DISS. ETH NO. 26064

OPTIMAL TRANSPORT INVESTMENT AND PRICING IN A MULTIMODAL CITY

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by

ALLISTER LODER
Master of Science ETH in Energy Science and Technology

born on 28.07.1989
citizen of Germany

accepted on the recommendation of

Prof. Dr. Kay W. Axhausen, examiner
Prof. Dr. Thomas F. Rutherford, co-examiner

2019

For the curious.

ABSTRACT

People's mobility behavior is often very ambivalent and selfish: Everyone wants to drive alone in his own vehicle, while at the same time wanting no one else to demand the same, as this would lead to congestion. In cities, transportation alternatives to the car exist and can be competitive in terms of travel times. However, transport policies that favor the car too much will create too many undesirable negative externalities at urban scale, for example, congestion and pollution. In the end, such a system will not provide the optimal level of mobility for everyone and the agglomeration economies that improve productivity. This brings us to the overarching question of this thesis: How should cities invest in infrastructure and price transportation to maximize everyone's mobility and productivity gains?

To this end, this thesis makes several methodological contributions. First, I investigate how individuals' mobility tool ownership (e.g., cars and public transport season-tickets) and travel activity (trip frequency) choices are a function of accessibility, a frequently used measure to analyze the level of mobility. I propose a methodology that captures multimodal aspects of accessibility necessary to understand multimodal travel behavior. I use this novel accessibility measure in a framework recently proposed by Chandra Bhat to model multi-dimensional choice environments. I find that our proposed multimodal accessibility measure is a strong predictor of mobility tool ownership as well as travel activity.

Second, considering urban traffic as a simple system, I contribute two elements: (1) With the first large-scale empirical comparison of urban-scale traffic of many cities, I provide evidence that such a system perspective is legitimate, as macroscopic urban traffic follows a deterministic relationship - the macroscopic fundamental diagram (MFD) - which is indeed a function of the underlying network's topology. (2) Such a system perspective requires a physically consistent, but simple multimodal congestion mechanism. With this thesis, I propose three different mechanisms to model the congestion effects of bus and car interactions, using the multimodal MFD or 3D-MFD: an empirical approach, a geometric approach and a Bose-Einstein condensate approach.

Third, I use the geometric approach to estimate the 3D-MFD to develop a static traffic assignment in a mixed complementarity problem formulation. With this assignment procedure, I propose two approaches to identify op-

timal pricing and investment decisions. One approach considers only the effects within the transportation system, with a network design problem, the other approach considers individuals' choices in an integrated model of urban housing, labor, and transportation markets, with a computational economic equilibrium model.

In closing, this thesis provides a macroscopic perspective on the question of how many cars are too many for a city and allows cities to identify their strategies subject to their policy constraints to improve traffic and mobility for everyone. As transport systems are sensitive to political decisions and public debate, this thesis provides approaches to derive quantitative arguments that may help to identify the optimal transportation policy for a city.

ZUSAMMENFASSUNG

Das menschliche Mobilitätsverhalten ist oft egoistisch und ambivalent: Jeder möchte am liebsten alleine in seinem Auto fahren, und gleichzeitig soll niemand sonst Auto fahren, damit kein Stau entsteht. In Städten gibt es üblicherweise Alternativen zum motorisierten Individualverkehr, welche ähnlich attraktive Reisezeiten ermöglichen können, jedoch kann eine Verkehrspolitik, welche zu stark das Auto bevorteilt, zu viele ungewollte negative Externalitäten wie Stau oder Umweltbelastungen hervorrufen. Schlussendlich wird so ein Verkehrssystem nicht die optimale Mobilität für jeden Einwohner bereitstellen und den ökonomischen Agglomerationsprozess nicht bestmöglich unterstützen, um Produktivitätsgewinne zu erzielen. Dieses Problem bringt uns zu der übergeordneten Frage dieser Doktorarbeit: Wie müssen Städte in Infrastruktur investieren und die Preise für Mobilität festsetzen, um Mobilität und Produktivitätsgewinne zu maximieren?

In diesem Zusammenhang macht diese Arbeit mehrere Beiträge. Erstens, erarbeiten wir wie der Mobilitätswerkzeugbesitz (Auto und Zeitkarte) und Reiseverhalten (Anzahl an Wegen) als Funktion von Erreichbarkeit dargestellt werden kann. Erreichbarkeit ist dabei ein häufig verwendetes Mass für das Level an Mobilität, welches das Netz bereitstellt. In dieser Arbeit schlagen wir eine Methode vor, um in der Betrachtung von Erreichbarkeit multimodale Aspekte abbilden zu können. Wir verwenden dieses neue multimodale Erreichbarkeitsmass in einem von Chandra Bhat vorgeschlagenen ökonometrischen Modell, um Mobilitätswerkzeugbesitz und Reiseverhalten zu erklären. Dabei finden wir, dass das vorgeschlagene Erreichbarkeitsmass eine sehr starke erklärende Variable für Mobilitätswerkzeugbesitz und Reiseverhalten ist.

Zweitens, in der Betrachtung von städtischem Verkehr als einfaches System, leistet diese Arbeit zwei wesentliche Beiträge: (i) mit dem ersten grossformatigen empirischen Vergleich von Verkehr in vielen Städten zeigen wir auf, dass eine einfache Systemperspektive mit dem makroskopischen Fundamentaldiagram (MFD) ein legitimer Ansatz ist, da wir für das MFD deterministischen Zusammenhängen mit der Systemkonfiguration identifizieren. (ii) eine solche Systemperspektive verlangt für den städtischen Verkehr auch einen physikalisch konsistenten multimodalen Staumechanismus. Diese Arbeit leistet dazu mit drei verschiedenen Ansätzen einen

Beitrag: einen empirischen datengetriebenen Ansatz, einen geometrischen Ansatz und einen Ansatz basierend auf der Bose-Einstein Kondensation.

Drittens, wir verwenden den geometrischen Ansatz für das 3D-MFD für den multimodalen Staumechanismus, um eine multimodale Verkehrsumlegung im mixed complementarity problem Format zu entwickeln. Basierend auf dieser Umlegung schlagen wir zwei Ansätze vor, um optimale Investitions- und Preisentscheidungen zu machen. Der erste Ansatz analysiert die Effekte nur innerhalb des Transportsystems und stellt die Frage nach der optimalen Netzkonfiguration von motorisierten Individualverkehr und dem öffentlichen Verkehr. Der zweite Ansatz analysiert in einem ökonomischen Gleichgewichtsmodell Preis- und Investitionseffekte im Transportsystem auf Produktivitätsgewinne.

Diese Thesis bietet eine makroskopische Perspektive auf die Frage «wie viele Autos sind genug für eine Stadt?». Die Beiträge dieser Arbeit erlauben es Städten, Strategien unter Berücksichtigung ihrer politischen Rahmenbedingungen abzuleiten, um optimale Mobilität für alle und Produktivitätsgewinne zu realisieren. Da viele Aspekte des Verkehrssystems sensitiv auf politische Entscheidungen sind, können die Beiträge dieser Doktorarbeit dabei helfen, quantitative Argumente zu erarbeiten, die schlussendlich dabei helfen, ein optimales und nachhaltiges Verkehrssystem zu schaffen.

ACKNOWLEDGEMENTS

Most likely, many will only read until this part of my thesis, which makes this section an important conclusion to my time at the institute. This section provides me an opportunity to express my gratitude to all who contributed and shaped thesis during almost four years of dedicated and cooperative research. Without their ideas, thoughts, discussions and feedback, this thesis would certainly read completely different.

First and foremost, I am especially indebted to my promoter and supervisor Kay W. Axhausen who gave me the opportunity and resources to commence this research. In particular, I am thankful to had plenty degrees of freedom to pursue my research and follow my own ideas as well as having the chance to travel to meet many inspiring people.

Second, during my PhD, I had the pleasure of many encounters with other researchers whose enduring support definitely shaped the contributions of my thesis, if not their thoughts made some of them even possible. Mostly, I have to thank Monica Menendez for her encouraging and engaging collaboration, discussions and feedback as well as her dedication in scrutinizing every written word of mine. I would like to thank my cosupervisor Thomas F. Rutherford for his curiosity and ingenuity in shaping and boosting our research. Then, from the other side of the globe, I want to thank Michiel C.J. Bliemer deeply for the creative and always methodologically accurate collaboration. I also have to thank Andrew Schreiber, Nan Zheng and Renger van Nieuwkoop for their support in shaping this thesis. Further, I wish to thank all members of the Institute for Transport Planning and Systems, especially Felix and Henrik Becker as well as Basil Schmid, for their advice, support and discussions. I could always rely on their expertise which advanced so many procedures and accelerated ideas beyond this thesis.

Third, I had the pleasure to work together on our MFD endeavors with my colleague Lukas Ambühl, who truly became a friend over the past years. Somehow, the two of us are perfect examples that creativity and output scales superlinear with the number of people involved. Without our interactions, without your sense of humor and your Bäääärner relaxedness, our research and everything around would certainly resulted in less productive outcomes. I am very grateful to had the opportunity to work and share time with you. Thank you.

Last, but – certainly – not least, I want to thank my parents, my family and friends for continuously supporting me in everything I did and hopefully will do. I always knew that plenty of places exist where I can feel at home. I dedicate this thesis to you.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOTATION

INDICES AND SUBSCRIPTS

SYMBOL	MEANING
i, j, k	Nodes and zones
r	Route
m	Mode
t	Mobility tool portfolio
c	Subscript for car
b	Subscript for public transport
v	Subscript for bicycles (velo)

FREQUENTLY USED SYMBOLS

SYMBOL	MEANING
A	Vehicle accumulation
B	Bus infrastructure length
C	Traffic signal cycle length
G	Green time
H	Bus headway
L	Network length
M	Minimum path costs
N	Flow of passengers
Q	Shares of mobility tool ownership
R	Red time
T	Travel time
k	Vehicle density
l	Intersection spacing
n	Two-fluid theory parameter
p	Bus stop spacing
s	Intersection saturation flow

q	Vehicle flow
v	Speed
w	Backward wave speed
Π	Total network travel production
Γ	Additional delay
Λ	Total demand
α	Bus network design parameter
η	Share of dedicated road space
ϑ	Fundamental diagram speed
κ	Jam spacing
λ	Network interaction parameter
μ	Elasticity
π	Prices
ζ	Dwell time
φ	Passenger car equivalent
∇	Public transport strategy

AUTHOR CONTRIBUTION STATEMENT

Chapter 3 of this thesis is a re-formatted and shortened version of Loder and Axhausen (2018). Kay W. Axhausen and Allister Loder had the idea for the analysis and wrote the manuscript, while Allister Loder did the data collection and analysis as well as wrote the code.

Chapter 4 contains several contributions from the collaboration with Lukas Ambühl, Kay W. Axhausen and Monica Menendez (Section 4.3, Section 4.4.1, and Section 4.4.2). In detail, Section 4.3. is based on a revised, shortened and re-formatted manuscript currently under review at Scientific Reviews. Lukas Ambühl and Allister Loder share first authorship and contributed equally in all stages of the data collection, method development, and manuscript preparation. Kay W. Axhausen and Monica Menendez had the idea for research design. All four authors contributed to the idea as well as wrote and revised the manuscript.

Section 4.4.1 is based on a revised and re-formatted version of Loder *et al.* (2017). Lukas Ambühl, Allister Loder and Monica Menendez had the idea for the method. Lukas Ambühl, Kay W. Axhausen, Allister Loder and Monica Menendez developed the design of research. Lukas Ambühl and Allister Loder did the data collection and analysis. Lukas Ambühl and Allister Loder did the code and method development. All authors wrote and revised the manuscript.

Section 4.4.2 is based on a re-formatted version of Loder *et al.* (2019). Allister Loder had the idea for the manuscript. Kay W. Axhausen, Michiel C.J. Bliemer, Allister Loder and Monica Menendez developed the method and research design. Lukas Ambühl, Igor Dakic and Allister Loder did the data collection. Lea Bressan and Allister Loder developed the code. All authors wrote and revised the manuscript.

Section 4.4.3 is based on a manuscript submitted to Transportation Research Part B: Methodological, written in collaboration with Kay W. Axhausen, Henrik Becker, Lea Bressan, Andy Emmonds, Martin Obee, Victor Knoop, Allister Loder, Monica Menendez, and Marie-Jette Wierbos. Henrik Becker and Allister Loder developed the idea of the study, and developed the research design and strategy for data collection. Allister Loder developed methodological foundations of the paper. Victor Knoop, Allister Loder, Monica Menendez and Marie-Jette Wierbos developed the method further. Lea Bressan and Allister Loder developed the code. The

strategy for data collection and processing involved Henrik Becker, Andy Emmonds, Martin Obee, Victor Knoop, Allister Loder, and Marie-Jette Wierbos. The initial manuscript has been written by Allister Loder and all authors revised the manuscript.

Chapter 5 of this thesis is a re-formulated part of the manuscripts presented in Chapter 6 and 7. Allister Loder developed the initial idea. Michiel C.J. Bliemer and Allister Loder developed the method.

Chapter 6 of this thesis is based on a conference contribution for the 10th Triennial Symposium on Transportation Analysis (TRISTAN) in collaboration with Kay W. Axhausen and Michiel C.J. Bliemer. All authors contributed to the idea as well as wrote and revised the manuscript. Michiel C.J. Bliemer and Allister Loder did the method development; and Allister Loder did the code development and analysis.

Chapter 7 of this thesis is based on a manuscript written in collaboration with Kay W. Axhausen, Thomas F. Rutherford, and Andrew Schreiber targeted towards Regional Science and Urban Economics. Kay W. Axhausen and Thomas F. Rutherford developed the idea. Allister Loder, Thomas F. Rutherford, and Andrew developed the code. Allister Loder did the data collection and model calibration. All authors wrote and revised the manuscript.

1

INTRODUCTION

Es irrt der Mensch so lang er strebt.

— Goethe (1808)

We build our transportation infrastructure for connecting people, exchanging goods and facilitating the transfer of ideas. The resulting *interactions* between people, traders and companies produce positive externalities for societies (Smith, 1776; Ricardo, 1817; Krugman, 1991; Fujita *et al.*, 2001). However, negative externalities (e.g., congestion and pollution) usually increase with the number of travelers and the tipping point of the system is usually considered the infrastructure's capacity: when loaded beyond capacity, the collective flow of travelers no longer increases with the number of travelers, but instead decreases (Schaar, 1925; Greenshields, 1935; Daganzo, 1997).

The positive externalities as well as negative externalities of human interactions can be seen as *forces* that together should balance, such that no traveler would benefit further from engaging in more interactions because the benefits would be consumed by more congestion externalities. It is not surprising then, that the question of an *optimal* equilibrium between these two forces is recurring and timeless (e.g., Vickrey, 1969; Small and Verhoef, 2007). Arguably, in a transport system, this equilibrium can be influenced and shifted to a more desired social optimum by, for example, investment, pricing and new technologies (e.g., Stopher, 2004; Prud'homme and Bocarejo, 2005; Arnott and Small, 1994).

Nevertheless, in the observed equilibrium, these two forces are of course not the only determinants. There is substantial evidence of at least two *universal laws* governing human behavior and choices influencing the equilibrium as well. Broadly speaking, the driving force beyond each of them is the density of people: When population density increases, human interactions and economic output scale in a superlinear manner, resulting from agglomeration economies and increased productivity (e.g., Fujita and Thisse, 2002, 2013; Melo *et al.*, 2009; Duranton and Puga, 2004; Bettencourt, 2013), and people tend to use transit more (e.g., Cervero and Kockelman, 1997; Ewing and Cervero, 2001, 2010). There is another density effect, namely, that many supply-side networks scale in a sublinear manner (Kühnert

et al., 2006; Samaniego and Moses, 2008), that is, in the economic terms economics of scale or density. The idea of an equilibrium in our complex cities might be unrealistic, but empirical observations show recurring and predictable mobility patterns (González *et al.*, 2008; Çolak *et al.*, 2016; Wang *et al.*, 2012; Lopez *et al.*, 2017) that somehow emphasize that individuals have no incentive to change their behavior, as if they were in an equilibrium.

All these forces *push* and *pull* the decisions of firms as well as households (Fujita and Thisse, 2002). When cities invest in infrastructure and adjust pricing, people will reconsider their choices and adopt their behavior. These reactions to policies have shaped our cities (Brueckner and Fansler, 1983; Louf and Barthelemy, 2014; Batty, 2008). One of the most notable reactions is summarized as the *fundamental law of road congestion* (Downs, 1962). Worldwide empirical evidence suggests that vehicle mileage increases with added road capacity, with a short- and long-run elasticity of below one (Goodwin, 1996; Cervero and Hansen, 2002; Graham and Glaister, 2004; Cervero, 2002, 2003; Weis and Axhausen, 2009), and more recently, close to one (Duranton and Turner, 2011; Hsu and Zhang, 2014). In either case, these elasticity values imply that the driving grows slower than the infrastructure, resulting in less congestion - at least in the short term.

Although the above mentioned forces and transport policies may have created the problems we currently attribute to our transportation system, they provide at the same time opportunities for solving these problems. However, with ongoing and expected further urbanization (Department of Economic and Social Affairs of the United Nations Secretariat, 2014), the pressure on transportation systems will increase (Schafer and Victor, 2000; Duranton and Turner, 2012). Consequently, the pressure to identify optimal transport policies to achieve the optimal balance of the mentioned forces is increasing and will be a key challenge of the 21st century. This thesis contributes with novel approaches to understand the interaction of forces in the equilibrium and to derive strategies for optimal policies.

