operate in the free flow conditions. Plane *III* describes the behavior in the network such that, when all roads where cars can circulate operate in the saturated state, running more buses on the dedicated bus network increases the total travel production. This contribution is usually rather small, but can become substantial in case of bus rapid transit systems and when the traffic performance is analyzed from a passenger perspective. Plane *IV* models the congested states in both modes and all sub-networks. We propose to use P_6 instead of possibly more obvious P_{10} for two reasons. First, once the multi-modal system reaches its capacity, we consider that the system in mixed traffic deteriorates faster to gridlock than if only cars would circulate. Second, with this approach, we require a fewer number of planes, making the entire formulation

of the 3D-MFD much simpler. We include planes V and VI to complement plane I in the domain of saturated and congested bus operations and to limit the travel production between the defined maximum points of buses and cars. Although these traffic states might be rarely observed in bus networks, considering these planes might be relevant to cities where bus lines at central hub locations are overlapping to a large extent, e.g. in Zurich or Brisbane. The last required plane is plane VII that closes the 3D-MFD and creates the familiar shape. It describes the trade-off between buses and cars when both are operating in the saturated state. We extend this plane to the congested regime of buses so that we can reduce the number of planes and thereby the complexity of the formulations.

In Figures 5(c) and 5(d) we then use the set of planes from Table 2 and Eqn. 7 to derive the functional form for the 3D-MFD and to obtain the familiar shape as introduced by Geroliminis *et al.* (2014). For simplicity, we assume an average value of λ that we set to $\lambda = 0.1$. This ensures a tight fit of the curve to the shape defined by the planes, resulting in a rather edged-shape 3D-MFD. Recall that the larger the value for parameter λ is, the smoother the entire shape would become.

It is important to underline that the definitions and assumptions for the points are optimistic and describe the highest possible travel production for any combination of bus and car accumulations. Although empirical observations will be, by definition, always below the defined curve of the 3D-MFD, the deviation can be substantial when the vehicle interactions (measured by λ) are rather strong and spatial heterogeneity is large (for details see Ambühl *et al.* (2018)). However, as Eqn. 7 is flexible to accommodate more realistic points and planes, especially in terms of bus operations, a more accurate 3D-MFD for the specific context can be obtained. For example, the method of cuts by Daganzo and Geroliminis (2008) can further be used - if applicable - to derive more points and planes, potentially resulting in a tighter 3D-MFD shape.

In this section, we implicitly assume that all points for the planes are defined based on external data and that λ is the only free parameter that can either be estimated from observations or simply derived from other studies. However, it is also possible to estimate some points for the planes from observations to obtain a tighter (and better) fit of the lower envelope (lower RMSE), but this may partially cut down the functional form's physical interpretation. In the end, the analyst has to decide and balance his/her priorities in how (s)he wants to work with this flexible functional form.

4 Derivation of speed functions

Unfortunately, in the definition of the 3D-MFD as given in Eqn. 7, bus speed $v_{b,MFD}$ and car speed $v_{c,MFD}$ cannot endogenously be derived from the 3D-MFD itself as the 3D-MFD only provides the average speed of the system with $v_{MFD} := \Pi/(N_c + N_b)$. Thus, to solve for each modes commercial speed, we require another constraint. Geroliminis *et al.* (2014) proposed to use a linear relationship between the speeds of both modes as provided by Eqn. 14.

$$v_b \cong \theta v_c + \beta \tag{14}$$

The parameter β is the intercept and θ is the slope of the linear relation and the parameters have to be estimated from data. Based on Eqn. 14, the speed of buses v_b and of cars v_c then follow according to Eqns. 15 and 16, respectively.

$$v_b = \frac{\Pi - \beta N_b}{N_c + \theta N_b}\theta + \beta \tag{15}$$

$$v_c = \frac{\Pi - \beta N_b}{N_c + \theta N_b} \tag{16}$$

For our analysis, we propose to approximate θ and β of Eqn. 14 with the parameters of the 3D-MFD shape as provided by Table 1 instead of an estimation from data. As we require only two points for a linear relationship, we evaluate bus speeds, first, at the maximum car speed ($v_{c,MFD} = v_c$) and, second, at the minimum car speed $v_{c,MFD} = 0$. First, in light traffic, i.e. $v_{c,MFD} = v_c$, buses are not obstructed by cars, and thus buses are running at $v_{b,MFD} = v_b$. Second, when the entire car network is jammed, i.e. $v_{c,MFD} = 0$, buses are only able to move on their dedicated lanes with v_b and are jammed everywhere else. Thus, we simply consider $v_{b,MFD} = v_b \eta_b/(1 - \eta_c)$ because only the fraction of buses running on dedicated lanes are producing vehicle kilometers. Then, the approximations

for β and θ follow according to Eqn. 17 and 18, respectively.

$$\beta = v_b \frac{\eta_b}{1 - \eta_c} \tag{17}$$

$$\theta = \frac{v_b}{v_c} \left(1 - \frac{\eta_b}{1 - \eta_c} \right) \tag{18}$$

Nevertheless, we consider that Eqn. 14 only holds approximately and not over the entire range of vehicle accumulations, especially at the interval borders of one of the variables. Thus, we require additional constraints for the speed of buses from Eqn. 15 and cars from Eqn. 16 to ensure that speeds do not exceed physical limits. In particular, we require that, for each mode, the speed is always within the fundamental diagram of that mode by using the minimum operator as given for bus speeds in Eqn. 19 and car speeds in Eqn. 20. Here, we consider a trapezoidal fundamental diagram (Daganzo, 1994) with parameters from Table 1. The speeds are obtained by evaluating the fundamental diagram at the respective density (dividing vehicle accumulations by network length) of each mode, for buses $v_{b,FD} \left(N_b / (1 - \eta_c) \right)$ and for cars $v_{c,FD} \left(N_c / (1 - \eta_b) \right)$. Additionally in Eqn. 19, we require that the bus speed equals the MFD speed in cases when the car network is approaching gridlock to avoid negative bus speeds.

$$v_{b,MFD} \left(N_b, N_c \right) = \min \left(v_{b,FD} \left(N_b \right) ; \frac{\Pi - \beta N_b}{N_c + \theta N_b} \theta + \beta ; v_{MFD} \left(N_b, N_c \right) \right) \tag{19}$$

$$v_{c,MFD} \left(N_b, N_c \right) = \min \left(v_{c,FD} \left(N_c \right) ; \frac{\Pi - \beta N_b}{N_c + \theta N_b} \right) \tag{20}$$

5 Applications of the functional form: The passenger 3D-MFD

So far, we have followed only a vehicle perspective, but since the goal of urban transportation systems is to transport people and not vehicles, we discuss in this section how the proposed functional form aligns with a passenger perspective, namely the passenger 3D-MFD (Geroliminis *et al.*, 2014; Chiabaut, 2015; Loder *et al.*, 2017). Generally, we can assume that the passenger travel production Π_{pax} follows from Eqn. 21.

$$\Pi_{pax} = N_b h_b v_{b,MFD} + N_c h_c v_{c,MFD} \tag{21}$$

where h_b and h_c are the vehicle occupancies for buses and cars, respectively.

Below, we show three different possibilities to estimate the passenger 3D-MFD. First, in Section 5.1 we discuss the passenger capacity 3D-MFD. This one describes the maximum number of passengers that could be accommodated in an urban network, using h_b and h_c as the vehicle passenger capacity, i.e. the maximum possible number of passengers inside a single vehicle. Second, in Section 5.2 we discuss the passenger preference 3D-MFD. In this case, h_b and h_c result from travelers preferences for a certain mode given a set of explanatory variables, e.g. travel times, network features, and vehicle crowding. Third, in Section 5.3 we discuss the measured passenger 3D-MFD. Here, we assume that h_b and h_c are available from observations, and can vary in time and space.

5.1 Passenger capacity 3D-MFD

In Figure 6 we provide an example for a passenger capacity 3D-MFD which is based on the 3D-MFD shown in Figure 5. We assume an average car occupancy of $h_c = 2$ passengers and an average bus occupancy of $h_b = 80$ passengers, corresponding to peak hour conditions, thus unrealistically high for low demand periods. These 3D-MFDs can be seen as the maximum possible production of the system if every public transport vehicle was used to capacity. Figure 6(a) shows the passenger capacity 3D-MFD based on the average speed v_{MFD} in the 3D-MFD and Figure 6(b) based on each modes' individual
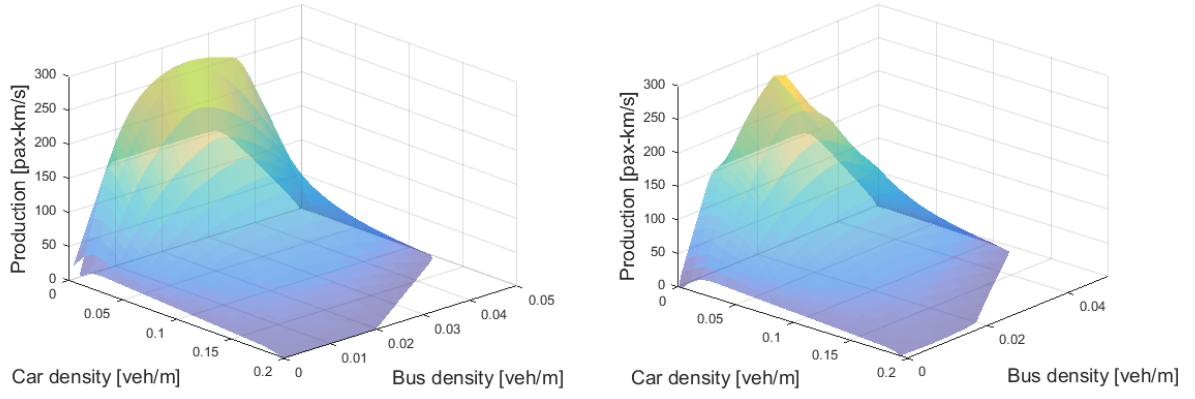
(a) Passenger capacity 3D-MFDs based on 3D-MFD average speed

(b) Passenger capacity 3D-MFDs based on individual mode's speed

Figure 6: Comparing passenger capacity 3D-MFDs. Car occupancy is assumed to be equal to two, while bus occupancy is assumed to be 80. The underlying 3D-MFD is identical to those shown in Figure 5.

speed $v_{c,MFD}$ and $v_{b,MFD}$.

Both passenger capacity 3D-MFDs clearly show a maximum at non zero density of buses. This is in accordance with the previous studies on the 3D-MFDs (Geroliminis *et al.*, 2014; Loder *et al.*, 2017). However, differences between both types of passenger capacity 3D-MFDs are visible. While the point of maximum travel production is similar, the peak is more distinct in the case of considering each modes' speed in Figure 6(b). This also leads to a stronger drop in production compared to the mean speed 3D-MFD in Figure 6(a).

5.2 Passenger-preference 3D-MFD

For the passenger preference 3D-MFD, we assume that h_b and h_c are a function of human preferences and a set of explanatory variables. In the context of the 3D-MFD, travelers make a discrete choice between buses and cars aiming to minimize the generalized cost for their trips. The generalized cost usually combines travel cost and travel time for both modes, as well as waiting time, access and egress time to the bus stop for the bus mode. The preferences then reflect the particular valuation of each of these elements. The list of elements above is not exhaustive as many other (personal) factors can influence mode choice as well.

For the generalized cost, we can obtain the travel time from the 3D-MFD itself. The

required bus network variables can be derived, for example, using the design approach by Daganzo (2010). Other preferences and the valuation of the cost elements require travel surveys. Using a multinomial logit (MNL) or similar function to translate all modes generalized cost into choice probabilities (Ben-Akiva and Lerman, 1985), allows us to determine what portion of the travelers choose the car or the bus (i.e. the modal split). In other words, this provides h_b and N_c, because h_c and N_b are usually fixed. The first one is fixed as cars are rarely occupied by more than 1-2 passengers. The latter is fixed because it is determined by the structure of the bus network.

To be more specific, let us consider the situation shown in Figure 7. We want to understand how for different levels of demand (in # pax) and different bus headways (in minutes) travelers make their mode choice. With this information, we can then determine the total passenger travel production accounting for preferences. We can clearly see that for very long headways, few travelers are transported at lower speeds compared to a situation with shorter headways. This is intuitive as shorter headways not only provide more capacity for travelers, but also lead to more attractive travel times in the network, and thus attract more passengers. Importantly, the parameters for human preferences are always context specific and must be carefully identified for each application. As a detailed discussion of this is out of scope of this paper and we only want to illustrate the idea of a passenger preference 3D-MFD based on our proposed functional form, we do not discuss details of the model any further.