In Figure 1.1 I illustrate the overarching perspective of this thesis. We generally consider that we can abstract the urban transportation system into the system shown in Figure 1.1**a**. Such a perspective is common in many engineering sciences such as mechanics or signal theory, while in transportation there have been a few early attempts (e.g., Smeed, 1968; Herman and Prigogine, 1979), but only the recent advances by Daganzo (2007) and Geroliminis and Daganzo (2008) follow up on this idea with a physically consistent system perspective using the *macroscopic fundamental*

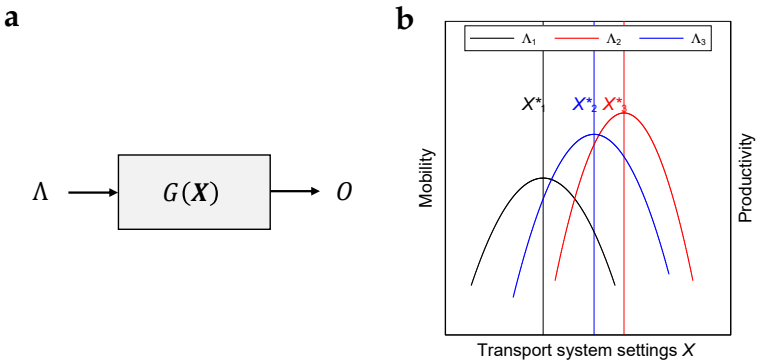


FIGURE 1.1: The overarching perspective of this thesis on the complex urban transportation systems. (**a**) illustrates the engineering system perspective with elements demand Λ, system settings $\mathbf{X}$, systems outflow (arrivals) O, all linked by function G. (**b**) summarizes the objective as finding $\mathbf{X}^*$ that maximizes mobility and or productivity given demand Λ.

diagram (MFD). The system has a characteristic function G that translates the demand Λ, depending on the system's parameter settings $\mathbf{X}$ into arrivals or system outflow O. $\mathbf{X}$ can describe a wide range of information, from the size and design of the network, to the pricing strategy and transport policies.

In urban transport policy making, the overall collective objective should be to improve mobility for everyone (Daganzo, 2007). This social optimum perspective is hard to identify due to the complexity of cities. However, among the key concepts of transport planning and politics is *accessibility*, which allows a quantitative and spatial assessment of mobility with just a single metric (Hansen, 1959). Further, accessibility also links to economic activity, and even is related to agglomeration economies (Vickerman *et al.*, 1999; Melo *et al.*, 2017; Weiss *et al.*, 2018). Thus, accessibility is not only a measure of the level of mobility the urban transportation system provides, but also an indicator of the output for which we, basically, build our infrastructure. Finally, bringing together the system perspective from Figure 1.1**a**, the overall collective objective of improving mobility for everyone and the concept of accessibility result in the overarching question of this thesis, illustrated in Figure 1.1**b**: Which system parameters $\mathbf{X}^*$ for given demand Λ

maximize mobility (accessibility), and then the agglomeration economies that increase productivity?

This link between congestion, accessibility, and agglomeration has been previously investigated (e.g., Zhang and Kockelman, 2016; Venables, 2007), but those approaches lack a multimodal congestion mechanism, in other words, a way to capture the interaction of cars and buses (and other modes), and a mechanism that can capture multimodal investment effects. Such a mechanism is important for analyzing the allocation of scarce urban resources. Thus, closing this gap with such a congestion mechanism is particularly important for improving mobility in our cities, as relying on cars alone for urban transportation is a blind alley. Further, our overall question is only one aspect in a wider perspective on the success of cities. Bettencourt (2013) proposes that cities exist when the balance of social benefits and costs is positive, most likely to have an optimum at a certain levels of infrastructure provision and economic output. Venables (2017a) pursues the question of which infrastructure and economic factors lead a city to produce internationally tradable goods that are potentially subject to increasing returns to scale. In other words, can we invest in infrastructure so that it increases productivity (Aschauer, 1989; Chatman and Noland, 2011; Rokicki and Stępniak, 2018) and "*spur[s] local economies*" (Hymel, 2009), that is economic growth? Consequently, the contributions of this thesis are helpful not only for transport policy decisions, but also for regions and cities in the global south (see Section 8.2.4).

As cities are complex, the function G as well as $\mathbf{X}$ can arguably be complex, too, if we do not focus on the most essential underlying dimensions of the transport system. Referring to the introduced forces mentioned at the beginning, I therefore focus only on congestion externalities for G and on two policy *levers* for $\mathbf{X}$: investment decisions in infrastructure and pricing.

The objective from Figure 1.1**b** is of course to simplify, and so it is necessary to break down this perspective into a sequence of research questions. First, the perspective introduced in Figure 1.1**a** suggests a fully mechanical system, which is of course the case with simply a vehicle perspective (e.g., Nagel and Schreckenberg, 1992; Daganzo, 1994), but from a traveler's perspective the system has to consider the choices people make that can most of the time, be only partially explained: where to go, which mode to use, which route to take, when to depart. Consequently, the first research question relates travel choices and accessibility:

1. How do people choose their multimodal travel behavior as a function of accessibility?

Second, determining accessibility of the road network is a well understood exercise (e.g., Geurs and van Wee, 2004; Handy and Niemeier, 1997; Morris *et al.*, 1979), but for the chosen system perspective, we wish to avoid modeling traffic on every link, and instead to abstract traffic into a system: the MFD (Daganzo, 2007; Geroliminis and Daganzo, 2008). However, the MFD so far lacks a thorough consideration of interaction with other modes. Consequently, we need a simple approach to model multimodal traffic with the MFD:

2. How can one model multimodal traffic in cities as a simple system?

Third and last, we then need to find a way to identify $\mathbf{X}^*$ from Figure 1.1**b**. Recall that $\mathbf{X}$ considers pricing and investment decisions. This encompasses not only aspects of finding $\mathbf{X}^*$ from a technical perspective, but also how it can be embedded in wider models that also account for nontraffic externalities.

In theory, the question of optimal investment and pricing has already been widely discussed (Pigou, 1920; Vickrey, 1969; Small and Verhoef, 2007), and when asked, many theoretical researchers would argue it has been solved. In reality, however, the theoretical marginal social cost pricing is impossible to achieve due to multiple price components, various (competing) agencies involved as well as the spatial and temporal heterogeneity of the transportation market. Consequently, we require a framework that reflects these aspects in the search for $\mathbf{X}^*$

3. How can one identify optimal prices and infrastructure provision with the system perspective of improving mobility and productivity?

Generally, it should be clear that there is no unique answer to the question of optimal investment and pricing in a city as each city has a different topology, various existing transport systems, different choice behavior of residents, and overall divergent demand levels, as shown in Figure 1.1**b**. However, the work presented in this thesis provides methods to identify the strategies for each city separately.

1.1 THESIS CONTRIBUTIONS

The contributions of this thesis are mainly in the methodological dimensions, because deriving strategies for optimal investment and pricing policies is context dependent and the strategies vary from city to city. Nevertheless, in this thesis, I provide numerical examples for London and Zürich and comment on how cities can use the proposed methods.

The particular contributions of this thesis are at least fourfold. First, I propose an approach to quantify multimodal accessibility as an indicator for describing travel behavior. Second, I present substantial empirical evidence for the existence and physical properties of the MFD, which implies that the urban road network can indeed be simplified into a system, as given in Figure 1.1**a**. Third, I present three multimodal congestion mechanisms for the MFD that capture the interactions between cars and buses. This allows modeling of passenger choices and vehicle interactions at the same time in the system described in Figure 1.1**a**. Fourth, I introduce a static multimodal network assignment based on the MFD. Its simple mathematical formulation underlines the required system perspective that I then illustrate in two macroscopic equilibrium models to identify optimal investment and pricing in a multimodal city.

This thesis has a strong emphasis on modeling traffic and transport issues in Switzerland; however, in the discussion in Section 8.2.4 I discuss the applicability and implications for developing countries. I further discuss the implications of this thesis for the expected large-scale vehicle automation in the upcoming decades, in Section 8.2.4.

1.2 OUTLINE OF THE THESIS

Chapter 2 provides an overview of the literature of the four overarching forces discussed in this chapter: why people demand mobility (Section 2.1.1), how their movement generally creates negative externalities (Section 2.1.2), and the ideas of agglomeration effects (Section 2.2.1) and how city structure (i.e., accessibility) influences travel behavior (Section 2.2.2). Subsequently, I summarize the policy levers of pricing (Section 2.3.1) and investment (Section 2.3.2).

In Chapter 3 I discuss the link between travel behavior and accessibility in a Swiss context. Chapter 4 presents the newly developed multimodal congestion mechanisms that can partially reflect investment effects. Then, Chapter 5 formulates mathematically the static traffic assignment based on the multimodal MFD, which is subsequently applied in a network design optimization problem in Chapter 6, as well as a general computational economic equilibrium model in Chapter 7, addressing urban housing, labor markets, and transportation markets. Finally, Chapter 8 discusses the implications of this thesis for the literature, policy makers, and practitioners before the summary concluding remarks in Chapter 9.

2

LITERATURE REVIEW

> *Not only does traffic have economic and safety effects, but it produces noise, it pollutes the atmosphere and it has adverse effects on the appearance of our towns. But on the other hand, it gives us extensive opportunities for enriching our lives.*
>
> — Smeed (1968)

This chapter aims at providing a comprehensive, but critical review on the overarching literature of the topics introduced in the previous chapter. In addition to this chapter, we provide in each following chapter further literature reviews more aligned to the actual content of each chapter.

We organize this chapter as follows. Firstly, Section 2.1 reviews the primary transport-related reasons and effects of *human interactions*. Here, following Smeed's 1968 quote, we discuss first in this chapter the "*opportunities for enriching our lives*" or in broader terms the demand for travel in Section 2.1.1. Thereafter, we introduce the reader to the negative effects of human interactions, namely congestion externalities in Section 2.1.2 that have been summarized by Wardrop (1952) with "*increases in the amount of traffic generally produce corresponding decreases in speed*". Second, this chapter then continues with discussing in Section 2.2 the two additional forces that affect the equilibrium and which commonly scale with population density. In detail, we first discuss agglomeration externalities in Section 2.2.1 which can be seen as the glue that allows cities to exist as higher output compensates for many disamenities experienced (e.g. congestion, pollution). Second, we introduce the reader to the vast research field of the relationship between the build environment (accessibility) and travel behavior in Section 2.2.2, which Ewing and Cervero (2017) summarized as "*Does Compact Development Make People Drive Less? The Answer Is Yes"*. We then end this chapter by discussing the constituents of the macroscopic system parameter settings $\mathbf{X}$, which are the transport policy levers. In particular, pricing in Section 2.3.1 and investment in infrastructure in Section 2.3.2. For completeness, we add traffic control in Section 2.3.3 to $\mathbf{X}$ as it is a frequently used mechanism. We summarize and synthesize the most relevant concepts and ideas of chapter in Section 2.4.

2.1 HUMAN INTERACTIONS

When people travel, they usually travel to obtain utility, a measure of pleasure or satisfaction (Ben-Akiva and Lerman, 1985), at the destination. In only few cases, e.g. hiking or taking the train up to the Jungfraujoch, traveling itself becomes part of creating utility. In transportation, the most notable source of utility reducing factors are travel time, cost and congestion delays (Small and Verhoef, 2007). The equilibrium is then that situation in which the utility obtained from more travel is consumed by more congestion delays. In the following, we therefore introduce the generation of utility in Section 2.1.1 and the generation of congestion externalities in Section 2.1.2.

2.1.1 *Opportunities*

Reasons for traveling are plentiful, making the term *trip purpose* a key element in the quantification of travel demand that we need for the analysis and modeling of our transport systems (Ortúzar and Willumsen, 2011). The most recent 2015 Swiss travel survey offered 13 different trip purposes (Swiss Federal Statistical Office and Swiss Federal Office for Spatial Development, 2017), e.g. working, going to school, shopping, leisure activity, accompanying other people. Figure 2.1 shows that commuting, shopping and leisure activities constitute the main reasons for travel.

With plentiful choices and opportunities to reach, a fundamental question for policy making and related fields is to which extent is mobility behavior predictable? At the individual level, Schlich and Axhausen (2003) noted that "*travel behaviour is neither totally repetitious nor totally variable*" using data from a six-week travel diary (Axhausen *et al.*, 2002). Analyzing collective dynamics of human behavior with anonymous mobile phone data then shows that human mobility behavior follows simple reproducible patterns (González *et al.*, 2008), which is further supported by evidence that the number of frequently visited locations is limited (Alessandretti *et al.*, 2018; Schönfelder and Axhausen, 2003), both of which in turn means a high level of predictability of human mobility behavior (Song *et al.*, 2010). Consequently, the answer to question of predictability is "*yes*", at least partially. Consequently, policy makers have to consider, but also can take advantage of these recurring patterns in their decision making.

In Sections 2.2.2 and 3.1 we provide an in-depth review on the relationship between the built environment - in other words accessibility - and the

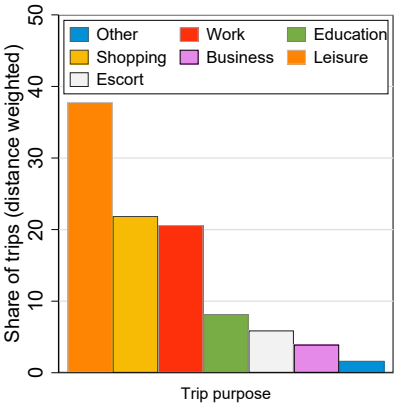


FIGURE 2.1: Distribution of trip purposes from the 2015 Swiss travel survey (Mikrozensus Mobilität und Verkehr).

demand for mobility. Therefore, we focus here on the research topics for travel demand. There are at least two topics related to information- and telecommunication technologies (ICT) which need to be mentioned. On the one hand, these technologies shift our demand to less active mobility: working from home, online shopping or ordering food online. On the other hand, ICT enabled innovative and radical car-sharing, ride-hailing and ride sourcing services (Farag *et al.*, 2003; Mokhtarian *et al.*, 2006; Cramer and Krueger, 2016; Schmid and Axhausen, 2018). However, these new services are disruptive to present business models, e.g. the taxi industry for which the earnings declined (Berger *et al.*, 2018), and public transport where some studies suggest a decline in ridership (e.g. Becker *et al.*, 2017a,b), even to that extent that it raises public concern as with London's Underground[1], while others report complementarity of these new services with public transport (Hall *et al.*, 2018).

Last, we have so far only considered the individual's utility from traveling, however, as individuals prefer to cooperate (Tomasello, 2009), the economic or utility effects from that interaction or cooperation between individuals must not be necessarily included in the travel utility. For example, trade interactions generate further benefits - the added value - than just giving the driver a salary (e.g. Levinsohn and Leamer, 1995); or fur-

1 https://www.wired.co.uk/article/tfl-finances-transport-for-london-deficit-passenger-numbers

ther the wider economic benefits from agglomeration (e.g. Venables, 2007; Graham, 2007a).

2.1.2 *Congestion externalities*

Traffic and all transport systems are physical multi-particle systems (Helbing, 2001; Treiber and Kesting, 2013), meaning that too many particles lead inevitably to crowding, jamming and congestion. In other words, congestion is occurring when the supply of road capacity is not sufficient for the travel demand. Important to note, as traffic is assumed to be in an equilibrium as pointed out by Wardrop (1952), congestion can be an equilibrium outcome although nobody wants it. The social importance of transport and its adverse effects in terms of congestion has resulted in a high degree of standardization of practices over the past seven decades (e.g. Forschungsgesellschaft für Straßen- und Verkehrswesen e.V., 2009; Transportation Research Board, 2016) as well as interest far beyond traffic engineers to understand the physical properties (e.g. Zhao *et al.*, 2005; Wu *et al.*, 2006; Nagatani, 1993; Li *et al.*, 2015).

The nature of these multi-particle systems implies that the time T to traverse a link is dependent on the number of flowing particles or vehicles N as expressed in Eqn. 2.1. The important physical property of the Eqn. 2.1 that annoys car drivers, planners and politicians is that the function is always increasing as shown in Figure 2.2.

$$T = T(N) \tag{2.1}$$

This increasing relationship has an important link to economics, where the delays imposed to all other drivers are negative externalities. In detail, an important question for policy making is how marginal costs, either social or private, evolve as a function of N. In other words, it is important to understand by how much the travel time for everybody on a link would increase if one additional traveller is present on the link? Mathematically, we can investigate this by Eqn. 2.2.

$$MC = \frac{\partial\, N\, T(N)}{\partial N} = T(N) + N\frac{\partial T(N)}{\partial N} \tag{2.2}$$

The two-terms on the right-hand side have two distinct meanings. The first one is the average costs on the link, while the second term corresponds the additional delay to all other vehicles by the vehicle at the margin (Small

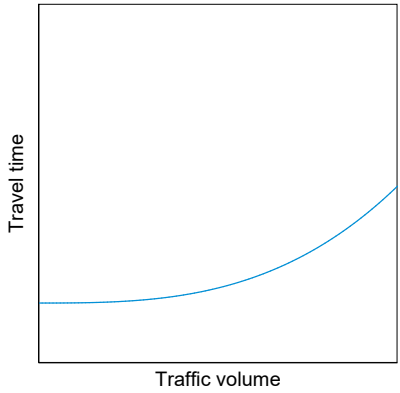


FIGURE 2.2: Monotonically increasing relationship between traffic volume and travel time.

and Verhoef, 2007; Ortúzar and Willumsen, 2011). The second term is always positive as Eqn. 2.1 is always increasing. There exist a substantial variety of functional formulations for Eqn. 2.1, where the most commonly used is the Bureau of Public Roads-function (Bureau of Public Roads, 1964) as given by Eqn. 2.3.

$$T = T^0 \left(1 + \alpha \left(\frac{N}{N^{max}}\right)^{\beta}\right) \tag{2.3}$$

Here, T^0 is the free-flow travel time on the link, N^{max} is the capacity, i.e. the maximum flow, and α and β are model parameters for calibration. It usually holds that $\beta \geq 2$. Figure 2.3 schematically shows the relationship between average and marginal link costs and strongly underlines that marginal costs are growing faster than average costs with traffic. For further insights, we refer the interested reader to (Branston, 1976; Spiess, 1990; Ortúzar and Willumsen, 2011) and to the sibling of volume-delay functions, the fundamental diagram (e.g. Del Castillo and Benitez, 1995a,b; Transportation Research Board, 2011; Bliemer and Raadsen, 2018).