5.3 Measured passenger 3D-MFD

Intuitively, the measured passenger 3D-MFD is the simplest to explain, but most likely the hardest to obtain. Here, h_c and h_b are measured, for example as Loder *et al.* (2017) did for Zurich. While one can make a fairly good assumption on h_c based on travel surveys, h_b requires extensive measurements. However, ticketing measures that require "*tapping in and out*" in buses provide a very good data source, but might be not available for many cities.
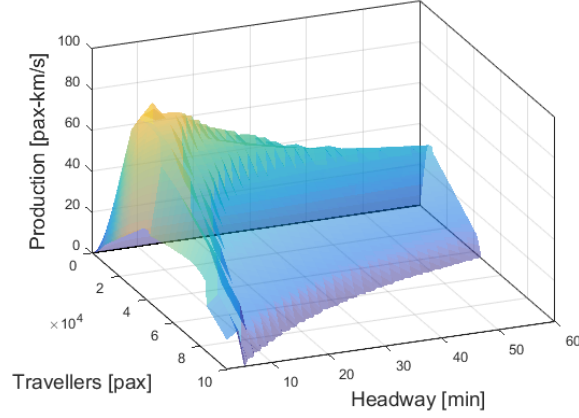
Figure 7: The passenger preference 3D-MFD in the dimensions of total number of travellers and design headway of the system. The underlying network is identical to those used in Figure 5. For the logit function, we choose a simple formulation for the generalized cost (utility) U and probability $\Pr = 1/\left(1 + \exp\left(-U\right)\right)$ of riding the bus with the commercial speed of buses (Eqn. 12), bus headway and vehicle crowding. The latter variable is the number of travellers divided by the maximum capacity of all buses in the network. Arguably, using congested travel times instead of the commercial speed of buses in the choice requires a traffic assignment based on Eqns. 19 and 20. We assumed typical values for the valuation of generalized cost elements. As the parameters for the valuation of the generalized cost elements differ from context to context and must be carefully identified for each application, we do not discuss this aspect here any further.

6 Comparison of the observed 3D-MFDs

Here, we first discuss in Section 6.1 how the functional form for the 3D-MFD can be constructed and estimated. Then, in Section 6.2 we validate the proposed functional form for the 3D-MFD with data obtained from a microscopic traffic simulator as well as empirical data from London and Zurich. Last, in Section 6.3 we validate the proposed speed functions from Section 4.

6.1 Derivation of the lower envelope and the smoothing parameter

The proposed functional form for the 3D-MFD must be estimated separately for each network by finding the values for each parameter listed in Table 1. Here, we have to distinguish between parameters that can be easily derived from the topology of the bus and road networks and the operational parameters of cars and buses.

First, network topology parameters L, η_c, η_b, l and p can be found using spatial data of the road and bus networks, e.g. from OpenStreetMap and the public transport agency. The network cutout for the 3D-MFD can be identified with network partitioning algorithms (e.g. Saeedmanesh and Geroliminis, 2016; Ambühl *et al.*, 2019). Then the network topology parameters can be calculated using spatial analysis tools. Importantly, η_b denotes the share of the total infrastructure devoted to buses; thus overlapping bus routes must not be counted multiple times, but only once.

Second, operational parameters might be obtained from prior traffic measurements, or approximated based on fundamental traffic principles. The jam spacing of cars l_c can be approximated using traffic departments or car clubs vehicle statistics, while the passenger car-equivalent φ can be calculated using the mean vehicle length of the entire fleet of the public transport agency. The average cycle length C and effective green time G can be observed on the real network or obtained directly from the traffic departments. For cars, detector and trajectory measurements can be used to estimate the fundamental diagram, including values for the cars free flow speed $v_{c,0}$ and backward wave speed $w_{c,0}$, as well as the saturation rate s_c. The average intersection delay δ_c can either be measured or calculated using the formula provided by Daganzo and Geroliminis (2008). For public transport, many agencies record the vehicle trajectories which allow to determine or approximate the fundamental diagram parameters $v_{b,0}$, $w_{b,0}$, and s_b, as well as the average dwell time Δ. The strategy parameter ζ must be either measured or derived from the bus-priority control algorithm.

Third, the smoothing parameter λ can be estimated from observations in at least two different ways: (i) If bus and car measurements are available and the empirical 3D-MFD can be estimated, λ can be obtained with a nonlinear regression minimizing the difference between the 3D-MFD from the calibrated lower envelope and the empirical 3D-MFD. (ii) If only the MFD from car measurements can be estimated, but we also have bus headway information and the desired commercial speed, then, Eqns. 1-4 provided by Daganzo (2010) can be used to approximate the number of buses and the bus travel production. This production can be added to the car MFD to obtain an approximation of the 3D-MFD. We illustrate this procedure in Section 6.2 using the Zurich data. With this approximated 3D-MFD we can calculate λ as before. In this case, however, we expect λ to be smaller, because the approximated 3D-MFD does not account for all interaction effects given that we use the scheduled and not the actual bus production. That being said, as we find in the next section that λ always falls within a certain range, the 3D-MFD for a given network can even be estimated without an observed (3D)-MFD, by assuming a λ value within the reported range in Table 3 scaled by network length (see Section 6.2).

6.2 Validation of the 3D-MFD

Using planes I-VII for the 3D-MFD, we now compare Eqn. 7 to the outcome of a microscopic traffic simulation and to two empirical data sets (see Appendix A for details). Here, we consider an average, i.e. constant, λ for the entire network and all vehicle accumulations. For the simulation, we use an abstract network with a 10×10 grid, with 180 links and the average block length of 150 m. For the simulation platform, we use a VISSIM microsimulation. Two network configurations are investigated: (i) a homogeneous network with identical links and no road hierarchy (Figure 8a); and (ii) a heterogeneous network with three levels of road hierarchy denoted as L1, L2, and L3 (Figure 8b). In the former configuration, each signalized intersection was modeled with a saturation flow of $s = 1800$ veh/h, cycle length of $C = 60$ sec, and $G = 30$ sec of green (including 3 sec of lost time) for all conflicting signal phases. In the latter configuration, we vary road capacity and signal timing parameters as follows: L1 has $s = 2000$ veh/h, $C = 80$ sec, and $G = 40$ sec; L2 has $s = 1800$ veh/h, $C = 70$ sec, and $G = 35$ sec; and L3 has $s = 1600$ veh/h, $C = 60$ sec, and $G = 30$ sec. For intersections between different capacity roads, the cycle and green time for the highest capacity road is used. Notice that this hierarchical network leads to more heterogeneous traffic conditions (Muhlich *et al.*, 2015). The tested traffic scenarios have public transport lines covering 20 % of the network length, where buses operate in a mixed-lane fashion, i.e. no dedicated lanes are allocated to public transport vehicles. For the empirical data sets, we use data collected from inductive loop detectors, providing vehicle flows and occupancy in London and Zurich. Figures 8c to d show the experimental sites in London and Zurich. The bus data is collected from the automated vehicle location devices (AVL), used to reconstruct the trajectories of vehicles and to estimate the averages of speed and density. The data from Zurich has been previously used by Loder *et al.* (2017) and Dakic and Menendez (2018). All required parameters for applying the proposed functional form on both simulation and empirical data sets are listed in Table 3, where those related to the network topology of London and Zurich are obtained from OpenStreetMap.

For all four data sets, we estimate λ with non-linear least squares. The estimated value for each 3D-MFD is given in Table 3. We show in Figure 9a to l the fitted functional form with the measurements given in red. Here we see that our proposed functional form aligns well with the observations for all four case studies. The third figure in each row in Figure 9 provides the residuals of the fit. We find residuals of around 30 % of the observed capacity for both simulation networks, with a slightly larger value for the heterogeneous network, and around 5-10 % for the empirical cases. Arguably, the residuals

Parameter	Simulation		London	Zurich
	Homogeneous	Heterogeneous		
Network length [km]	27	27	354	46
Dedicated bus network [-]	0	0	0.08	0.32
Dedicated car network [-]	0.8	0.8	0.69	0.46
Car length [km]	0.0065	0.0065	0.005	0.006
Bus-car length ratio [-]	3	3	3	4
Car free flow speed (w. intersection delay) [km/h]	30	30	28	27
Car backward wave speed (w. intersection delay) [km/h]	10	10	6	6
Free flow speed for buses (w/o intersection delay) [km/h]	25	25	21	22
Backward wave speed for buses [km/h]	8	8	5	5
Average saturation rate [veh/h-lane]	1800	1760	1600	1600
Average block length [km]	0.15	0.15	0.23	0.28
Average stop spacing [km]	0.45	0.45	0.33	0.35
Average cycle length [s]	60	72	62	55
Average effective green time [s]	30	36	19	16
Average dwell-time [s]	20	20	18	28
ζ [-]	1	1	0.4	0.1
δ_c [h]	0.006	0.006	0.006	0.006
λ [-]	0.585 (0.048)	0.661 (0.050)	7.81 (0.063)	1.15 (0.006)
Normalized λ [-]	0.022	0.024	0.022	0.025

Table 3: Network parameters for empirical validation of the 3D-MFD functional form. Standard error for λ in parenthesis.
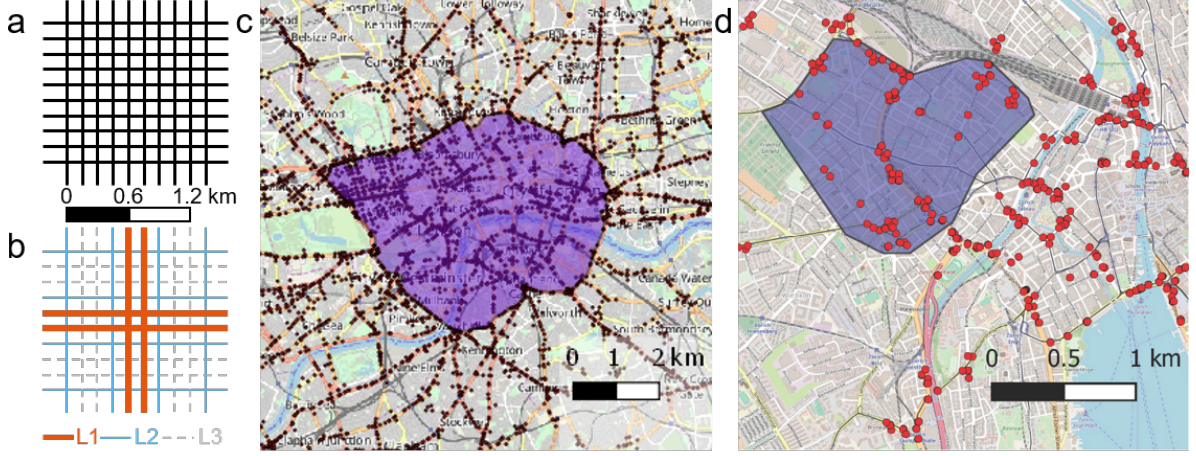
Figure 8: Network exhibits for the estimation of the 3D-MFDs. (**a**) shows the grid network in a homogeneous road configuration, while (**b**) shows a heterogeneous network with road hierarchy L1>L2>L3. (**c**) shows the neighborhood in London (UK) and (**d**) the neighborhood selected in Zurich (CH).

for the simulation are larger because the wider range of densities amplifies two effects not accounted for in the lower envelope: nonlinearities and network dynamics. When further comparing the estimated 3D-MFDs from both simulation networks, we find two different patterns. In the homogeneous network, the curve describes well the observed capacity and then overestimates the congested branch, while it underestimates the capacity and congested traffic states in the heterogeneous case. Arguably, this difference is caused by the difference in the two networks. With more heterogeneity, the similarity between the observed traffic states and the lower envelope decreases. Capturing this in our functional forms single parameter λ is difficult and consequently, the fit of the functional form in Figure 9 is less ideal in the heterogeneous case compared to the homogeneous case. Nevertheless, even in the case of a heterogeneous network, our proposed functional form stills provides a meaningful 3D-MFD shape.