There is widespread consensus that the provision of public transport is a congestion relief (e.g. Sherman, 1971; Small and Verhoef, 2007; Beaudoin *et al.*, 2015; Arnott and Yan, 2000; Harford, 2006). The rationale is that in a *two mode* case with cars and buses the equilibrium travel costs of both modes are identical. In other words, with less transit services available,

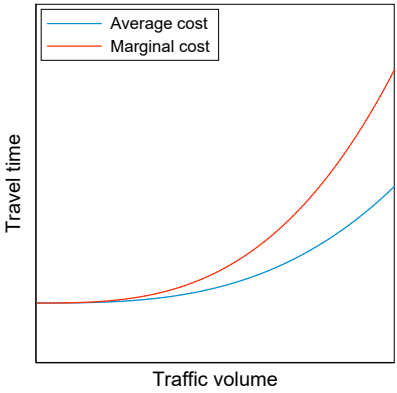


FIGURE 2.3: Comparing average and marginal cost curve of travel times based on the BPR function.

more cars will be used until the car costs match again the transit costs (Small and Verhoef, 2007). In real world experiments these effects have been revealed, either in context of intermittent closures of transit services, e.g. in Los Angeles (Anderson, 2014), Rotterdam (Adler and van Ommeren, 2016), and Melbourne (Nguyen-Phuoc *et al.*, 2018), or in context of transit line extensions, e.g. in Beijing (Yang *et al.*, 2018a).

Last, we already mentioned the idea of two *natural laws* of congestion in the first chapter. The first *natural law* is *induced demand*, meaning that adding capacity to the road network is increasing vehicle kilometers traveled (VKT). Here, in order to evaluate whether the investment is a congestion relief, one has to see at which rate VKT grows compared with the network expansion, i.e. in economic terms the elasticity. Empirically, short-term and long-term values have been found to be well-below one (Downs, 1962; Goodwin, 1996; Mogridge, 1997; Hansen and Huang, 1997; Cervero and Hansen, 2002; Weis and Axhausen, 2009), with the notable expection of the recent work by Duranton and Turner (2011) who provide a value close to one. In other words, an elasticity close to unity means that the investment did not result in congestion relief. Importantly, as it sometimes difficult to identify causality in demand-supply analyses, one can also see the problem from the supply side, namely *induced supply* (Levinson and Karamalaputi, 2003; Cervero and Hansen, 2002). The second *natural law* is about, despite literally nobody liking congestion, that people collectively recurringly create similar congestion patterns (e.g. Wang *et al.*, 2012; Lopez

et al., 2017; Olmos *et al.*, 2018), and they create congestion patterns according to predictable features such as road network topology, demand (e.g. Çolak *et al.*, 2016; Wang *et al.*, 2015).

2.2 DENSITY EFFECTS

It seems like a natural law that individual animals' and plants' use of resources is scaling approximately by 3/4 of their size or body mass, implying some sort of economies of scale (West *et al.*, 1997; Enquist *et al.*, 1998). A substantial amount of literature suggests evidence that cities also exhibit such scaling relationships in their infrastructure that result from the intense competition for urban space (Batty, 2008; Bettencourt, 2013; Bettencourt *et al.*, 2010). In particular, Bettencourt *et al.* (2007), Kühnert *et al.* (2006) and Levinson (2012) report a variety of scaling relationships, e.g. that the number of petrol stations, length of power cables and road surface is sublinear scaling with city size.

The size of a city then usually also has effects on the topology of road networks and then accessibility. Levinson (2012) reports that larger cities are usually more interconnected, quantifying that a one percent increase in accessibility is reducing average commute time by 90 seconds in the US. Comparing to biological systems, Samaniego and Moses (2008) report that urban road networks are less centralized making per capita road capacity independent of the size cities. Scaling effects, however, also are observed for transit networks that allow to predict with city area, population and income the length, number of stations and ridership of subways and rail networks (Louf *et al.*, 2014).

In this thesis, we focus on two (scaling) effects of population density that *push* and *pull* households and firms in their decisions (Fujita and Thisse, 2002). Consequently, we discuss first agglomeration effects in Section 2.2.1 and second the built environment effects in Section 2.2.2.

2.2.1 *Agglomeration effects*

There exist a vast amount of empirical evidence that output of individuals and firms per unit, i.e. productivity, is increasing or is scaling superlinear with density, as measured in, e.g., wages and patents at a scaling factor of about 1.05 to 1.2 (e.g. Henderson, 2003; Bettencourt *et al.*, 2007; Melo *et al.*, 2017; Holl, 2016; Ciccone, 2002; Melo *et al.*, 2009; Hanlon and Miscio, 2017; Axhausen *et al.*, 2015). In other words, doubling density, is not only

doubling output, but is generating additional output according to the scaling relationship. These productivity gains are externalities resulting from agglomeration or urbanization economies (Fujita *et al.*, 2001; Fujita and Thisse, 2002; Duranton and Puga, 2004; Rosenthal and Strange, 2004, 2001; Krugman, 1991). Mathematically speaking, output y is produced from inputs x with a production function f and the process is subject to external productivity scaling that result from z and follow function h. Total output is then $y = h(z) f(x)$. Consequently, an increase in productivity $h(z)$ requires less input for the same output (Graham, 2007a).

Levinson (2012) provides an intriguing explanation as it is "*the efficiency of interaction (and hence the number of contacts per unit time, and the amount of time spent with contacts rather than in transport) that brings about that superlinear scaling*". Using mobile phone data, Schläpfer *et al.* (2014) then actually showed that the number of interactions indeed scale with city size. However, there are economic reasons for agglomeration: Marshall (1920) named labor market pooling, input sharing and knowledge spillovers, while already Thünen (1826) formulated an expression on the forces that load to agglomeration when asking whether firms are better off when they locate in cities (Fujita and Thisse, 2002). We refer the interested reader to the work by Rosenthal and Strange (2001, 2004) as well as Duranton and Puga (2004) that provides a throughout review on the (micro) mechanisms underlying agglomeration economies. Consequently, the reasoning is that improvements in infrastructure will increase the number interactions per unit time as less time is spent in transport, which affects z in such a way that productivity is increased.

The accessibility approach by Hansen (1959) is a widely accepted measure of proximity of economic activity as it basically quantifies the price of access to markets, both for households and firms (Vickerman, 2008; Rokicki and Stępniak, 2018). The Hansen (1959) measure of accessibility is usually expressed as the sum of opportunities weighted by the generalized transportation costs between locations (see also Section 3.2). Consequently, improving speeds at one location in the network is improving accessibility of the entire network and as Graham (2007a) reports that "*if transport investment changes the densities available to firms [...] then there are likely to be positive gains from agglomeration*". In other words, reducing the generalized transportation costs or increasing accessibility can lead to a substantial relocation of economic activity and thus accelerate the agglomeration process (Krugman, 1991; Venables, 2007; Glaeser, 2008; Vickerman *et al.*, 1999; Vickerman, 2008). However, when a region already has a high

quality transportation system, more investment to improve accessibility must not necessarily leads to increasing productivity gains (Banister and Berechman, 2001), but can only provide additional capacities. Generally, the reported elasticity for agglomeration economies is around, and often less than 0.1, but the effect is robust and substantial (Graham, 2007a; Melo *et al.*, 2009, 2017).

This superlinear relationship implies that allocating more resources to urban areas compared to non-urban areas results in larger productivity gains (Melo *et al.*, 2017). Consequently, it is not surprising that many countries initiated programs to promote collaboration in agglomeration of cities[2]. Here, it is then important to consider how the agglomeration effects attenuate over distance (e.g. Rosenthal and Strange, 2004; Graham *et al.*, 2010; Melo *et al.*, 2017).

2.2.2 *Travel behavior*

Travel behavior is frequently considered to be not uniform across space: Neighborhoods with large population density, land use diversity, accessibility and better access to public transport are generally found to reduce driving (Ewing and Cervero, 2017). As the quantification of travel demand is key in transport modeling and research (Ortúzar and Willumsen, 2011), the analysis of the influence of the built environment or land-use on travel behavior has attracted substantial research interest and is one of the most studied subjects in the field (for comprehensive revies we refer here to the work by Ewing and Cervero, 2001, 2010).

Cervero and Kockelman (1997) summarized their findings into the concept of "*three Ds*" of the build environment, namely density, diversity and design. Ewing and Cervero (2010) follow their classification, but also include two further "*D*" variables: destination accessibility and distance to - and quality of - public transport. In the following, we focus on accessibility as it at the same time provides a link to agglomeration externalities as introduced above in Section 2.2.1. Further, it can be seen as a social indicator, because as Geurs and van Wee (2004) define accessibility as "*the extent to which land-use and transport systems enable (groups of) individuals to reach activities or destinations*". As already mentioned in the introduction, this makes accessibility a good indicator to model mobility (see also Daganzo, 2007).

2 For example in Switzerland: `https://www.are.admin.ch/are/de/home/staedte-und-agglomerationen/strategie-und-planung/agglomerationspolitik-des-bundes-2016-.html`

In their review, Geurs and van Wee (2004) identify four components of accessibility to which an accessibility measure should be sensitive. First, it should result from travel times or costs. Second, the accessibility measure should describe the amount, quality and spatial distribution of opportunities. Third, it should capture temporal constraints regarding the opportunity availability of individuals to access these, e.g. congestion. The last component may include an individual's needs and constraints. However, the authors themselves concluded "*applying the full set of criteria would imply a level of complexity and detail that can probably never be achieved in practice*". In a similar vein, Handy and Niemeier (1997) concluded that no best approach exists to measure accessibility, whereas Owen and Levinson (2015) stated that the location based measure of cumulative opportunities is the most probably simplest approach. The cumulative opportunities measure counts all opportunities at destinations that can be reached within a travel time radius and is consequently simpler as Hansen (1959) as opportunities are not weighted by the generalized cost of travel.

2.3 POLICY LEVERS

The aforementioned forces in Sections 2.1 and 2.2 are likely to result in an equilibrium that is rarely considered optimal, i.e. too many vehicles, resulting in too much congestion. Thus, we discuss in the following sections how the market failures of too many vehicles can be corrected with pricing (Section 2.3.1), investment (Section 2.3.2) and control (Section 2.3.3). Although out of scope of this thesis, we include the latter for reasons of completeness, as it is generally important

2.3.1 *Pricing*

In economic theory, the mechanism to control and correct inefficient human behavior is pricing (Pigou, 1920; Small and Verhoef, 2007). The price results from the equilibrium between supply and demand, i.e. the intersection of the demand and supply curves. Using taxation, either of the two curves can be shifted, e.g. as a policy to correct for a less optimal market outcomes (Pigou, 1920).

With congestion being considered the most costly externality of road transportation (Small and Verhoef, 2007), consequently, substantial interest in the literature emerged aiming to correct for this market failure, i.e. internalizing the congestion externalities. Further, Vickrey (1969) noted that

prices have the advantage over infrastructure investment that they affect demand at short notice. The theory of *road pricing* is to charge a toll for driving in a network exactly at the cost of the additional delays the driver at the margin will impose on all other drivers. As this is again a timeless issue in transport research, we refer the interested reader to the textbooks and reviews by Small and Verhoef (2007), Parry (2009), Anas and Lindsey (2011) and de Palma and Lindsey (2011). Although, many argued that these road pricing schemes work only in theory, but not in practice, some cities had the courage to implement such systems successfully: Most notably are the systems in London (the Congestion Charge), where evidence suggests that congested levels declined in London (Leape, 2006; Prud'homme and Bocarejo, 2005; Santos and Shaffer, 2004), similarly in Stockholm (Eliasson, 2009), Milano (Rotaris *et al.*, 2010), as well as in Singapore (the electronic road pricing (ERP) system). As the system exists for so long and Singapore is strongly limiting car ownership, now it is difficult to access quantitatively its success independently.

As public transport is considered a congestion relief, how should transit be priced in a multimodal city? In a simplified context with a substantial preference for cars as well as usually lower travel times by car without congestion, in theory, the fares for the transit system should be subsidized from the revenue collected from car drivers (Small and Verhoef, 2007). Empirical evidence suggests that current subsidy levels are justified and economically efficient, while further fare reductions would increase welfare further (Parry and Small, 2009). However, in case public transport would be free, it is not expected to attract as many car drivers as needed to alleviate congestion (see for experiments Cats *et al.*, 2014; Thøgersen, 2009), and thus most likely not to maximize mobility for everybody.

In transportation systems, however, the idea of marginal cost based pricing and taxation is for a variety of reasons not applicable or too difficult. For example, different modes are competing and interacting in a dynamic system which makes the marginal (external) cost functions also dynamic. In addition, in many cities, the prices and taxes have several elements (e.g. fixed and variable components) with many involved agencies. Therefore, in this thesis, we use the structure of existing pricing schemes to correct for the less optimal equilibrium and consequently do not follow the idea from economic theory of *first* or *second*-best pricing (Small and Verhoef, 2007). With the rationale of this thesis to maximize mobility for everyone (Daganzo, 2007), the prices of all modes should be adjusted while accounting

for mode interactions (Kraus, 2012; Tirachini and Hensher, 2012; Tirachini *et al.*, 2014).

2.3.2 *Investment*

Investment in infrastructure, either for roads or public transport might be more appealing than increasing prices as it not only means cutting ribbons, but also is supposed to provide benefits for travelers and society. Related to this is the paper "*Roads to prosperity?*" by Fernald (1999) that summarizes the question whether we can achieve with infrastructure investment not only improvements for the convenience of traveling, but also in terms of productivity gains.

As already mentioned in Section 2.1.2 for congestion, building roads does not necessarily improve traffic due to induced demand. Undoubtedly does more infrastructure increase the total capacity of networks, however, as Smeed (1966) analytically derives are the returns of investment decreasing. Unfortunately, the analysis by Wong and Wong (2016) of the network topological effects on the network performance does not consider this feature, while they intuitively recover that intersection density is reducing the free-flow speed in networks.

With investment made to reduce travel time, people are likely to increase not only their range of accessible destinations, but also to cover larger distances as they now can drive faster (Metz, 2008). This is among the effects that lead to the widely observed urban sprawl (Brueckner and Fansler, 1983), where especially highway expansions led to a substantial decentralization of metropolitan areas in the US (Baum-Snow *et al.*, 2017).

When investing in public transport, especially underground lines that run independently of car traffic, the rationale is to provide alternative transport modes and thus congestion relief. Empirical studies indeed show that this short-term effect was present after the opening of line extensions (e.g. Beaudoin *et al.*, 2015; Yang *et al.*, 2018a), while similar effects have been observed during transit strikes (e.g. Adler and van Ommeren, 2016). However, Stopher (2004) summarizes in his report on reducing congestion that "*achievement [to attract more public transport travellers] would be relatively small on the overall congestion of the road system, and that these effects would also be likely to be fairly short-lived*". Then, at some point transit networks become more complete and as Roth *et al.* (2012) conclude that there exist a long-term limit of underground networks' shapes such as it would be a universal mechanism that is governing investment decisions. This

is further in line with Louf *et al.* (2014) findings on the predictability of network design. In other words, at some point, investing further in underground lines might not help much as the entire network is already quite complete. Last, similar to the freeway expansion, Gonzalez-Navarro and Turner (2018) noted an decentralization effect as well, although in this case the effect is smaller than the effect from highways.

Coming back to Fernald's 1999 question on the productivity gains from infrastructure investment, he himself concluded that "*at the margin, however, road investments do not appear unusually productive. Intuitively, the interstate system was highly productive, but a second one would not be. Road-building thus explains much of the productivity slowdown through a one-time, unrepeatable productivity boost in the 1950's and 1960's*". This view is supported by Aschauer (1989) who reported a positive relationship between investment and productivity gains, but also noted that the decline in productivity gains in the 1970s is due to the decline in investment. Regarding the effects, elasticities of around 0.2 between investment and productivity improvements have been reported (Graham, 2007a; Duranton and Turner, 2012). Thus, we can conclude that investment can be indeed productive and this has been formalized and analytically explained by Venables (2007), but still further research is necessary, especially for public transport, to evaluate impacts on agglomeration (Chatman and Noland, 2011).

2.3.3 *Traffic control*

Traffic control is are a further mean to influence the choices of people as it manages which mode has to wait longer and has higher generalized transportation costs. In other words, which mode is getting allocated more time and space? In contrast to investment, traffic control is less radical than building roads, but one can argue that the effects will not dramatically change the system. To get a wider perspective on urban traffic control, we refer the interested reader to textbooks (Daganzo, 1997; Guberinić *et al.*, 2008) and reviews (e.g. Papageorgiou *et al.*, 2008; Stevanovic, 2010; Hamilton *et al.*, 2013).

In this thesis, we are concerned with a system perspective of urban traffic, in particular the MFD, and its subsequent applications. We accordingly focus in the following on network-wide traffic control measures. As a general notice, it should be clear that not all measures can be combined as some of them are evidently contradicting, e.g. transit priority and signal propagation for cars.

The first measure is perimeter control (Haddad and Geroliminis, 2012; Geroliminis *et al.*, 2013; Yang *et al.*, 2018b; Haddad and Mirkin, 2017; Kouvelas *et al.*, 2017; Aboudolas and Geroliminis, 2013). Here, the idea is to define an area, usually the CBD, within the network that should be protected against congestion. This protection is ensured as follows. The traffic states within the protected area are measured by the MFD, once the density of the network reaches the critical density, the signals on the inbound arterials reduce the green time to such a level that the amount of inflow equals the outflow (leaving the protected area or end their trip). This approach of course creates congestion outside the perimeter, but *storing* cars along arterials is simpler, less costly in terms of space and less conflicting routes and modes are usually present. So far, no city has such a MFD-based perimeter control, however, the city of Zürich implemented a similar scheme but instead of density it operates with flow thresholds (Dönier *et al.*, 2013; Ambühl *et al.*, 2018c).

The second strategy is to implement a system with adaptive signals. In other words, the offsets between signals adapt according to the flow of vehicles, i.e. the size of platoons, to maximize the flow of vehicles (Webster, 1958; Hunt *et al.*, 1982; Daganzo *et al.*, 2018; Daganzo and Knoop, 2016; Lämmer and Helbing, 2008). These systems are widely used nowadays as several companies provide large scale solutions for monitoring and control. So far, most of the network-wide adjustments are done manually and none is based on the MFD so far. However, the work by Daganzo *et al.* (2018) provides starting points.