For the value of λ, we normalize the accumulations to vehicle densities, while we keep the vertical axis in terms of production. As a result, and as previously noted by Ambühl *et al.* (2018), λ is not scale-invariant; hence it cannot be directly compared across networks. However, scaling λ by L brings λ to the order of magnitude of $\lambda \approx 10^{-2}$ in this study, which is in the same range reported by Ambühl *et al.* (2018) for the unimodal case. Therefore, even if no observations are available for the λ estimation, it is possible to use the previous finding of $\lambda \approx 10^{-2}$ (normalized for network length) to approximate it with reasonable certainty. We further conclude, that this holds even in heterogeneous network configurations as shown in Figure 8b. That being said, if the heterogeneity increases significantly, the network should be further partitioned to guarantee that the
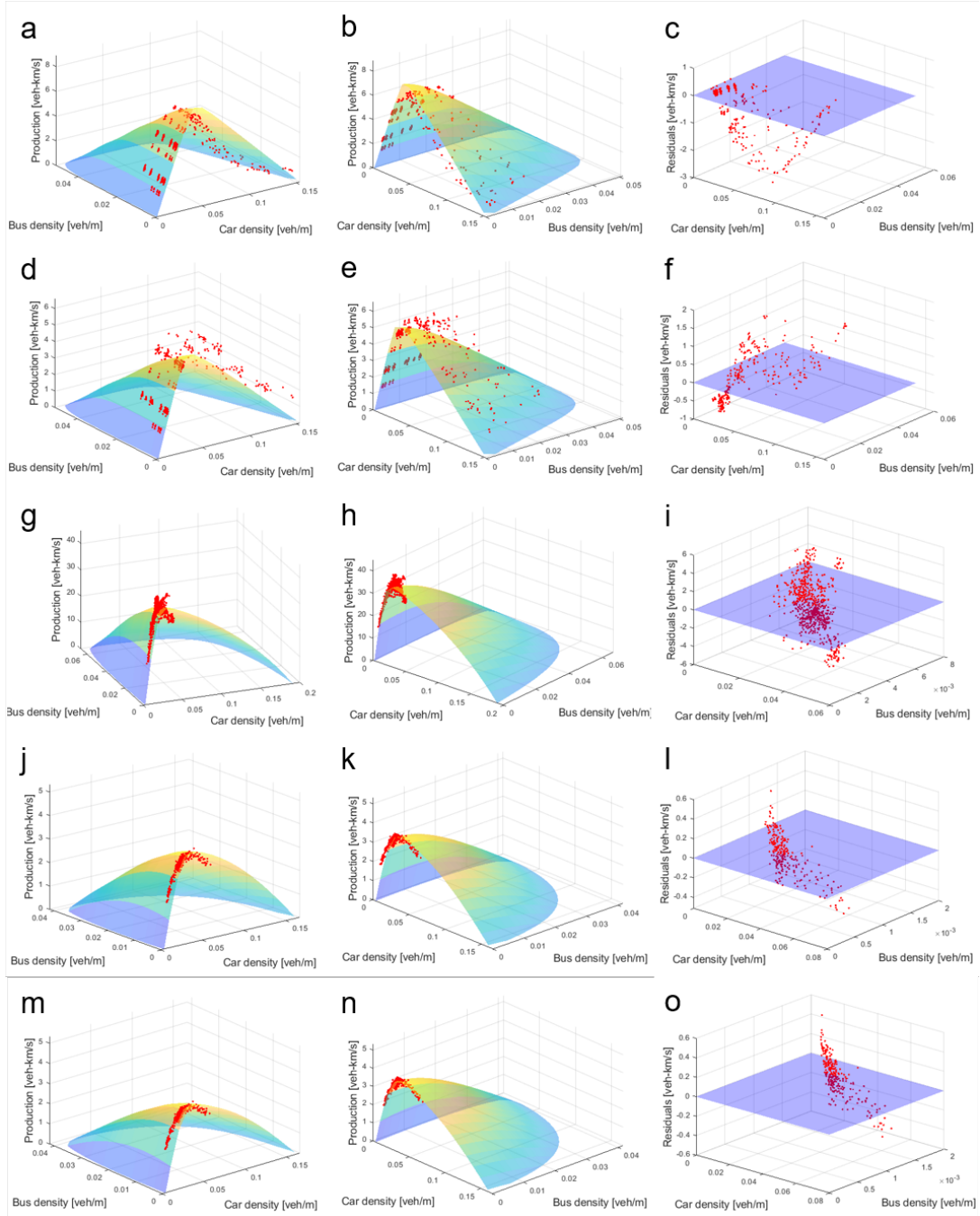
Figure 9: Validating the functional form for the 3D-MFD from simulation data for a homogeneous network (**a** to **c**), from simulation with a heterogeneous network (**d** to **f**), empirical traffic data from London (**g** to **i**), and empirical traffic data from Zurich (**j** to **o**). The first two columns show the estimated 3D-MFD from two different angles, while the third column shows the residuals of the λ estimation. For Zurich, **j** to **l** show the estimated 3D-MFD if traffic data from buses and cars is available, while **m** to **o** show the estimated 3D-MFD if no bus data is available and is approximated using Eqns. 22 and 23.

MFD homogeneity requirements are satisfied. Further research is required to investigate whether λ can be analytically derived as a function of different explanatory variables.

We now discuss the estimation of λ if no public transport measurements, like the ones we used for the validation of London and Zurich, are available. To do so, we follow Daganzo's 2010 idea and we use the Zurich case as an example. In this case, we require the total network length for buses B and the headway H. Using GIS tools, B can be calculated from spatial bus network data, which is usually available from the local agency. For our Zurich network, we have a total network length of $B = 43.6$ km. Note that this is more than the total infrastructure for buses in Table 3 as many routes overlap. According to the time table of Zurich, during peak hour, most services run with a headway of $H = 0.1$ h. Then, the production of buses during peak hour is given by Eqn. 22.

$$\pi_b = \frac{B}{H} = 436 \text{ veh-km h}^{-1} \tag{22}$$

With our observed data from Zurich, we find that during peak hour the maximum public transport production is $\pi_b \approx 470$ veh-km h^{-1}. This means an error of around $7\,\%$ with Eqn. 22 and the proposed approach; so we can conclude that this is a reasonable approximation.

We now need to calculate the number of buses during peak hour N_b with Eqn. 23 as used by Daganzo (2010).

$$N_b = \frac{\pi_b}{v_b} \tag{23}$$

Here, v_b is the commercial speed of buses, which can be either derived using Eqn. 12 from this paper, Eqn. 4 from Daganzo (2010), or based on measurements during peak hour. Suppose that we estimate v_b during peak hour at $v_b \approx 11$ km h^{-1} as Loder *et al.* (2017) reported. This then results in $N_b = 436/11 \approx 40$ veh, which is close to the observed $N_b \approx 43$ veh during peak hour (with an error of about $7\,\%$). Adding π_b and N_b to the observations of car traffic and then estimating the 3D-MFD function leads to $\lambda = 1.120$ with a standard error of 0.006 (normalized $\lambda \approx 0.024$). This value is very close to the λ estimated for the measured 3D-MFD (see Table 3). Consequently, the resulting

31

estimated 3D-MFD in Figures 9 m to o looks very similar to the 3D-MFD estimated using all empirical data in Figures 9 j to l.

We then compare our proposed functional form with Drakes generalized exponential function proposed by Geroliminis *et al.* (2014). The latter function has six model parameters that need to be fitted to the data and cannot be defined a priori, as they do not have any physical meaning. Our functional form, on the other hand, has a single parameter, λ, and as we already showed, it can be reasonably assumed given that it seems to be always within the same range when normalized for network length. Not surprising, we find a better fit of Drakes generalized exponential function to the data compared to our functional form. In particular, we first obtain for the simulation an RMSE of 0.323 for Drakes generalized exponential function and 0.770 for our proposed functional form for the homogeneous network, as well as 0.306 for Drake's function and 0.820 for our functional form in case of heterogeneous network configuration. This increase in RSME (error) of our proposed function is intuitive. The similarity between the theoretical best-case situation and the observed traffic states decreases with heterogeneity, which makes it difficult to describe with just a single parameter the observed traffic states. Second, for London we obtain an RMSE of 1.80 for Drakes generalized exponential function and 2.28 for our proposed functional form, although MATLABs fmincon solver exited for Drakes generalized exponential function with the status local minimum possible (see the resulting issue in Figure 10). Third, for Zurich we obtain an RMSE of 0.110 for Drakes generalized exponential function and 0.183 for our proposed functional form. However, Figure 10 emphasizes that it is difficult to obtain a satisfying 3D-MFD shape with Drakes generalized exponential function: Only with a full range of densies as exclusively available from a simulator, the 3D-MFD shape is satisfying. For empirical observations, the resulting 3D-MFD shapes are inevitably unsatisfying and consequently not applicable without concerns: As the parameter estimation for London converged only to a possible local optimum, the familiar 3D-MFD shape is clearly not recovered, although the proposed constraints are satisfied. For Zurich, we see that at least a shape similar to a 3D-MFD is recovered, but production decreases towards zero for higher bus accumulations faster than what is observed in reality, and production is non-zero at jam density of cars and zero accumulations of buses. None of these issues can arise with our proposed functional form, which in addition, can be potentially constructed from scratch without any traffic data, in contrast with Drake's exponential function that needs the full range of observations across the two modes.
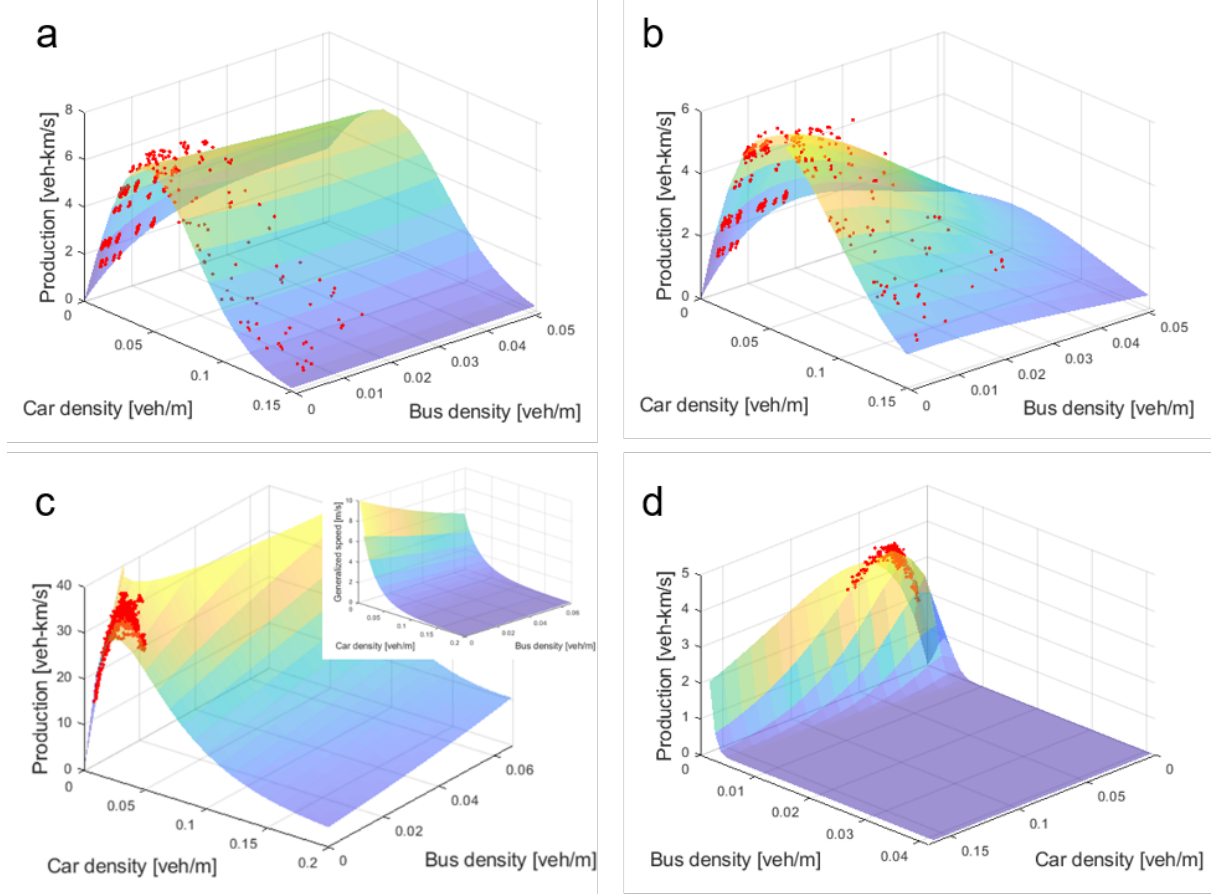
Figure 10: Estimation of Drake's exponential function as proposed by Geroliminis *et al.* (2014): (**a**) results from the simulation with the homogeneous network configuration, (**b**) results from the simulation with the heterogeneous network configuration, (**c**) shows the function estimated for London with the inset showing that the speed constraint is satisfied, and (**d**) shows the function estimated for Zurich.