A third measure is transit priority (Garrow and Machemehl, 1999; Baker *et al.*, 2002; Smith *et al.*, 2005; He *et al.*, 2016; Guler and Menendez, 2014; Guler *et al.*, 2016; Kohla and Fellendorf, 2015), where transit vehicles not only operate on their dedicated lanes, but also have priority at intersections. In other words, once a bus or tram approaches an intersection, signals for all vehicles except the public transport vehicle turns red and the public transport vehicle can drive over the intersection with almost no delays. Again, these kind of systems are already widely operating similarly as adaptive signals. For example, the city of Zürich operates such as system where in most cases, a transit vehicle has to wait less than three seconds at a red signal.

Last, a rather novel idea are intermittent bus lanes (Eichler and Daganzo, 2006; Chiabaut *et al.*, 2012). The idea is that once a bus actually requires a dedicated bus lane, the bus lane is dedicated to buses and in all other cases, the space is available for cars. Operating bus lanes only during peak

hour, as it is currently done in London, is a less dynamic alternative of this idea, but already allows for some flexibility in sharing space. So far, however, an actual intermittent system has only been tested in Lyon and first results show that this kind of control measure is promising (Chiabaut and Barcet, 2018).

2.4 SUMMARY

In this chapter, we have outlined the most relevant ideas within literature for this thesis. In particular, we see how this thesis is aligned with recurring and timeless questions from economics, engineering and social sciences. Nevertheless, this review also emphasizes that many mobility problems we face today will persist as long as individuals populate the city. Consequently, the social optimum will most likely mean that travel times will still be unsatisfactory for most travelers.

3

ACCESSIBILITY AND TRAVEL BEHAVIOR

> *Humans follow simple reproducible patterns. This inherent similarity in travel patterns could impact all phenomena driven by human mobility, from epidemic prevention to emergency response, urban planning and agent-based modelling.*
>
> — González *et al.* (2008)

As mentioned in Section 2.2.2, it seems like a universal law governing individual's mobility behavior: more dense environments, i.e. better accessible areas, make people drive less and take other modes of transportation (Ewing and Cervero, 2017). Although this pattern has been observed around the world, the investigation has mostly been from a car **or** transit perspective. However, for countries like Switzerland (or the Netherlands or Singapore) with a competitive public transport system, this universal law requires a comprehensive multimodal perspective. This chapter contributes with a procedure to quantify multimodal accessibility and to incorporate the newly derived accessibility measure in the analysis of multimodal travel behavior. As this thesis deals with a Swiss context, the dimensions of multimodal travel behavior are car ownership, season-ticket ownership as well as the number of car, transit and non-motorized trips.

Therefore, we deepen our understanding of this universal law from Chapter 2 with focus on multimodal aspects in Section 3.1. We introduce our approach to model multimodal accessibility in Section 3.2 with an empirical application to Switzerland. Thereafter, we propose an econometric model, introduced by Bhat (2015), to analyze multimodal travel behavior in Section 3.3 where we use the multimodal accessibility measure as explanatory variable. We then present the results of the model estimation for the 2010 Swiss travel survey in Section 3.4. Last, we summarize the findings and conclusions of this chapter in Section 3.5.

3.1 BACKGROUND

The analysis of the relationship between travel behavior and the built environment is predominantly focusing on the influence of car-related travel

behavior. Accordingly, most existing methods have been developed for cars, for example, ownership models (e.g. de Jong *et al.*, 2004; de Jong and Kitamura, 2009; Anowar *et al.*, 2014) and driving distance (e.g. Hamilton Ailsa, 1982), or car ownership and use (e.g. Bhat and Sen, 2006; Tanner and Bolduc, 2014; Jäggi *et al.*, 2012). Other mobility tools such as public transport season tickets or bicycles, however, have been far less of interest, especially the jointness of choices (e.g. Scott and Axhausen, 2006; Yamamoto, 2009; Simma and Axhausen, 2001).

The lack of interest in analyzing multimodal choices of travel behavior is then also reflected in low interest of capturing multimodal aspects in measuring accessibility. So far, most studies focused either on accessibility by car or public transport (see for an overview Ewing and Cervero, 2010) and only a few combined both modes (e.g., Kuzmyak *et al.*, 2006; Shen, 2000; Scott and Axhausen, 2006; Jäggi *et al.*, 2012). Capturing multimodal aspects in accessibility is important to describe the nuances in the competition of modes, e.g. when at almost similar levels of accessibility transit offers a small advantage, we would expect that individuals will make slightly more decisions for public transport, while we expect the opposite if transit has a small disadvantage. Further, this idea is not yet comprehensively linked to Hansen's 1959 notion of accessibility which offers a simple quantification of the level of mobility as well as a link to economic impacts (see Section 2.2.1).

As we focus in this thesis on multimodal travel activity of mobility tool ownership and use, it is necessary to consider briefly the relevant literature as well. First, mobility tool ownership is usually a discrete question, either a car is owned or not. There is a vast number of textbooks that introduce to the field of discrete choice modelling and thus we refer the interested reader to this work (e.g. Ben-Akiva and Lerman, 1985; Train, 2009; Greene, 2012). In most of the discrete choice models, the underlying idea is that the decision maker has a latent propensity score, i.e. a utility function, which is affected by individual attributes (e.g. income, age) and outcome-related attributes (e.g. price, travel time). This score is mapped to discrete choices with a specific function, e.g. the logistic function for the logit model or the cumulative normal distribution function for the probit function. The parameters are usually estimated from observations with maximum likelihood (see Appendix D.1 for details).

Second, travel activity is a continuous (*how far will I drive?*) or ordered choice (*How many trips will I make?*). Thus, the model can be a simple linear model (e.g. Bento *et al.*, 2005), but the nature of only positive driving dis-

tance can be accounted for in a truncated Tobit model (e.g. Redmond and Mokhtarian, 2001). As causality is not trivial in the question of how multimodal travel behavior is affected by land use, structural equation modelling (SEM) can be used (e.g. Cervero and Murakami, 2010). If travel activity is count data or ordered, but not continuous, Poisson or ordered Probit models (e.g. Redmond and Mokhtarian, 2001) can be used. In addition, the sample might be corrected for sample selection as people without a car available might not be eligible to answer on how many kilometers they traveled (Heckman, 1979).

Third, Bhat (2005) proposed an econometric model to jointly model several discrete-continuous choices in the multiple discrete-continuous extreme value (MDCEV) model. This modeling framework has been frequently applied for household's vehicle holdings (e.g. Bhat and Sen, 2006; Jäggi *et al.*, 2012), time-use models (e.g. Bhat, 2005) and household's expenditure models (e.g. Ferdous *et al.*, 2010). However, also much simpler structural equation models can be used to estimate multiple discrete-continuous choices (e.g. Simma and Axhausen, 2001).

3.2 MULTIMODAL ACCESSIBILITY CALCULATION

The methodology to a obtain multimodal Hansen (1959)-based measure of accessibility first requires as an input travel times or costs c_{ijm} from origin i to destination j for a set of modes m. The set of modes has at least two elements: cars and public transport. We further require spatial information about the distribution of a set of opportunities o at each destination O_{jo}. This set has usually at least one element, but here we use two elements: population and employment to illustrate how this method generalizes into the dimensions of modes and opportunities. In transport planning, origin and destinations are typically zones (i.e. traffic analysis zones (TAZ)) that overlap fully or partially with statistical zones. The travel costs then are obtained from each (arbitrarily chosen) centroid to other centroids of TAZs.

Second, with given travel costs and opportunities, we then calculate the Hansen (1959)-based accessibility measure A_{imo} as given by Eqn. 3.1.

$$A_{imo} = log\left(\sum_{j=1}^{N} O_{jo} \cdot exp\left(\beta_{mo} c_{ijm}\right)\right) \tag{3.1}$$

A_{imo} is a measure of destination accessibility of TAZ i to all other zones N with mode m and for opportunities o. The distance decay parameter β_{mo} takes into account that more distant destinations are less attractive.

For Switzerland, β_{mo} has been estimated for each mode by Sarlas *et al.* (2015): $\beta_c = -0.261$ and $\beta_b = -0.034$. Typically, the generalized cost of travel c_{ijm} are equal to the in-vehicle time from i to j for each mode, but for public transport additionally contains access & egress time, waiting time and transfers.

Third, the previously computed $|m| \cdot |o|$ accessibility measures expected to correlate strongly. Using each of them in further econometric analysis provides an inevitable risk of multicollinearity. Reasons for this are that at larger scale, the distributions of opportunities and travel costs are likely to be similar, but not identical, e.g. both modes frequently share infrastructure and residential areas overlap with work places. Thus, to reduce the available accessibility measures to their most defining underlying structure, we carry out a principal component analysis following the idea of Jäggi *et al.* (2012). The obtained principal components then recover the essential directions of accessibility in the data (Jolliffe, 2002).

We apply this procedure to Switzerland with two modes: cars and public transport and two opportunities: employment and population. The travel times are available from the 2010 national macroscopic transport models for car and public transport. In both transport models, TAZ zoning follows the municipality boundaries, except for large cities that are further subdivided, but still boundaries follow the statistical zoning. In total, both models have 2949 zones within Switzerland. From the same model and accordingly for the same zoning, we have information on the population and employment numbers for each TAZ. Thus, we compute for each TAZ i accessibility A_{imo} with Equation 3.1.

Four accessibility measures are then available: (1) population accessibility by car, (2) job accessibility by car, (3) population accessibility by public transport, and (4) job accessibility by public transport. In the next step, we uncover the essential directions of accessibility in the four accessibility measures using a principal component analysis. Results of the analysis are presented in Table 3.1.

The first component explains more than 90 % (as measured in the proportion of the Eigenvalue) of the variation in the data and we interpret it as general accessibility. The second component explains 7.6 % of the variation and describes comparatively better accessibility by public transport and the third component explains 0.3 % of the variation in the data and describes comparatively better job accessibility. The fourth component does not have a meaningful interpretation for this analysis and is thus omitted. A prominent criterion for the selection of the number of principal com-

a) Summary statistics

	Component 1	Component 2	Component 3	Component 4
Eigenvalue	3.67715	.310054	.0112005	.00159724
Proportion of the Eigenvalue	0.9193	0.0775	0.0028	0.0004
Cumulative proportion	0.9193	0.9968	0.9996	1
N = 2949				

b) Loadings

Population accessibility by car	0.5019	- 0.4697	-0.6857	-0.2394
Job accessibility by car	0.4969	- 0.5306	0.6427	0.2419
Population accessibility by public transport	0.4997	0.5099	-0.2160	0.6660
Job accessibility by public transport	0.5015	0.4877	0.2647	-0.6638

c) Correlations of factors and items

Population accessibility by car	0.9624	-0.2615	-0.0726	-0.0096
Job accessibility by car	0.9529	-0.2955	0.0680	0.0097
Population accessibility by public transport	0.9582	0.2839	-0.0229	0.0266
Job accessibility by public transport	0.9616	0.2716	0.0280	-0.0265

TABLE 3.1: Results of the principal component analysis of the four accessibility variables.

ponents is the Eigenvalue criterion. All components with an Eigenvalue of equal or greater than one should be selected. In this analysis, only the first principal component satisfy this criterion. Nevertheless, for three reasons we do not follow this criterion and select the first three principal components. First, we identify for the first three components a meaningful interpretation. Second, if we would consider only the first principal component, which is highly similar to each of the four accessibility measures, the principal component analysis would be pointless because we would ignore the differences in the accessibility measures. Third, we compared different model specifications including either only the first, the first two or the first three principal components with a likelihood ratio test and found that using all three components improves the model significantly.

After the estimation of the principal components, we calculate for each TAZ the score values of the general accessibility, comparatively better accessibility by public transport and better job accessibility. The calculation multiplies for each zone the four original accessibility values with the loadings from Table 3.1 for the considered principal component. We illustrate the spatial distribution of the score values in Figures 3.1 and 3.2. Figure 3.1 shows that the general accessibility is highest in metropolitan regions and the densely populated Swiss plateau, but low in Alpine regions. We have added the Swiss motorway (purple lines) and railway (black lines) network to the map. The zones with high levels of general accessibility overlap with motorways and dense railway networks in large parts of the country. Figure 3.2 shows spatial distribution of the second component, comparatively better accessibility by public transport. Again, we have added the motorway and railway network to the map. The value distribution does not follow the population distribution, as in the case of the general accessibility, but we observe that many municipalities close to the motorway network score low in this accessibility measure. The values do not score highest in centers of metropolitan regions, but in the agglomeration and countryside/Alpine regions. We can, for example, explain high values in Alpine regions by citing existing railway and limited car networks.

3.3 ECONOMETRIC MODEL

As aforementioned, we want to model the multimodal travel behavior in Switzerland in its five dimensions: Car and season-ticket ownership as well as the number of car, transit and non-motorized trips. Thus, we model mobility tool ownership and use with a multivariate probit-based model for

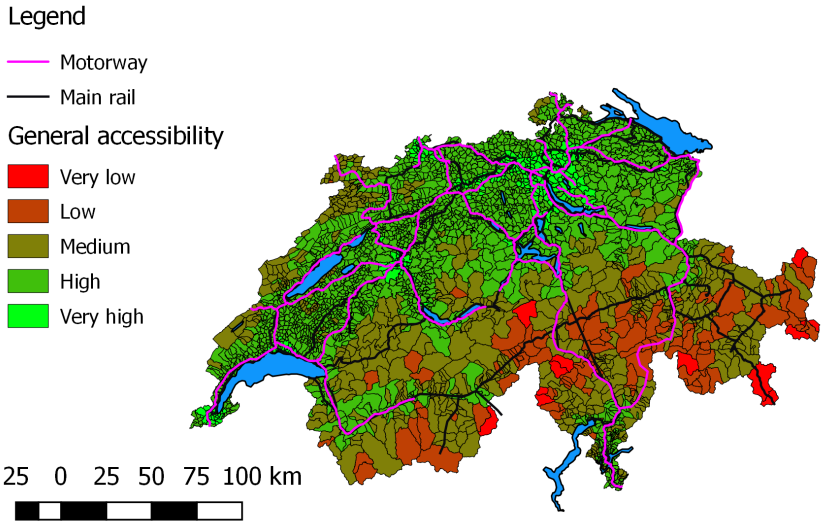


FIGURE 3.1: General accessibility levels in Switzerland. The values correspond to the scores calculated from the accessibility values of each municipality and the loading from Table 3.1. Higher values mean greater general accessibility. The purple lines show the Swiss highway network while the black lines correspond to the main railway network.

mixed type of outcomes, as introduced by Bhat and his colleagues (Paleti *et al.*, 2013; Bhat, 2015; Bhat *et al.*, 2016). For a detailed description, we refer the interested reader especially to Bhat *et al.* (2014). In this model, relationships between choice outcomes are established by allowing for correlations of error terms and endogenous variables' structural effects. This probit based model is an extension of the traditional multivariate probit, e.g. (e.g. Scott and Axhausen, 2006; Yamamoto, 2009; Andrés and Gélvez, 2014).

For the readers' convenience, we omit in all equations the subscript for number of the outcome equation. The choice of owning a mobility tool is modeled with a binary probit. We define a latent propensity $Y^* = \beta x + \varepsilon$, with β a vector of coefficients to be estimated, x a vector of exogenous covariates and the normally distributed error term ε. If $Y^* > 0$, the observed outcome is chosen, i.e. $Y = I(Y^* > 0)$. The outcome of number of trips is modeled as a generalized ordered probit (Bhat *et al.*, 2014; Bhat,

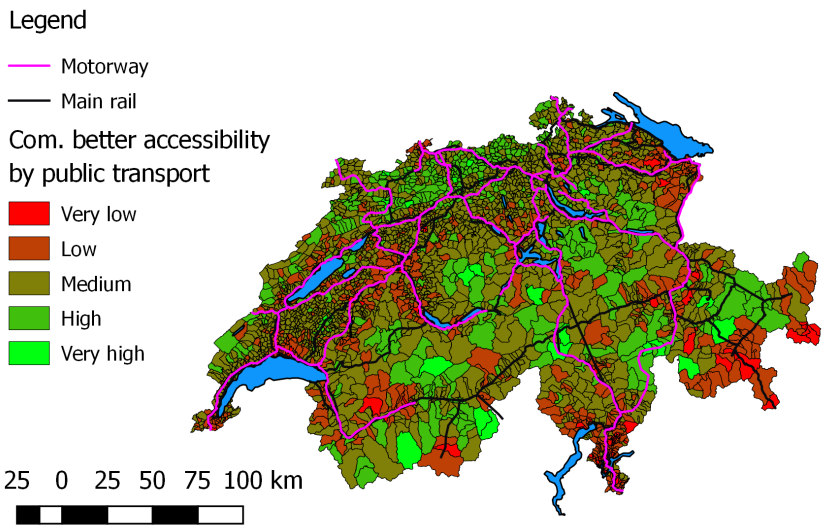


FIGURE 3.2: Comparatively better accessibility by public transport. The values correspond to the scores calculated from the accessibility values of each municipality and the loading from Table 3.1. Higher values mean comparatively better accessibility by public transport. The purple lines show the Swiss highway network while the black lines correspond to the main railway network.

2015). The generalized ordered probit also has a latent propensity $Y^* = \varepsilon$, which is mapped to the observed count outcome j by threshold parameters ψ_n. For the observed count value $j = n$, the following condition holds $\psi_{n-1} < Y^* < \psi_n$. The threshold parameters ψ_n are determined by the function

$$\psi_n = \Phi^{-1}\left(\frac{(1-c)^\theta}{\Gamma(\theta)}\sum_{r=0}^{n}\left(\frac{\Gamma(\theta+r)\,c^r}{r!}\right)\right) + \varphi_n \tag{3.2}$$

with

$$c = \frac{exp(\beta x)}{exp(\beta x) + \theta} \tag{3.3}$$

Dispersion parameter θ and flexibility parameter φ in Equations 3.2 and 3.3 allow flexible count distribution modeling. Φ is the cumulative normal distribution function, Γ is the gamma function, x is a vector of exogenous and endogenous covariates and β a vector of parameters to be estimated.