6.3 Validation of speed functions

Here, we validate the approximation of the speed relationship between buses and cars proposed by Geroliminis *et al.* (2014) in Eqns. 14-16 as well as our proposed speed model in Eqns. 19 and 20. Figure 11 shows the bivariate scatter plots of the observed speeds of cars and buses for the four previously used data sets. Recall that these are the average speeds in the network, so they account for all delays including that from stopping at intersections and/or bus stops as well as congestion. We add to the scatter plots a linear fit (dashed lines) and our proposed speed model (solid line). Generally, we observe a linear trend between both modes speeds despite the scatter. Even though there are some differences between the proposed model and both the linear fit and the observations, the overlap is still significant. Therefore, we can reasonably assume that
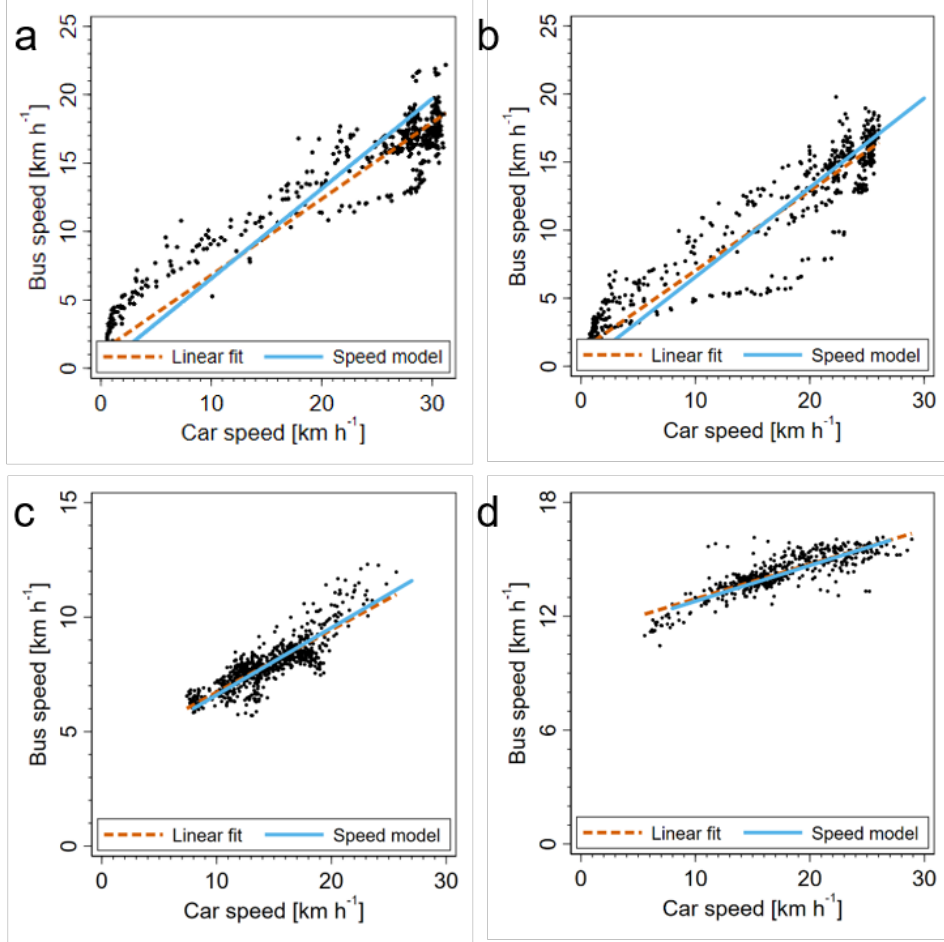
Figure 11: Validation of the proposed speed functions: (**a**) results from the simulation with the homogeneous network configuration, (**b**) results from the simulation with the heterogeneous network configuration, (**c**) shows the relationship for London, and (**d**) shows the relationship for Zurich.

the proposed linear relationship by Geroliminis *et al.* (2014) as well as our proposed speed model approximates the observed behavior relatively well.

7 Conclusions

This paper proposes a new general functional form for multi-modal MFDs from a geometric perspective with an detailed application to the 3D-MFD. The proposed functional form has two components. First, a lower envelope limits travel production to the theoretical best-case situation for any given vehicle accumulation. Second, a smoothing parameter quantifies how interactions between all vehicles types reduce travel production from the theoretical best-case.

For the 3D-MFD, the shape defining parameters are derived from the physical and operational characteristics of the multi-modal urban networks. The smoothing parameter can either be estimated from 3D-MFDs observed from traffic data or if such data is not available, our finding that the smoothing parameter is of similar value across networks allows a good approximation of its value. The proposed methodology is flexible to accommodate further city-specific characteristics in the traffic operations such as taxi and mini bus operations. MFD analysts can use the proposed functional form for the 3D-MFD to analyze changes in the urban resource allocation on the overall productivity of urban road networks. By doing so, this approach allows a rather quick determination whether a city is *full* and demands other means of transportation such as an underground system.

The proposed macroscopic approach of multi-modal urban transportation describes interactions under steady-state conditions, but lacks microscopic detail and does not consider dynamic aspects of urban traffic. Future research needs to investigate these issues. Moreover, the current definition of functions for the lower envelope considers only the main aspects of the analytical approaches to the MFD (Daganzo and Geroliminis, 2008; Leclercq *et al.*, 2014) and bus operations (Daganzo, 2010; Eichler and Daganzo, 2006), but accounting for more detailed aspects might improve the fit of the functional form. Further, considering the effects of bus network design (e.g. Daganzo, 2010; Saade *et al.*, 2018; Chen *et al.*, 2015) on the shape of the 3D-MFD would generate important contributions.

The discussed application of the functional form for the 3D-MFD, and the passenger 3D-MFD, emphasizes the practical and policy relevant implications of this research. First, it allows to understand multi-modal urban traffic even with few data requirements. Second, it allows to consider infrastructure variables as endogenous in optimization problems in a continuous representation of multi-modal urban traffic.

Acknowledgement

References

8 References

Ambühl, L., A. Loder, M. C. Bliemer, M. Menendez and K. W. Axhausen (2018) A functional form for the macroscopic fundamental diagram with a physical meaning, *Transportation Research Part B*, **(in press)**, 12–19, ISSN 0191-2615.

Ambühl, L., A. Loder, N. Zheng, K. W. Axhausen and M. Menendez (2019) Approximative network partitioning for MFDs from stationary sensor data, *Transportation Research Record: Journal of the Transportation Research Board*, **accepted**.

Amirgholy, M., M. Shahabi and H. O. Gao (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, ISSN 13665545.

Ampountolas, K., N. Zheng and N. Geroliminis (2017) Macroscopic modelling and robust control of bi-modal multi-region urban road networks, *Transportation Research Part B: Methodological*, **104**, 616–637, ISSN 01912615.

Ben-Akiva, M. E. and S. R. Lerman (1985) *Discrete choice analysis: theory and application to travel demand*, vol. 9, MIT press, ISBN 0262022176.

Bliemer, M. C. J., J. M. Rose and C. G. Chorus (2017) Detecting dominance in stated choice data and accounting for dominance-based scale differences in logit models, *Transportation Research Part B: Methodological*, **102**, 83–104, ISSN 01912615.

Boyaci, B. and N. Geroliminis (2011) Estimation of the network capacity for multimodal urban systems, *Procedia - Social and Behavioral Sciences*, **16**, 803–813.

Castrillon, F. and J. Laval (2018) Impact of buses on the macroscopic fundamental diagram of homogeneous arterial corridors, *Transportmetrica B: Transport Dynamics*, **6** (4) 286–301.

Chen, H., W. Gu, M. J. Cassidy and C. F. Daganzo (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, ISSN 01912615.