Error terms of each outcome equation correlate pairwise with ρ and constitute the correlation matrix P. For identification, we set $\varphi_{-1} = -\infty$, $\varphi_0 = 0$ and $\varphi_{n>0} = \varphi$ for each count outcome. The model parameters β, θ, φ and P are estimated with maximum likelihood. For each observation the likelihood is defined by

$$L(\beta, \theta, \phi, P) = \int_{\gamma_{low}}^{\gamma_{upp}} \phi_5(\tilde{u}|P)\, d\tilde{u} \tag{3.4}$$

The probability is obtained by integrating the five-dimensional normal density distribution ϕ_5 from γ_{low} to γ_{upp}, both five-dimensional vectors. For the binary outcome, the lower integration bound is $-\infty$ and the upper integration bound is determined by evaluating the corresponding outcome equation for Y^*. For the count outcome, the integration domain is determined by individual threshold values ψ_{n-1} and ψ_n. For the estimation of the model parameters, we use the maximum approximate composite marginal likelihood (MACML) (Bhat and Sidharthan, 2011) for which Bhat (2011) reported that the MACML approach recovers estimates just as well - but faster - as the simulation approach and that reduction in efficiency by the marginal compared to the simulation approach is "in the range of nonexistent to small".

3.4 RESULTS

We use the newly derived accessibility measure from Section 3.2 and the econometric model from Section 3.3, to investigate the link between our derived multimodal accessibility measures and mobility tool ownership and use. For this, we present in Section 3.4.1 the data that we use for the model estimation and thereafter the model estimates in Section 3.4.2.

3.4.1 *Travel behavior data*

Data on mobility tool ownership, number of trips and accompanying sociodemographic information is provided by the Swiss national transportation microcensus for the year 2010. The transportation microcensus is a large-scale survey carried out every five years with approximately 1 % of the Swiss population. In 2010, 59'771 households and - within these households, 62'868 individuals - were interviewed about their travel behavior (Swiss Federal Statistical Office and Swiss Federal Office for Spatial Development, 2012). We exclude anyone who can only move with outside sup-

Mobility tool		Season ticket					
		No		Yes		Total	
		N	%	N	%	N	%
Car	No	9'496	18.10	8'309	15.83	17'805	33.93
	Yes	29'364	55.96	5'307	10.11	34'671	66.07
Total		38'860	74.05	13'616	25.95	52'476	100

TABLE 3.2: Jointness in mobility tool ownership, illustrated by the cross tabulation of car and season ticket ownership.

port, all cases where we cannot impute the income and all cases younger than 18 from the sample. When two persons of a household reported on their travel behavior in the census, the second observation was in most cases a child. The final sample has 52'476 complete observations.

This analysis models individuals' decision making. For each individual in the data set, we extract five dependent variables of interest, car and season ticket ownership and the number of car, public transport and non-motorized trips as follows: car ownership is defined as having a car exclusively available. All individuals without driver's license are coded as having no car available. Season ticket ownership is defined as having any kind of season ticket subscription offering unlimited use of public transport, on either a regional or national scale. The number of trips is taken from the microcensus' travel diary, encompassing a single day. In each of the three trip variables, we pool the count outcomes of 1 and 2 trips into a single outcome and all outcomes larger than 11 to the outcome of 11. We did the first because just one trip was rarely observed and the latter because we wanted to avoid long tails in the distribution.

Table 3.2 shows that 55.96 % of all observations only have a car, 18.1 % have neither a car nor a season ticket, 15.83 % have only a season ticket and 10.11 % have both mobility tools available. Table 3.3 shows the univariate distributions of the number of trips by car, public transport and non-motorized modes. The total share of immobile persons in the dataset is 10.6 %. However, we cannot ignore the potential influence of soft-refusal (not reporting of shorter trips), especially for the non-motorized trips (Madre *et al.*, 2007). The accumulation of zero trips is highest for public transport trips and lowest for car trips, which also shows the longest tail. In

Mobility tool		Number of trips		
		Car	Public transport	Non-motorized
Car	No	0.489	0.477	1.103
	Yes	1.660	0.130	0.727
Season ticket	No	1.489	0.088	0.831
	Yes	0.616	0.701	0.922
Total		1.262	0.247	0.855

TABLE 3.3: Average number of car public transport and non-motorized trips conditional on mobility tool ownership. The ownership of a car or a season ticket corresponds to an increase in car or public transport trips, respectively.

Table 3.3 we present the average number of trips distinguished by mobility tool ownership. Intuitively, car ownership increases the number of car trips and reduces the number of public transport and non-motorized trips, while the opposite occurs for season ticket ownership. For season ticket ownership, we observe a slight increase in the number of non-motorized trips.

Besides the multimodal accessibility measure from Section 3.1, we include further spatial control variables and common socio-econometric control variables. Section C.1 in the Appendix summarizes all extracted variables as well as the sample's summary statistics in Table C.1. In particular, we use the spatial topology on a three level scale (living in the city, the agglomeration or countryside) and the quality of public transport at the household location on a five level scale(A to E) as further spatial control variables. The latter we use as an instrument for the structural effects of mobility tool ownership on travel activity.

3.4.2 *Model estimates*

Table 3.4 lists the model estimates for the independent land-use variables of interest, while Table 3.5 presents the additional parameters of the correlation matrix. For readability, the socio-economic control variables are not shown, but are given in Loder and Axhausen (2018). As we find most of these parameters statistically significant, we conclude that univariate mod-

els cannot provide the behavioral insights generated by the multivariate model. Further, we find all count parameters to be significantly different from zero. Therefore, the count models are between a traditional negative binomial and a Poisson count model.

Each of the three new derived measures of accessibility - general accessibility, comparatively better access by public transport and comparatively better job accessibility - show a negative effect on car ownership. The effects on season ticket ownership are positive for all three variables. We find, for gradually decreasing quality of public transport at household locations, likelihood of car ownership increases and likelihood of season ticket ownership decreases. In the agglomeration, car ownership is greater than in the urban center and the countryside, while car ownership is greater in rural areas than in the city center. Living in the city center shows a greater likelihood of subscribing to a season ticket than living in the agglomeration and the countryside.

For each of the three count outcomes of the number of trips we find significant structural effects of the two mobility tools. The observed differences for the number of car trips and season ticket ownership in Table 3.3 are replicated by the model estimates in Table 3.4, except for the effects of season ticket on the number of non-motorized trips. Table 3.3 show a slightly greater average of non-motorized trips for season tickets holder, but the effect in Table 3.4 is negative. This is anticipated because in Table 3.3 further control variables are not considered. The number of car trips increases in the countryside and even more in the agglomeration. The number of public transport trips is highest in the city center and decreases in the agglomeration and even more so in the countryside. This pattern is also observed for the number of non-motorized trips. With increasing general accessibility, the number of car trips declines and the number of public transport trips increases. The effect of general accessibility on the number of non-motorized trips is insignificant.

Table 3.5 lists the cross-equation parameters of all five outcomes. Except for the correlations between equations of season ticket ownership and number of non-motorized trips as well as between equations of the number of public transport trips and non-motorized trips, all correlations are significant. A positive correlation implies common unobserved factors act in a similar way, while a negative correlations implies that they act in the opposite way. In other words, a positive correlation indicates that both outcomes are complementary goods, while a negative correlation means substitute goods. Here, the negative correlation between car and season ticket

	Mobility tool ownership				Number of trips					
	Car		Season ticket		Car		Public transport		Non-motorized	
Quality of public transport at household location										
Level A: very good (base)										
Level B: good	0.163^{***}	(0.010)	-0.098^{***}	(0.011)						
Level C: moderate	0.307^{***}	(0.011)	-0.245^{***}	(0.012)						
Level D: low	0.411^{***}	(0.011)	-0.347^{***}	(0.012)						
Level E: very low	0.548^{***}	(0.012)	-0.473^{***}	(0.014)						
Spatial typology at household location										
City (base)										
Agglomeration	0.237^{***}	(0.007)	-0.174^{***}	(0.008)	0.205^{***}	(0.007)	-0.152^{***}	(0.011)	-0.319^{***}	(0.007)
Countryside	0.149^{***}	(0.010)	-0.164^{***}	(0.012)	0.132^{***}	(0.009)	-0.282^{***}	(0.021)	-0.342^{***}	(0.009)
General accessibility	-0.040^{***}	(0.002)	0.104^{***}	(0.003)	-0.051^{***}	(0.002)	0.173^{***}	(0.004)	a	
Comp. better accessibility by pub. transp.	-0.071^{***}	(0.005)	0.025^{***}	(0.005)						
Comp. better job accessibility	-0.572^{***}	(0.025)	0.848^{***}	(0.029)						
Log household income	0.438^{***}	(0.006)	0.040^{***}	(0.006)						
Car always available					0.973^{***}	(0.008)	-0.680^{***}	(0.013)	-0.431^{***}	(0.009)
Subscription to season ticket					-0.527^{***}	(0.009)	1.749^{***}	(0.014)	-0.053^{***}	(0.008)
Constant	-4.166^{***}	(0.047)	-0.747^{***}	(0.053)	-0.984^{***}	(0.012)	-2.438^{***}	(0.020)	0.514^{***}	(0.011)
Dispersion parameter θ					1.353^{***}	(0.012)	1.447^{***}	(0.050)	0.817^{***}	(0.007)
Flexibility parameter φ					0.105^{***}	(0.004)	0.319^{***}	(0.004)	0.319^{***}	(0.004)
Observations	52476									
Log likelihood at convergence	-983768									
Log likelihood constant only model	-1051388									
Pseudo R^2	0.064									

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
[a] Estimated, but not significant different from zero.

TABLE 3.4: Multivariate estimation results. Estimates are obtained using the MACML procedure. Socio-demographic control variables are not shown, but are given in Loder and Axhausen (2018).

Between the mobility tools			
ρ_{21}	Car and season ticket	-0.489^{***}	(0.007)
Between mobility tools and the number of trips			
ρ_{31}	Car and car trips	-0.022^{**}	(0.008)
ρ_{41}	Car and public transport trips	0.036^{***}	(0.010)
ρ_{51}	Car and non-motorized trips	0.016^{**}	(0.006)
ρ_{32}	Season ticket and car trips	0.028^{***}	(0.008)
ρ_{42}	Season ticket and public transport trips	-0.037^{***}	(0.009)
ρ_{52}	Season ticket and non-motorized trips	-0.006	(0.008)
Between the number of trips by			
ρ_{43}	Car and public transport	-0.355^{***}	(0.007)
ρ_{53}	Car and non-mot. modes	-0.281^{***}	(0.005)
ρ_{54}	Public transport and non-mot. modes	-0.013	(0.008)

Standard errors in parentheses
$^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

TABLE 3.5: Estimates of the correlation between unobserved factors.

ownership indicates that both are substitutes. This finding is intuitive and consistent with previous findings (Scott and Axhausen, 2006).

We expect that the counter-intuitive negative correlations between car ownership and use as well as season-ticket ownership and use might capture unobserved factors such as the impetus to use the mobility tool due to a large financial commitment. For positive correlations of structural effects, we assume they might describe a general factor of demanding mobility. We validate these counter-intuitive correlations with Poisson and linear regression models with endogenous covariates and the same sign for the correlations. Thus, we conclude that joint modeling is necessary to capture all of the unobserved effects.

3.5 SUMMARY

This chapter contributes with an approach to reduce highly-correlated accessibility measures of different transport modes to their most essential underlying structure. We then use the new multimodal accessibility measures to explain multimodal travel behavior in the dimensions of ownership and use in Switzerland.

We generally find that the land-use effects widely reported in literature are also present in Switzerland (Ewing and Cervero, 2010), but further we

find that our newly derived accessibility measures are strong predictors of travel behavior: The first dimension captures general accessibility and follows the universal law of people driving less in denser (more accessible) neighborhoods. The second dimension describes comparatively better accessibility by public transport and emphasizes that people choose those mobility tools that are of comparatively lower generalized cost of travel.

4

MODELING CONGESTION IN CITIES

> *The feeling that something should be done to mitigate the harmful effects of motor vehicles is almost universal, but the opinions on what should be done are widely conflicting. There are groups in our society who would like to travel everywhere by car, there are other groups who believe that nobody should be allowed to travel by private transport - at any rate in our larger towns - and there are groups with every variety of view in between.*
>
> — Smeed (1968)

Road traffic congestion is of considerable concern across the globe as nicely summarized by Smeed (1968) in the quote. Its physical and mathematical understanding is key to derive strategies to improve accessibility and thus mobility for everyone. The interest started in the era of motorization after the first world war, notably with the work by Schaar (1925) and Greenshields (1935). Both where interested in understanding how many vehicles a road can carry, i.e. in modern terms the capacity of a street. Their research lead to the foundation of what has now become the *Fundamental Diagram* (FD) of a street.

Since then, we have seen a substantial growth in the understanding of the physical mechanisms of congestion (e.g. Lighthill and Whitham, 1955; Underwood, 1961; Nagel and Schreckenberg, 1992; Treiber and Kesting, 2013; Helbing, 2001; Nagatani, 2002; Daganzo, 1994) which may lead us to the conclusions that we understand street level traffic and congestion. However, this is less the case for understanding urban traffic at the network level. In cities, traffic is often disturbed by, for example, traffic signals, spillback, buses, pedestrians and cyclists.

In the past fifty years, we have seen a few theories of collective vehicular traffic movement in cities at the network level (e.g. Smeed, 1961; Godfrey, 1969; Herman and Prigogine, 1979). The rationale of such theories is to predict the average network speed (and its distribution) as a function of demand and infrastructure constraints. It was only recently that this perspective on urban traffic regained interest, after almost two decades of mi-

croscopic research, with the formulation of the macroscopic fundamental diagram (MFD) by Daganzo (2007).

In this chapter, we review first in Section 4.1 the theories of urban network traffic prior to the MFD before we discuss the MFD throughoutly in Section 4.2. Thereafter, we present the contributions this thesis makes: The empirical evidence for the influence of network topology on the MFD in Section 4.3 and the proposal of three multimodal macroscopic congestion mechanisms based on the MFD in Section 4.4 that account for the influence of buses, bicycles and other disturbances in the flow of cars. Last. we summarize this chapter in Section 4.5.

4.1 MACROSCOPIC CONGESTION IN CITIES

The inevitable benefit of a macroscopic or network-wide perspective on urban traffic is that, compared to microscopic and link-based perspectives, it is usually much simpler in terms of mathematical complexity, but at the same time insightful for a long-term oriented policy decision making (Williams *et al.*, 1987; Williams, 2001).

Intuitively, with a robust speed-flow-density relationship on every link as seen in Figure 2.2, it is not surprising to expect a similar relationship also for an entire network (as observed by, e.g. Thomson, 1967; Olszewski *et al.*, 1995; Wong and Wong, 2016). The understanding of this "*mechanism of a road network*" (Godfrey, 1969) was initiated by traffic engineers in London, where first key questions were to identify the capacity of a network and how journey speeds are linked to the number of vehicles (Smeed, 1961, 1966; Wardrop, 1968; Godfrey, 1969). Slightly later, Zahavi (1972) proposed that all features of a read network (apart the network density for normalization, measured in length or area of roads per unit area) are captured in the single parameter α that defines the relationship between network-wide speed and flow. Importantly, Williams (2001) concludes that "*α may serve as a measure of the combined effects of the network characteristics and traffic performance, and can possibly be used as an indicator for the level of service*", from which we conclude that it would be more desired to disentangle network characteristics from traffic performance a network congestion mechanism. However, all these models consider a monotonically decreasing relationship between flow and speed, but "*monotonicity only makes sense if traffic is light, since it cannot capture crowded states with very low speeds and flows; e.g., approaching gridlock*" (Geroliminis and Daganzo, 2008). Consequently, these models are less reliable in the congested regime, but with our pressing con-

gestion problems, we require physically reliable mechanisms to identify the optimal resource allocation in cities.

A first macroscopic model that physically accounts for overcrowding is the *two-fluid theory* proposed by Herman and Prigogine (1979). In essence, their theory is that the average speed in the network is a function of the fraction of stopped vehicles, where the fraction of stopped vehicles is considered to follow a power law as a function of vehicle density. With traffic data becoming more available, Herman and Ardekani (1984), Ardekani and Herman (1985), and Ardekani *et al.* (1992) empirically show how the two-fluid parameters are likely to be influenced by network topology. However, "*the idea was not sufficiently developed to create a macroscopic model with variable inputs and outputs that could describe a rush hour dynamically*" (Geroliminis and Daganzo, 2008). In other words, the two-fluid model cannot be used for dynamic traffic models.

All mentioned approaches, except for the two-fluid model have been proposed by traffic engineers, while there has been - somehow unnoticed (or unreported) by the engineers - interest on understanding network traffic traffic by the physicist community. In particular, how the network structure, e.g. intersection spacing and network topology, influences the jamming transition in networks (e.g. Maniccam, 2003; Nagatani, 1993; Sasaki and Nagatani, 2003; Mendes *et al.*, 2014).

The last and seminal step that combines a simple macroscopic approach with a physical representation of congestion, captures the effect of network topology and models network dynamics has been taken by Daganzo (2005, 2007).

His approach relates the trip end rate or network exit rate to the accumulation of vehicles in the network. As the vehicle accumulation A inside a network can be managed or controlled, this approach is particularly useful for maximizing trip ends and thus directly helps to derive strategies to improve accessibility for everyone. This approach is centered around Eqns. 4.1 and 4.2. In detail, Eqn. 4.1 states that the change of vehicle accumulation A equals to the balance of the inflow to the system by demand $\Lambda(t)$ and the outflow of the system $O(A)$. Importantly, outflow $O(A)$ is considered to be only a function of A, but not t.

$$\frac{\partial A}{\partial t} = \Lambda(t) - O(A) \tag{4.1}$$

The outflow O - the trip ends or network exit rate - per unit time interval follows from Eqn. 4.2 and is the total travel production Π per unit time interval divided by the (average) trip length $\bar{l}$ in the network. Importantly,

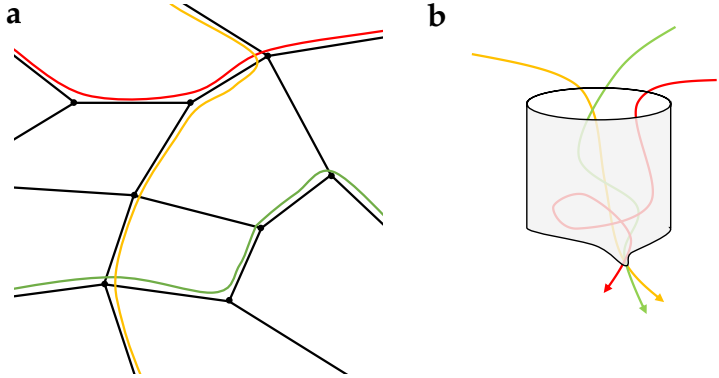


FIGURE 4.1: Comparing the modeling ideas for traffic routing in cities. (**a**) shows conventional routing in urban road networks, (**b**) shows the idea of routes in the MFD context (reservoir idea).

it is only the number of vehicles inside the network and the trip length that determine determine how many vehicles can exit the network due to congestion effects.

$$O(A) = \frac{\Pi(A)}{\bar{l}} \tag{4.2}$$

The relationships $O(A)$ and $\Pi(A)$ are defined as the *Macroscopic fundamental diagram* (MFD). The thinking of an urban network as a *reservoir* or bathtub with inflow and outflow with a MFD (Arnott, 2013; Fosgerau, 2015) is diametrical to most existing perspectives on (urban) traffic which consider vehicles always mapped to a sequence of connected links. Figure 4.1 illustrates the different modeling perspectives of how a vehicle propagates through the network. Figure 4.1**a** shows the common, conventional or traditional thinking of a vehicle being routed on links through the network, while Figure 4.1**b** shows that the network and route can be abstracted into a simple reservoir or system where the trip end rate and journey speeds are still key measures of the system.

In the remainder of this chapter, the focus is on the static behavior of $\Pi(A)$, the MFD. The interested reader is directed to the work by Mariotte *et al.* (2017), Aghamohammadi and Laval (2018) and Lamotte and Geroliminis (2018) for the dynamic aspects of Eqn. 4.1.

4.2 THE MACROSCOPIC FUNDAMENTAL DIAGRAM

This section introduces to the MFD in two ways. First, we provide a comprehensive review of the MFD theory and adjacent literature in Section 4.2.1. Second, we summarize the approaches to estimate analytically as well as empirically MFDs in Section 4.2.2.

4.2.1 *Theory*

The MFD as formulated with $\Pi(A)$ provides a well-defined link between travel production (and with $\bar{l}$ trip end rate or network exit rate) and the accumulation A of vehicles in the network. Intuitively, A, $O(A)$ and $\Pi(A)$ and their relationships are difficult obtain, neither empirically with data nor analytically. Therefore, it was the seminal contribution by Daganzo (2007) and subsequently by Geroliminis and Daganzo (2008) to show how $O(A)$ and $\Pi(A)$ can be approximated by a relationship between the network's average flow q and the average vehicle density k, where both measures and their relationship can be obtained[1]. Therefore, in the remainder of this thesis, we always refer to the MFD in average network flow q versus average network density k if not stated otherwise.

In detail, the MFD links the network average vehicle density k (in vehicles lane-km^{-1}) to the network average vehicle flow q (in vehicle h^{-1} lane^{-1}) with a smooth, concave and reproducible curve with a maximum flow q^*, defined as capacity, at the so called critical density k^*. In the following, we define that k^* and q^* mark the critical point of a network. If $k > k^*$, the network is congested and vehicular flow decreases; in turn, if $k < k^*$, the network is uncongested. Figure 4.2 shows MFD from Zürich, Switzerland, where we emphasize by color whether the network is congested or not. Accordingly, to maximize vehicle throughput, traffic control has to act at the critical point to avoid network-wide congestion (Daganzo, 2007; Haddad and Geroliminis, 2012).

In theory, the MFD shape is related to the underlying fundamental diagram, network topology, public transport operations, and origin and destination streams, i.e. route and mode choice (e.g. Boyaci and Geroliminis, 2011; Geroliminis and Boyaci, 2012; Leclercq and Geroliminis, 2013; Leclercq *et al.*, 2014; Castrillon and Laval, 2018). For a well-defined MFD to exist in an urban road network, more or less homogeneous traffic condi-

1 Assuming that the considered network has a total length of L, it follows $A = L \cdot k$ and $\Pi = L \cdot q$. For the outflow O then Eqn. 4.2 still holds

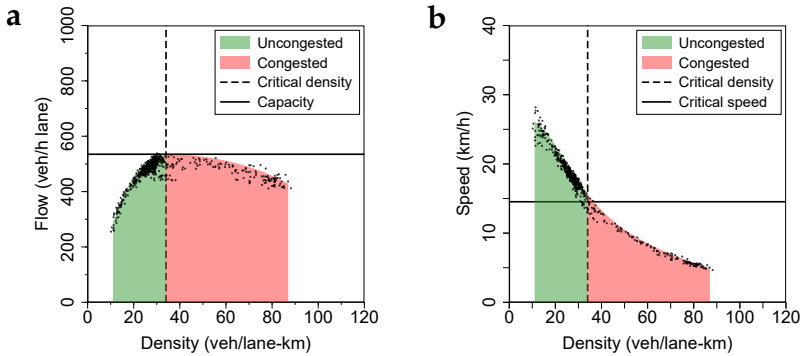


FIGURE 4.2: MFD from Zürich Wiedikon (see Figure B.2) in the flow density (**a**) as well as speed density (**b**) representation. The critical point and macroscopic traffic states are indicated.

tions in the network are required (Daganzo and Geroliminis, 2008; Geroliminis and Sun, 2011). Therefore, MFDs typically exist for multiple, smaller regional networks, around 10 km² in size, within a city. Regional networks are identified with network partitioning algorithms (Ji and Geroliminis, 2012; Saeedmanesh and Geroliminis, 2016, 2017), which, unfortunately, so far only apply to Lagrangian observations, i.e. trajectory or simulation data, but not to Eulerian observations, i.e. stationary detectors. Section 4.2.2 compares the MFD estimation for the different data sources.

The shape of the MFD can be modelled in different ways. While some only use curve-fitting with a polynomial function (e.g. Kouvelas *et al.*, 2017; Ramezani *et al.*, 2015) or exponential functions (e.g. Amirgholy and Gao, 2017; Ampountolas *et al.*, 2017) or step-wise linear functions (e.g. Mariotte *et al.*, 2017; Gao and Gayah, 2017). Helbing (2009) and Ambühl *et al.* (2018a) propose functional forms for the MFD with meaningful parameters that allow not only curve-fitting, bot also to estimate the MFD shape using additional measurements or estimates of the parameters instead. The function by Ambühl *et al.* (2018a) has only five parameters, where most relate to the method proposed by Daganzo and Geroliminis (2008).

Last, which applications aside of dynamically modeling traffic at the network level can be build or advanced using the MFD? First and foremost, due to the ease of identifying the critical point of a network, the MFD allows a variety of traffic control applications, in particular perimeter control approaches that aim to control traffic inside the perimeter at or before the

critical point to maximize vehicle or passenger throughput when considering buses as well (e.g. Yang *et al.*, 2018b; Aboudolas and Geroliminis, 2013; Kouvelas *et al.*, 2017; Haddad and Geroliminis, 2012; Haddad and Mirkin, 2017). Second, as the MFD further allows to measure the distance between the actual traffic state and the optimal or desired traffic state, while at the same time quantifying caused congestion externalities, it can be used to calculate optimal road pricing to improve mobility (e.g. Zheng *et al.*, 2012; Gonzales and Daganzo, 2012; Geroliminis and Levinson, 2009; Amirgholy and Gao, 2017). Third, as the MFD links network topology to the system performance, the MFD is a novel tool to discuss (multimodal) space allocation and optimal city and network design from a macroscopic perspective (e.g. Tsekeris and Geroliminis, 2013; Zheng and Geroliminis, 2013; Zheng *et al.*, 2017; Amirgholy *et al.*, 2017).

4.2.2 *Estimation*

At the time of writing this thesis, three different methods to estimate the MFD have been established in the literature (Leclercq *et al.*, 2014). In this section, we summarize the existing methods: One of these methods is an analytical method, i.e. the method of cuts in Section 4.2.2.1. The other two methods estimate the MFD from Eulerian observations in Section 4.2.2.2, Lagrangian observations in Section 4.2.2.3, and from simulation data in Section 4.2.2.4.

However, their seminal work, Daganzo and Geroliminis (2008); Geroliminis and Daganzo (2008) actually combined Eulerian with Lagrangian observations to improve the MFD estimation and to reduce the bias resulting from empirical measurements. This approach has subsequently been defined as *data fusion* in the MFD (e.g. Ambühl and Menendez, 2016; Ambühl *et al.*, 2017; Dakic and Menendez, 2018).

4.2.2.1 *Method of cuts*

The idea of the method of cuts is to approximate the upper bound of the MFD with a set of linear functions in the (k, q) plane. The method has been developed by Daganzo and Geroliminis (2008) and all subsequent mathematical formulations have been taken from Daganzo and Geroliminis (2008) and Leclercq *et al.* (2014) to which we refer the interested reader for the full description and discussion of this method. The linear functions are defined as *cuts* and are parameterized as given by Eqn. 4.3. Each cut corresponds to a moving observer that has no mass, no dimension and is

not delayed by traffic signals. In Eqn. 4.3, u is the associated speed of the moving observer in the network, while $R(u)$ is the associated maximum passing rate that a particular moving observer can encounter at speed u.

$$q = ku + R(u) \tag{4.3}$$

The upper bound notion of the MFD means that all observed flow traffic states must be less or equal to this upper bound as given by Eqn. 4.4.

$$q \leq ku + R(u) \tag{4.4}$$

Instead of evaluating $R(u)$ for all possible values of u, Daganzo and Geroliminis (2008) proposes to use discrete values of u and thus obtaining three families of "*practical*" cuts. The cuts are based on observers that either do not move at all $u = 0$, move forward $u > 0$ and move backward $u < 0$. The first family of cut relates to stationary observers and the practical cut is the observer standing, i.e. $u = 0$, at the most constraining intersection as formulated in Eqn. 4.5, where s_i is the maximum discharge rate of intersection i with cycle length C_i and green time G_i.

$$q \leq \min_i \left(s_i \frac{G_i}{C_i} \right) \tag{4.5}$$

The second family relates to forward moving observers. Here, $R(u)$ is evaluated at γ different speeds, where γ_{max} is the maximum number of signals a fast moving observer can pass without stopping at extended red light phases. The resulting u_γ is calculated as if the observer is moving with free flow speed u_f, but needs to stop at every γ-th signal, where the fraction of time being stopped at that signal is f_γ (see the Appendix in Daganzo and Geroliminis (2008) for exact formulae for u_γ and f_γ). Then the cuts for the forward moving observer is given by Eqn. 4.6, where q_m is the capacity of the associated fundamental diagram.

$$q \leq ku_\gamma + q_m f_\gamma \tag{4.6}$$

The third and last family of cuts is for the case when the observer is moving backward. This case is very similar to the second family of cuts, but where the observer is now moving at speeds $u = -w_\gamma$, i.e. in the opposite direction of the traffic stream. Then, the third families of cuts is given by Eqn. 4.7 where b_γ is the fraction of time the observer is stopped at signals, while w_γ / w is the fraction of time the observer is moving and r is

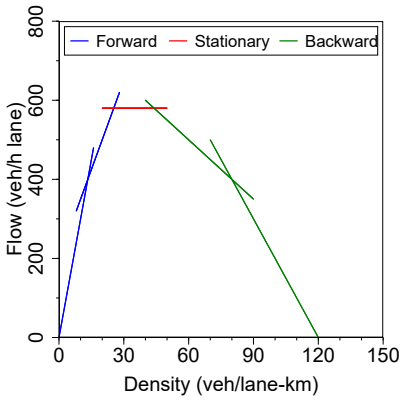


FIGURE 4.3: The three cut families: forward, stationary and backward cuts

the maximum passing rate when the observer is moving (see the Appendix in Daganzo and Geroliminis (2008) for exact formulae for w_γ and b_γ).

$$q \leq -kw_\gamma + q_m b_\gamma + r\frac{w_\gamma}{w} \tag{4.7}$$

In Figure 4.3 we show the cuts and thus the creation of the upper bound and clearly shows the familiar shape of a MFD as given, for example, in Figure 4.2. Although Daganzo and Geroliminis (2008) show that their method is also applicable in heterogeneous networks, Leclercq and Geroliminis (2013), Leclercq *et al.* (2014) Geroliminis and Boyaci (2012) and Boyaci and Geroliminis (2011) clearly emphasize that in complex and heterogeneous networks the estimation of cuts is not trivial anymore and the recovery of the parameters is difficult, even impossible.

To overcome the issue of inhomogeneous corridors with the analytical approximation, Laval and Castrillón (2015) proposes a stochastic approximation to the MFD when the block length and signal timings are considered random variables with mean and standard deviation. The authors conclude that the MFD's shape can be well described using the mean block length to green ratio and the mean red to green ratio.

4.2.2.2 *Eulerian observations*

The Eulerian observations of vehicular flow or in general any particle flow in motion are measurements from a specific, i.e. fixed, location in space

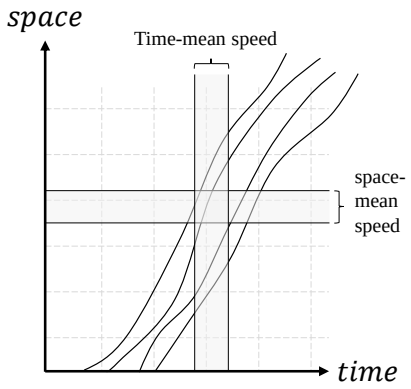


FIGURE 4.4: Comparing Eulerian and Lagrangian observations or speed measurements. This figure is adopted from Bliemer (2001).

bounded by $[x, x + \Delta x]$. Figure 4.4 illustrates the measurement procedure: During an observation interval $[t, t + \Delta t]$, a certain number of vehicles or particles flows through that window and spends a certain time inside this space-time window. In traffic, these observations result from fixed or stationary detectors (see Appendix A.1.1 for further details), e.g. inductive loop detectors or ultrasonic detectors. This data has been frequently been used in empirical MFD estimation (e.g. Geroliminis and Daganzo, 2008; Buisson and Ladier, 2009; Loder *et al.*, 2017) (also in Figure 4.2). In a network with links i of length l_i, the MFD from Eulerian observations is then estimated by averaging flows q_i and densities k_i of all links weighted by their length as given by Eqn. 4.8.

$$q_{MFD} = \frac{\sum q_i l_i}{\sum l_i} \tag{4.8}$$

The density is estimated in a similar way as given by Eqn. 4.9.

$$k_{MFD} = \frac{\sum k_i l_i}{\sum l_i} \tag{4.9}$$

Last, we obtain the time-mean speed of the MFD by applying the fundamental traffic equation as given by Eqn. 4.10 (Daganzo, 1997).

$$v_{MFD} = \frac{q_{MFD}}{k_{MFD}} \tag{4.10}$$

Importantly, stationary traffic detectors have their own issues as they measure only the traffic conditions at their exact location in space and only measure occupancy for technical reasons instead of reporting density (Bickel *et al.*, 2007). Thus, we refer the reader to Appendix A.1.1 for a detailed summary of measuring traffic with stationary detectors.

4.2.2.3 *Lagrangian observations*

Contrary, Lagrangian observations of vehicles or particles follow an individual vehicle through space and time. In traffic research, this corresponds to the trajectory of the object (see Appendix A.1.2 for further details) and is frequently provided by taxis and other fleet-operating vehicles. This data source has also already been used in empirical MFD estimation (e.g. Ji *et al.*, 2014). Consequently, for each vehicle j the travel distance d_j as well its travel time τ_j inside the space-time windows defined by by $[x, x + \Delta x]$ and $[t, t + \Delta t]$ is known. Then, the MFD space mean speed is simply obtained by dividing the total travel time of all vehicles by the total distance traveled of all vehicles.

$$v_{MFD} = \frac{\sum_j d_j}{\sum_k \tau_j} \tag{4.11}$$

The density is then obtained by the total travel time of all vehicles divided by the length of the observation interval Δt and the total monitored network length L.

$$k_{MFD} = \frac{\sum_k \tau_k}{L \delta t} \tag{4.12}$$

Last, we obtain the MFD flow by applying again the fundamental traffic equation as given by Eqn. 4.10. Importantly, usually not vehicles provide trajectory data, which might not be a substantial problem for the speed estimation, but it is for the density and flow estimation. Consequently, additional information is required to scale flow and density accordingly with the probe penetration rate.

4.2.2.4 *Edie's definition*

In the special case of obtaining data from a traffic simulator, all vehicle trajectories are available and thus the unbiased MFD can be estimated using Edie's generalized definitions Edie (1963). This approach follows two steps (Leclercq *et al.*, 2014). First, from the trajectories, the link averages of flow

q_i and density k_i for link i in observation interval $[t, t + \Delta t]$ are calculated as given by Eqns. 4.13 and 4.14.

$$q_i = \frac{\sum_j d_j}{l_i \Delta t} \tag{4.13}$$

$$k_i = \frac{\sum_j \tau_j}{l_i \Delta t} \tag{4.14}$$

Second, we then average the link flows as in case of Eulerian observations with Eqns. 4.8 and 4.9.

4.3 UNDERSTANDING TRAFFIC CAPACITY OF URBAN NETWORKS

The MFD theory predicts that the shape of the MFD, and thus ultimately networks' capacities depend on the network topology (Daganzo and Geroliminis, 2008; Leclercq and Geroliminis, 2013; Geroliminis and Boyaci, 2012). This is intuitive and not surprising as for many other macroscopic flow models, the influence was not only theoretically predicted (e.g. Smeed, 1961), but also empirically proven (e.g. Wong and Wong, 2016; Ardekani *et al.*, 1992). So far, however, the empirical evidence for the MFD and thus the empirical quantification of the effects is still missing. Therefore, in this section we provide this first comparative empirical evidence of such relationship.

We analyze critical points, with data from 41 cities worldwide, including 107 regional networks (see Table B.3). We use a newly assembled, rich empirical data set of billions of vehicle observations from stationary traffic detectors (i.e. Eulerian observations as introduced in Section 4.2.2.2; see Appendix A for further information) to estimate critical points. We pair this data with networks from OpenStreetMap to explain how network topology drives the critical point.