Chiabaut, N. (2015) Evaluation of a multimodal urban arterial: The passenger macroscopic fundamental diagram, *Transportation Research Part B: Methodological*, **81**, 410–420, ISSN 01912615.

Cook, J. (2011) *Basic properties of the soft maximum*, University of Texas, Houston.

Daganzo, C. F. (1994) The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory, *Transportation Research Part B: Methodological*, **28B** (4) 269–287.

Daganzo, C. F. (2007) Urban gridlock: Macroscopic modeling and mitigation approaches, *Transportation Research Part B: Methodological*, **41**, 49–62, ISSN 0191-2615.

Daganzo, C. F. (2010) Structure of competitive transit networks, *Transportation Research Part B: Methodological*, **44** (4) 434–446, ISSN 01912615.

Daganzo, C. F., V. V. Gayah and E. J. Gonzales (2011) Macroscopic relations of urban traffic variables: Bifurcations, multivaluedness and instability, *Transportation Research Part B: Methodological*, **45** (1) 278–288, ISSN 01912615.

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

Daganzo, C. F. and V. L. Knoop (2016) Traffic flow on pedestrianized streets, *Transportation Research Part B: Methodological*, **86**, 211–222, ISSN 01912615.

Daganzo, C. F., L. J. Lehe and J. Argote-Cabanero (2017) Adaptive offsets for signalized streets, *Transportation Research Part B: Methodological*, 1–9.

Dakic, I., L. Ambühl, O. Schümperlin and M. Menendez (2019) On the modeling of passenger mobility for stochastic bi-modal urban corridors, paper presented at the *23rd International Symposium on Transportation and Traffic Theory*.

Dakic, I. and M. Menendez (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, ISSN 0968090X.

Eichler, M. and C. F. Daganzo (2006) Bus lanes with intermittent priority: Strategy formulae and an evaluation, *Transportation Research Part B: Methodological*, **40** (9) 731–744, ISSN 01912615.

Gayah, V. V. and C. F. Daganzo (2011) Clockwise hysteresis loops in the macroscopic fundamental diagram: An effect of network instability, *Transportation Research Part B: Methodological*, **45**, 643–655, ISSN 0191-2615.

Geroliminis, N. and C. F. Daganzo (2008) Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings, *Transportation Research Part B: Methodological*, **42**, 759–770.

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

Haddad, J. and N. Geroliminis (2012) On the stability of traffic perimeter control in two-region urban cities, *Transportation Research Part B: Methodological*, **46**, 1159–1176.

He, H., M. Menendez and S. Ilgin Guler (2018) Analytical evaluation of flexible-sharing strategies on multimodal arterials, *Transportation Research Part A: Policy and Practice*, **114**, 364–379, ISSN 09658564.

Ji, Y. and N. Geroliminis (2012) On the spatial partitioning of urban transportation networks, *Transportation Research Part B: Methodological*, **46** (10) 1639–1656, ISSN 01912615.

Leclercq, L., N. Chiabaut and B. Trinquier (2014) Macroscopic fundamental diagrams: A cross-comparison of estimation methods, *Transportation Research Part B: Methodological*, **62**, 1–12.

Loder, A., L. Ambühl, M. Menendez and K. W. Axhausen (2017) Empirics of multimodal traffic networks Using the 3D macroscopic fundamental diagram, *Transportation Research Part C: Emerging Technologies*, **82**, 88–101, ISSN 0968090X.

Muhlich, N., V. V. Gayah and M. Menendez (2015) An examination of MFD hysteresis

patterns for hierarchical urban street networks Using micro-simulation, *Transportation Research Record: Journal of the Transportation Research Board*, **2491**, 117–126.

Saade, N., J. Doig and M. J. Cassidy (2018) Scheduling lane conversions for bus use on city-wide scales and in time-varying congested traffic, *Transportation Research Part C: Emerging Technologies*, **95**, 248–260, 10 2018, ISSN 0968-090X.

Saeedmanesh, M. and N. Geroliminis (2016) Clustering of heterogeneous networks with directional flows based on Snake similarities, *Transportation Research Part B: Methodological*, **91**, 250–269, ISSN 01912615.

Smeed, R. J. (1968) Traffic studies and urban congestion, *Journal of Transport Economics and Policy*, **2**, 33–70.

Zheng, N., T. Dantsuji, P. Wang and N. Geroliminis (2017) Macroscopic Approach for Optimizing Road Space Allocation of Bus Lanes in Multimodal Urban Networks Through Simulation Analysis, *Transportation Research Record: Journal of the Transportation Research Board*, **2651**, 42–51, ISSN 0361-1981.

Zheng, N. and N. Geroliminis (2013) On the distribution of urban road space for multi-modal congested networks, *Transportation Research Part B: Methodological*, **57**, 326–341.

A Empirical data preparation

The bus and car traffic data for London and Zurich come from different kind of sources. The bus data comes from *Automatic vehicle location* (AVL) systems that record buses' locations with a time-stamp that allows to generate their trajectories in the network. From these trajectories, we calculate in each analysis interval T the distance traveled d_i and travel time t_i within that interval T and the analyzed region. With these measurements we are able to obtain the measurements of space-mean speed of buses v_b by evaluating Eqn. 24.

$$v_{b,MFD} = \frac{\sum d_i(t)}{\sum t_i(t)} \tag{24}$$

Additionally, we derive the total accumulation of buses with Eqn. 25.

$$N_b(t) = \frac{\sum t_i(t)}{T} \tag{25}$$

Then, we can calculate the total travel production of buses with Eqn. 13.

$$\Pi_b(t) = N_b(t)\, v_b(t) \tag{26}$$

The car traffic data comes from inductive stationary sensors in the road network. During an observation interval T, each detector i reports the count of vehicles, q_i, as well as the fraction of time, where the detector was physically occupied with vehicles, o_i. Commonly, occupancy is understand as an indicator for traffic density. We then obtain the network average for the MFD by calculating the weighted averages of flow and occupancy weighted by the link length l_i of each monitored lane according to Eqns. 27 and 28.

$$q_{c,MFD}(t) = \frac{\sum l_i q_i(t)}{\sum l_i} \tag{27}$$

$$o_{MFD}(t) = \frac{\sum l_i o_i(t)}{\sum l_i} \tag{28}$$

We then calibrate our MFDs, in particular traffic density, by using refernce space-mean speeds in the same regions obtained from the google directions API. In detail, we identify the on average hour in the MFD and the network, typically in the early morning, and request 1000 random directions from the API at the same hour of the day. We then ensure that the means of the direction API speeds match with the MFD speeds from the same hour of the day.