The estimation of critical points requires first the identification of suitable regional networks where the MFD, critical points as well as the network topology are estimated and analyzed. We define regional networks heuristically, as existing partitioning algorithms (e.g., Ji and Geroliminis, 2012; Ji *et al.*, 2014; Saeedmanesh and Geroliminis, 2016, 2017) cannot be applied to our data and limited detector coverage usually prevents a systematic and automatic zoning at large scale across all cities. Second, we estimate MFDs using a re-sampling methodology to account for heterogeneity in the network because we want to clearly identify the boundary

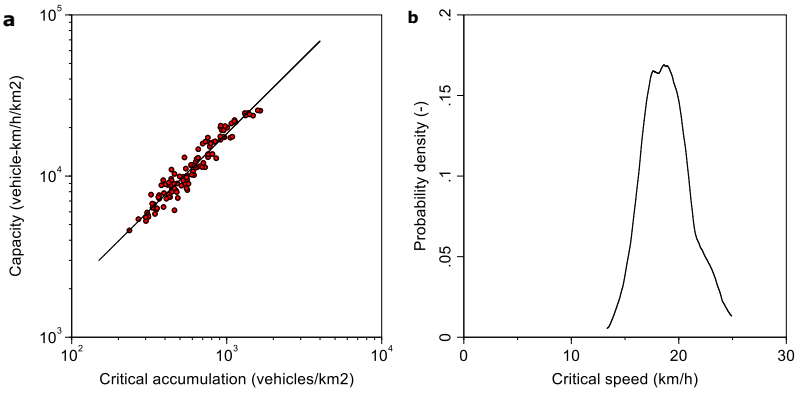


FIGURE 4.5: Comparison of urban traffic capacity across 107 urban networks. (**a**) Positive relationship of network capacity against critical accumulation (R^2=0.92). (**b**) Kernel density estimate of critical speeds of all networks emphasizes that congestion starts at different speeds in different networks (mean 18.8km/h).

of traffic states as independent as possible of demand and traffic irregularities (Ambühl *et al.*, 2018b). Third, we fit an MFD function18 in the uppermost 20-quantile (quantile regression approach) of the estimated MFDs to functionally describe the boundary of traffic states (Ambühl *et al.*, 2018a). The critical point is then the maximum of the estimated function. This allows a more robust estimation of the critical point because this procedure borrows information from all observations in the MFD.

Figure 4.5 compares the critical points of all 107 networks with sample statistics $\overline{\Pi}^* = 12236 \pm 551$ vehicle-km h^{-1} km^{-2} (mean $\pm$ standard deviation) and $\overline{A}^* = 653 \pm 308$ vehicles km^{-2} (mean $\pm$ standard deviation). Figure 4.5**a** shows the intuitive positive relationship between critical accumulation and network capacity with $R^2 = 0.91$. $R^2 < 1$ implies – ignoring measurement uncertainties - that other factors create variation in critical points as well. This variation is also emphasized by the kernel density estimate of the critical speed in Figure 4.5**b**, revealing a substantial variation of critical points across regions. This implies that in some regions, car drivers can experience low speeds, but against intuition, the network is not macroscopically congested. The distribution of critical accumulation in our sample is consistent with the traffic physics literature (Chowdhury *et al.*,

2000) that predicts that critical accumulation does not exceed one third of the jam accumulation.

To quantify the critical point relationship in Figure 4.5**a** and the influence of bus and road network topology on the critical point, we use a two-stage least squares regression analysis because the influence of critical accumulation on capacity is endogenous and requires instrumental variables. We focus on exogenous variables linked to the full network properties of bus and road network topology that are not only macroscopic, but also measurable from OpenStreetMap, common to all cities, and clearly interpretable. We do not consider city-specific factors, e.g. bicycle policies, although they undoubtedly also influence the critical point. On a broader level, the exogenous variables should describe conflicts between vehicles that cause delays. For example, delays appear when two vehicles want to be at the same time at the same place (i.e. at intersections), or when cars must follow buses or cyclists. All these conflicts influence the total travel time of drivers or even the travelled distance of vehicles, both of which affect the location of the critical point. Accordingly, we define four variables. First, road network density R (lane-km km^{-2}) with sample statistics $\overline{R} = 23.46 \pm 10.36$ lane-km km^{-2} (mean $\pm$ standard deviation). Second, network redundancy, measured in the network average betweenness centrality b_c (-) with sample statistics $\overline{b_c} = 0.084 \pm 0.039$ (mean $\pm$ standard deviation). Third, the average distance between or spacing of signalized intersections (LSA) in the network I (LSA network-km^{-1}) with sample statistics $\overline{I} = 0.385 \pm 0.188$ LSA network-km^{-1} (mean $\pm$ standard deviation). Fourth, the bus production density B (bus-km h^{-1} km^{-2}) that combines the density of the bus network per unit area with the average headway with sample statistics $\overline{B} = 147 \pm 85.5$ bus-km h^{-2} km^{-2} (mean $\pm$ standard deviation). We expect that R, b_c and I affect the critical accumulation (via total travel time) and B affects capacity (via total travelled distance). Consequently, the first three variables correspond to the required instruments for the critical accumulation. In particular, we expect that A^* increases with R as more road space per unit area generally allows more vehicles to circulate, but denser networks provide more opportunities for conflicts (at intersections, through lane changes, etc.), resulting in a sublinear scaling (Smeed, 1968; Nagatani, 1993; Miller, 1970) between R and A^*. In low redundancy networks with few alternative routes, the concentration of vehicles on these routes reduces the critical accumulation compared to networks with many alternative routes (Maniccam, 2003; Mendes *et al.*, 2014; Ortigosa and Menendez, 2014), i.e. A^* decreases with b_c. In a network of density R, we expect that

a denser intersection spacing increases travel times due to more waiting compared to a network of the same density R with a larger intersection spacing (Transportation Research Board, 2016; Ardekani *et al.*, 1992; Geroliminis and Boyaci, 2012; Wong and Wong, 2016), consequently we expect that the interaction of I and R has a negative effect on Π^*. Last, buses are usually larger than cars and can behave as either stationary bottlenecks at bus stops, or as moving bottlenecks when driving, consequently, Π^* should decrease with B (Loder *et al.*, 2017; Nagai *et al.*, 2005; Castrillon and Laval, 2018; Geroliminis *et al.*, 2014).

Table 4.1 shows the model estimates. We have to reject the hypothesis that critical accumulation is exogenous, consequently, the two stage least squares is required. We find that we can significantly explain around 90 % of the observed critical point variation with just four exogenous variables describing bus and road network topology. We cannot explain the entire variation because we do not include city-specific factors, nor those that are not clearly measurable. Further, we see that the elasticity of A^* is not statistically significantly different from one, meaning that efforts to increase the critical accumulation affect capacity proportionally. Contrary, we find that the elasticity of R on the critical accumulation is significantly different from one, with $\varepsilon \approx 0.8$.

In Figure 4.6, we illustrate the effects underlying the results from Table 4.1 graphically to show how the variation in the data drives the critical point variation. We conclude that even with the rather small empirical variation in bus and road network topology across cities, there is substantial variation in the critical point.

The results emphasize macroscopic trade-offs for urban transportation policies. Clearly, Figure 4.6**a** shows that more roads in the network increase the total production of the network, but the elasticity of road network density from Table 4.1 points towards the important policy implication of decreasing marginal returns of road network expansion. Figures 4.6**b-c** describe the macroscopic trade-off between vehicular and non-vehicular traffic and walkability of a city. More intersections provide more right of way to pedestrians and a concentration of vehicular traffic on few routes could mean less negative car externalities for non-vehicular modes elsewhere in the network. Figure 4.6**d** emphasizes a universal trade-off between buses and cars allowing cities to optimize passenger throughput under further assumptions on vehicle occupancy levels and bus services. The results from Table 4.1 and the illustrations in Figure 4.6 imply that cities can now un-

derstand how investment measures benefit or harm the capacity of entire networks. Importantly, these measures go beyond just building roads.

Our findings have practical implications. First, the possibility to identify a city's optimal infrastructure may help to attain the full benefits of urban scaling (Bettencourt *et al.*, 2007). Second, the estimates provide means for discussing urban space allocation. However, our study also faces limitations. First, our sample is still relatively small and mostly located in Europe, limiting our findings' (geographic) validity. Second, we cannot consider some topological and operational features (e.g. intersection design, transit priority) as they are either unavailable or not clearly measurable. Last, we do not cover traffic control. However, since control strategies are typically derived from between-vehicle conflicts, which we do consider, we expect to explain many traffic signal effects with our four variables (Webster, 1958; Daganzo and Geroliminis, 2008; Geroliminis and Boyaci, 2012; Laval and Castrillón, 2015). Future research can use the collected data to analyze the dynamics of network traffic, in particular to investigate at network level the determinants of when they reach congestion and for how long cities stay congested.

In closing, the variables derived from bus and road networks, significantly explain the critical point and the traffic capacity of urban networks, making it predictable. This has profound implications for transportation investments. Cities can now identify their optimal infrastructure level and the macroscopic effects of investment decisions on the (multimodal) network performance. This new understanding of traffic capacity of urban networks may not solve congestion problems, but it is crucial for the development of new strategies to improve traffic.

4.4 MODELING MULTIMODAL INTERACTIONS AT THE NETWORK LEVEL

In cities, car traffic streams are rarely homogeneous as they are frequently disturbed by other modes of transportation, notably pedestrians, buses, motorbikes and bicycles. Geroliminis *et al.* (2014) proposed to account for the bus-car interactions in the MFD by adding a *third* dimension to the MFD, to obtain the so called 3D-MFD. Typically, in the 3D-MFD, the accumulation of cars A_c and the accumulation of buses A_b are related to the total travel production Π of vehicles or passengers (Geroliminis *et al.*, 2014; Loder *et al.*, 2017). Note that the accumulation has units of vehicles, whereas the travel production has units of either vehicle or passenger kilometers per unit time.

		Dependent variables			
		Critical accumulation A^* (veh km^{-2})		Capacity Π^* (veh-km h^{-1} km^{-2})	
Covariates	Unit	β	t-value	β	t-value
Constant		-292.73	(-3.46)	1282.29	(1.56)
A^*	(veh km^{-2})			18.63	(18.87)
R	(lane-km km^{-2})	24.40	(15.31)		
$R \times I$	(lane-km km^{-2} $\times$ LSA km^{-1})	-5.34	(-2.00)		
b_c	(-)	-140.67	(-4.02)		
B	(bus-km h^{-1} km^{-2})	0.307	(2.12)	-6.64	(5.19)
N		107		107	
Adj. R^2		0.88		0.91	

Elasticity ε	95 % CI	$H_0 : \varepsilon = 1$		
$\partial \log \Pi^* / \partial \log A^* = 0.968$	[0.828;1.10]	F(1,40) = 0	Prob > F = 0.6	not reject
$\partial \log A^* / \partial \log R = 0.8$	[0.727;0.95]	F(1,40) = 215.60	Prob > F = 0.0	reject

TABLE 4.1: The critical point model. All effects are statistically significant at least at the 5 % level of statistical significance. The probability that A^* is exogenous is $p = 0.000$.We cluster standard errors at the city level and weigh observations by a measure of certainty in the critical point estimation. As we cluster standard errors, the test for endogeneity is based on Wooldridge's robust score test (the equivalent in case of unadjusted standard errors would be the Durbin-Wu-Hausman test). The lower part of the table shows elasticity estimates (Delta-method) with the 95 % confidence intervals between critical accumulation and capacity as well as road network density and accumulation. For each elasticity, we test the hypothesis of being different from one.

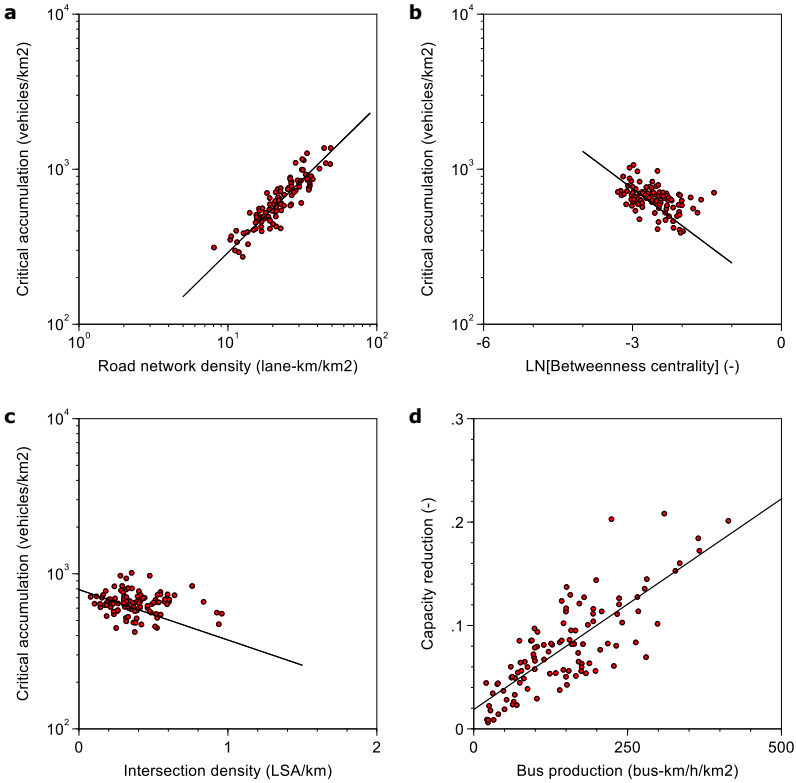


FIGURE 4.6: Illustration of the effects in the critical point model. (**a**) Influence of road network density on critical accumulation. (**b**) Influence of betweenness centrality (network redundancy) on critical accumulation. (**c**) Influence of intersection density on critical accumulation. (**d**) Capacity reduction by bus operations in the network. (**a**)-(**c**) are constructed as follows. We first residualize the x- and y-variables on the other control variables of the critical point model as given in Table 1. Secondly, we add back the unconditional mean of the x- and y-variables to the residuals. (**d**) is constructed by calculating each network's capacity at zero bus production and then calculating the relative difference to the observed capacity. The solid black line in each panel corresponds to the linear fit of the scatterplots.

So far, the understanding of the other modes' local interaction effects on the MFD is growing, but mostly for buses, descriptive and generated from simulations (e.g. Boyaci and Geroliminis, 2011; Castrillon and Laval, 2018; Zhang *et al.*, 2018), except for pedestrianized streets (e.g. Daganzo and Knoop, 2016). Therefore, we propose in this section three approaches to estimate the 3D-MFDs as a multimodal congestion mechanism for the system perspective of urban traffic: (i) an empirical approach in Section 4.4.1, (ii) an analytical, but geometric approach in Section 4.4.2, and (iii) a Bose-Einstein condensation approach in Section 4.4.3.

4.4.1 *An empirical approach*

In this empirical approach, we focus on the city of Zürich. The data is described in Section B.1 and the two regions (*City center* and *Wiedikon*) we analyze are given in Figure B.2 in the Appendix.

The overall Zürich network can be split into three parts. In the first part, only cars circulate, in the second part only public transport vehicles circulate, and in the third part both modes share the same road space. It is important to note that even when both modes share the same link they do not necessarily share the same lane. Consequently, the total network length is not simply the sum of the car network length and the public transport network length. The total length of primary road network is 39 km in the City Center and 31 km in Wiedikon. In the City center, the public transport network length includes 22 km for trams and 12 km for buses (approx. 75 % dedicated), while in Wiedikon, 7 km are for trams and 18 km for buses (approx. 60 % dedicated). Given that both regions have a high share of dedicated lanes for public transport, interactions between both modes occur particularly at intersections and curbside stops.

The interactions between the private and public mode affect the speeds of both modes. Thus, we compare the speed of both modes - aggregated into 15 min intervals to avoid systematic variation due to the relatively rigid timetable in Zürich - in Figure 4.7**a** during Tuesday, October 27th 2015. In the City center, we observe that the car speeds drop during the morning peak to 11 km/h and during the afternoon peak to 9 km/h, but recover during the day up to 13-14 km/h. The speed of public transport vehicles is around 12 km/h all the time, but decreases slightly to around 10 km/h during the afternoon. In Wiedikon, speeds of both modes are generally higher than in the City center and show a greater range, where in

This section is taken in large parts from Loder *et al.* (2017).

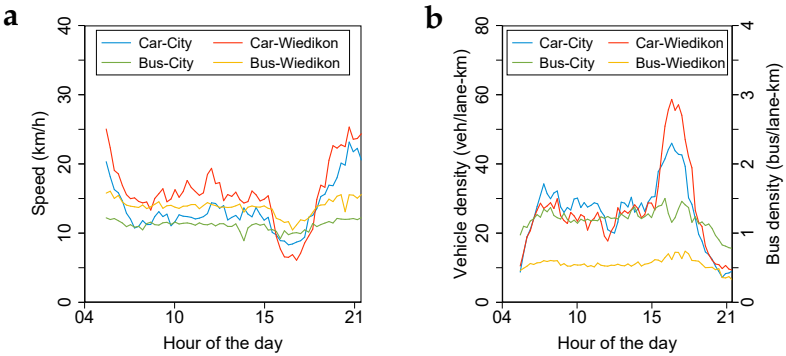


FIGURE 4.7: Bi-modal relationships in the city of Zürich. (**a**) compares the speeds and (**b**) compares the vehicle densities. Recall that the speeds of public transport vehicles contain the dwell times.

the afternoon, car speeds drop remarkably below public transport speeds. We use the vehicle densities instead of total vehicle accumulations, to normalize for the different regions. We compare in Figure 4.7**b** the vehicle densities of both modes over time, where the effect of the rigid timetable can be clearly seen.

In Figure 4.8, we present the observed 3D-MFDs for the City center and Wiedikon. The horizontal plane represents the accumulation of both modes and the vertical axis the total production. We calculate total production by adding the production of each mode. Figure 4.8 shows trends similar to those found in Geroliminis *et al.* (2014) for the 3D-MFD using simulation data, but our data does not exhibit similar experimental variation because it is empirical, especially for public transport densities due to the constant timetable (see also Figure 4.7**b**). However, we can still compare the two similarly-sized regions. In general, we register higher public transport accumulation for the City center than for Wiedikon, due to the very high number of public transport lines in the City center, and the longer public transport network.

We use a statistical model to describe the available, but limited data. We use both modes' speeds, i.e. car speed, v_c, and bus speed, v_b, as dependent variables. As we do not observe a full range of public transport densities, we cannot use the exponential function proposed by Geroliminis *et al.* (2014) because we must arbitrarily define extreme values to fit the curve.

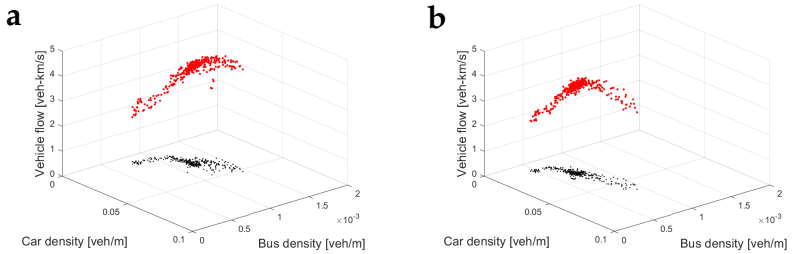


FIGURE 4.8: 3D-MFDs in Zürich. (**a**) is from the region *City center*, and (**b**) is from the region *Wiedikon*. The red points indicate the measurements. The black points indicate the projection of total production onto the density plane.

Therefore, for simplicity and as a first order approach, we model v_c with Greenshields's 1935 linear model between density and speed that has also been applied at the network level (Mahmassani *et al.*, 1987). Thus, the first equation as defined in Eqn. 4.15 links car speeds v_c to the free flow speed of cars $\beta_{c,0}$, the density of cars k_c, and the density of public transport vehicles k_b. β_c and β_b are coefficients to be estimated and represent the marginal effect of each mode on car speeds.

$$v_c = \beta_{c,0} + \beta_c k_c + \beta_{pt} k_b \tag{4.15}$$

The speed of public transport v_b is modeled as a linear function of v_c, as proposed by Geroliminis *et al.* (2014) and Zheng and Geroliminis (2013) as defined in Equation 4.16. Thus, public transport speeds do not explicitly depend on vehicles' densities k_c and k_b. However, v_c is a function of vehicle densities and therefore v_b depends implicitly on vehicle densities as well. The coefficients to be estimated are $\beta_{b,0}$ and $\beta_{c,b}$. $\beta_{c,b}$ captures the aspect that public transport vehicles typically move slower than cars due to frequent stops, and $\beta_{b,0}$ adjusts for the fact that public transport speeds might exceed car speeds during congested times due to dedicated lanes. This effect is observed in both regions during the evening peak, see Figure 4.7**a**.

$$v_b = \beta_{c,b} v_c + \beta_{b,0} \overset{Eq.4.15}{=} \beta_{c,b}(\beta_{c,0} + \beta_c k_c + \beta_b k_b) + \beta_{b,0} \tag{4.16}$$

In Table 4.2 we show the model results. All estimates are statistically significant and show the expected signs. Wiedikon shows a slightly higher

	City center		Wiedikon	
	Car	Public transport	Car	Public transport
Car density (veh/km)	-0.288		-0.345	
Bus vehicle density (veh/km)	-5.659		-6.726	
Car speed (km/h)		0.116		0.197
Constant (km/h)	27.933	9.574	28.448	10.794
R^2	0.96	0.49	0.94	0.79
N	360	360	360	360

TABLE 4.2: Model estimates for the 3D-MFD for the data from Zürich (5 weekdays, 06:00-24:00, 15min intervals; total 360 observations) estimated with OLS. The dependent variables are the space-mean speeds of cars and public transport vehicles. All estimates are significant at 1 % level of significance.

free flow speed than the City center and a marginal effect of car density in the same order of magnitude. The higher free flow speed can be attributed to longer links and fewer signalized intersections in Wiedikon. Comparing the marginal effects of public transport vehicles, we find a decrease in speeds of 5.7 km/h per additional public transport density of 1 veh/km in the City center whereas this decrease amounts to 6.7 km/h in Wiedikon. At a density of one public transport vehicle per kilometer, the elasticity of public transport vehicles on car speeds in the City center is -0.4 and in Wiedikon -0.7. We expect that the lower impact of public transport vehicles in the City center is due to the larger share of dedicated lanes. Considering public transport speeds, we observe 10.8 km/h in Wiedikon and 9.6 km/h in the City center for the constant $\beta_{b,0}$ and around 0.12 km/h change in public transport speeds for 1 km/h change in car speed in the City center and 0.2 km/h in Wiedikon. These findings are in line with Figure 4.7**a**.

With the model estimates for the 3D-MFD we can then predict the full shape of the MFDs of both regions. For this, the speeds $\hat{v}_c$ and $\hat{v}_b$ are calculated over a range of densities k_c and k_b. We obtain the total production as the sum of both modes' productions[2] where L is the mode's respective infrastructure length, i.e. $A_c \cdot \hat{v}_c + N_b \cdot \hat{v}_b$. Then, Figures 4.9**a** and 4.9**b** show the predicted 3D-MFD. It should be clear that our predicted MFDs are more accurate for points closer to, or within the region of the empirical

2 We obtain the production of each mode by $v\ k\ L$ = v N

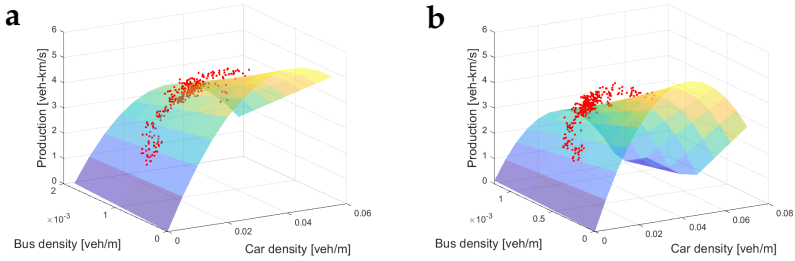


FIGURE 4.9: Predicted 3D-MFD shape. The red dots represent the range of empirical observations. (**a**) is from the region *City center*, and (**b**) is from the region *Wiedikon*.

data. Any extrapolation of our model beyond the observed range of data should be considered with caution. Nevertheless, our model has a strong indicative character and gives a first empirical insight regarding the interactions of multimodal traffic at a network level. The model is intuitive and when we compare the results of our extrapolated 3D-MFD with the ones from simulations, such as the one from San Francisco, estimated by Geroliminis *et al.* (2014), we see similar trends. Both, the simulation and our prediction model show concave curves with a maximum vehicular production for zero public transport accumulation, and a similar skewness. Arguably, the curves' shapes are mainly determined by the functional form of the model.

4.4.2 *A geometric approach*

The idea of the geometric approach is to find a lower envelope or functional form for the 3D-MFD that *encapsulates* the observations from the observed 3D-MFD as shown in Figure 4.8. In other words, it provides an upper limit for the travel production, for any given combination of accumulations of both modes. Here, we first introduce in Section 4.4.2.1 to the new functional form in general. Then, in Section 4.4.2.2 we apply this general functional form to the 3D-MFD case. Thereafter, we show how to derive each mode's speed from the functional form in Section 4.4.2.3 and discuss how the lower envelope can be estimated in Section 4.4.2.4. We end this section with a validation of the lower envelope in Section 4.4.2.5 and the proposed speed functions in Section 4.4.2.6.

4.4.2.1 *A new functional form for the multimodal macroscopic fundamental diagram*

The proposed geometric approach to derive a functional form for multimodal 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 mode's vehicle accumulation is A_m, and we define that $\mathbf{A}$ in $\mathbb{R}_+^M$ is the vector of vehicle accumulations with elements $(A_1; A_2; \ldots; A_M)$. 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 $\mathbf{A}$. 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 A_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 $\Pi_{j \in J} \colon \mathbf{A} \rightarrow \Pi$, where each Π_j provides a physically meaningful constraint of one feature of multimodal traffic (e.g. gridlock, saturation) relating $\mathbf{A}$ 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. 4.17. Each Π_j function can be seen as a hyperplane of $M+1$ dimensions.

$$\Pi(\mathbf{A}) = \min\left(\Pi_1(\mathbf{A}); \Pi_2(\mathbf{A}); \ldots; \Pi_J(\mathbf{A})\right) \tag{4.17}$$

The definition of functions $\Pi_j(\mathbf{A})$ is context specific (e.g. in our application to the 3D-MFD in Section 4.4.2.2, we use seven linear functions, i.e. planes, to create the lower envelope). Nevertheless, the function resulting from Eqn. 4.17 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 $\mathbf{A}$ across all modes, because different modes might share a portion of the infrastructure, η_{mixed}. Eqn. 4.18 shows this constraint for κ_m. Assuming l_m is the jam spacing for mode m, then $\kappa_m(\mathbf{A})$ 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 $\frac{L}{l_m}\eta_m \leq \kappa_m \leq \frac{L}{l_m}(\eta_m + \eta_{mixed}), \forall m \in M$.

$$\kappa_m(\mathbf{A}) = \frac{L}{l_m}(\eta_m + \eta_{mixed}) - \sum_{m' \in M \setminus \{m\}} \frac{l_{m'}}{l_m}\left(\kappa_{m'} - \frac{L}{l_{m'}}\eta_{m'}\right), \quad \forall m \in M \tag{4.18}$$

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. 4.19.

$$\Pi(\mathbf{A}) > 0 \quad \text{for } 0 < A_m < \kappa_m, \quad \forall m \in M \tag{4.19}$$

Third, we define the average network speed as $V \equiv \Pi / \sum_m A_m$, so that it must always decrease with any accumulation of vehicles, as given by Eqn. 4.20.

$$\frac{\partial V}{\partial A_m} < 0, \quad \forall m \in M \tag{4.20}$$

Fourth, we require that the production is zero, i.e. $\Pi(\mathbf{A}) = 0$, 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. 4.21.

$$\Pi(\mathbf{A}) = 0, \text{if } A_m = \kappa_m \vee A_m = 0, \quad \forall m \in M \tag{4.21}$$

Fifth, we require that that the total production Π (in Eqn. 4.22) equals the sum of all modes' productions π_m. As Eqn. 4.17 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.

$$\Pi\left(\mathbf{A}\right) = \max \sum_{m \in M} \pi_m\left(A_m\right) \tag{4.22}$$

Eqns. 4.17-4.22 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 $\lambda\left(\mathbf{A}\right)$, as formulated in Eqn. 4.23 with the smooth approximation of the minimum operator (Cook, 2011; Bliemer *et al.*, 2017; Ambühl *et al.*, 2018a).

$$\Pi(\mathbf{A}) = -\lambda(\mathbf{A}) \log\left(\exp\left(-\frac{\Pi_1(\mathbf{A})}{\lambda(\mathbf{A})}\right) + \exp\left(-\frac{\Pi_2(\mathbf{A})}{\lambda(\mathbf{A})}\right) + \ldots \right.$$
$$\left. \cdots + \exp\left(-\frac{\Pi_J(\mathbf{A})}{\lambda(\mathbf{A})}\right)\right) \tag{4.23}$$

The smoothing parameter $\lambda\left(\mathbf{A}\right)$ 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 $\lambda\left(\mathbf{A}\right)$ is, the stronger the vehicle interactions become, resulting in higher reduction of the total production. In case $\lambda\left(\mathbf{A}\right) = 0$, no vehicle interactions in addition to those captured by functions Π_j take place.

To be precise, as Ambühl *et al.* (2018a) show, $\lambda\left(\mathbf{A}\right)$ is also a measure of heterogeneity in the network. Thus, $\lambda\left(\mathbf{A}\right)$ 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 $\lambda\left(\mathbf{A}\right)$.

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

4.4.2.2 *A functional form for the 3D-MFD*

Here, as a particular application of the multimodal 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. 4.24 defines plane j in the three-dimensional space given by A_c, A_b, and Π, where $\left(car_{0,j}, bus_{0,j}, \pi_{0,j}\right)$ are the coordinates of a point in plane j, and $\left(car_{n,j}, bus_{n,j}, \pi_{n,j}\right)$ is the corresponding normal vector.

$$\left(\begin{pmatrix} A_c \\ A_b \\ \Pi \end{pmatrix} - \begin{pmatrix} car_{0,i} \\ bus_{0,i} \\ \pi_{0,i} \end{pmatrix}\right)^T \cdot \begin{pmatrix} car_{n,i} \\ bus_{n,i} \\ \pi_{n,i} \end{pmatrix} = 0 \tag{4.24}$$

After solving Eqn. 4.24 for Π, we obtain Eqn. 4.25 for plane Π_j as a function of both modes' vehicle accumulations. Recall that each plane $\Pi_j\left(A_c, A_b\right)$ is a function required for the general functional form as given in Eqn. 4.23.

$$\Pi_j\left(A_c, A_b\right) = \pi_{0,j} - \left(car_{n,j}\left(A_c - car_{0,j}\right) + bus_{n,j}\left(A_b - bus_{0,j}\right)\right) / \pi_{n,j} \tag{4.25}$$

The planes for Eqn. 4.23 will be defined from eleven points that we will introduce below. Figure 4.14 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 multimodal 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 A_c, A_b and Π. All resulting planes are summarized in Table 4.4.

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 4.3 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 4.10. 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:

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

Name	Key	Unit
Variables		
Accumulation of cars	A_c	(veh)
Accumulation of buses	A_b	(veh)
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	(-)
Total network length	L	(km)
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)
Average link length	l	(km)
Average bus stop spacing	p	(km)
Average dwell time	ΔT_d	(s)
Average cycle length	C	(s)
Average effective green time	G	(s)
Public transport strategy	∇	(-)

TABLE 4.3: 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.

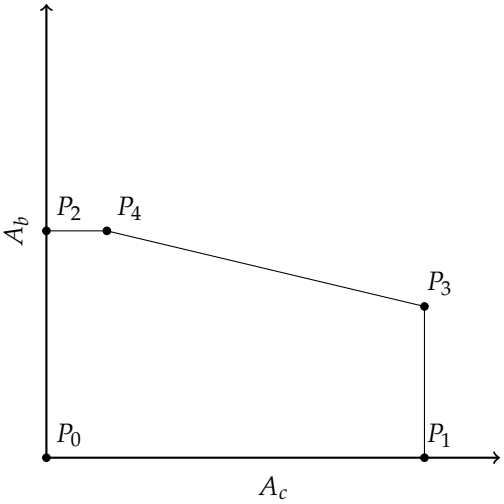


FIGURE 4.10: Top-down view on the 3D-MFD which shows the bimodal vehicle accumulation plane. For the defined points, total travel production is zero. The lines connecting the points also represent all combinations of accumulations where travel production is zero.

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 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. $A_b = 0$ (Figure 4.11). 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

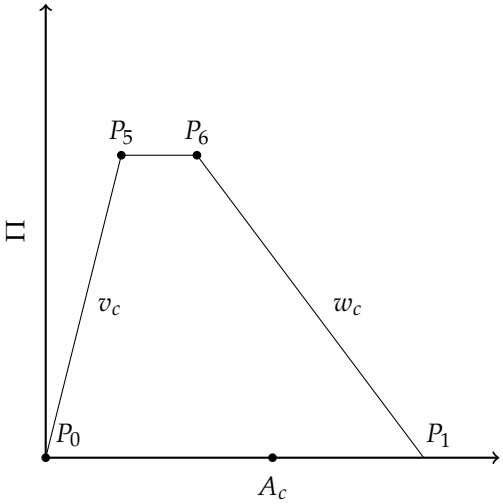


FIGURE 4.11: Side view on the 3D-MFD on the total travel production and car accumulation plane. For the defined points, no public transport operates.

v_c. Here, we approximate the network-wide average free-flow speed v_c as by Eqn. 4.26, 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. 4.26. 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 signal offsets (for simplicity not included here).

$$v_c = \frac{l}{\frac{l}{v_{c,0}} + \delta_c} \tag{4.26}$$

Figure 4.11 shows the discussed situation with $A_b = 0$ 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 4.10) and only points P_5 and P_6 remain to be defined. Both points share the common total travel production given by Eqn. 4.27.

$$\Pi_c = s_c \frac{G}{C} (1 - \eta_b) L \tag{4.27}$$

Equation 4.27 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 A_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 4.11. Then, points P_5 and P_6 equal to:

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

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

We now follow the same rationale for the case with only buses operating and no cars circulating, i.e. $A_c = 0$. Figure 4.12 exhibits this situation where the points P_0 and P_2 are already defined (see Figure 4.10). 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 total dwell time period ΔT_d that has a constant or fixed component and a demand-dependent component (see Eqn. 4.74 in Section 4.4.3 for details). For simplicity, we assume that the dwell time includes deceleration and acceleration, and that the public transport operator has defined for each stop a scheduled arrival and departure time so that ΔT_d 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 Zürich'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. Zürich, 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

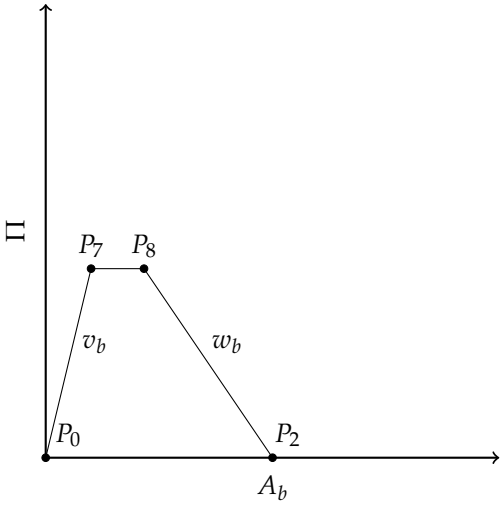


FIGURE 4.12: Side view on the 3D-MFD with the total travel production versus the accumulation of buses. For the defined points, no cars are operating.

strategies, we define the public transport strategy parameter $\nabla \in [0, 1]$. 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. 4.28 that follows the idea by Daganzo (2010), but accounts for the public transport priority strategy ∇ and simplifies the dwelling behavior. The bus backward wave speed w_b is then obtained using the same procedure as for cars in Eqn. 4.26.

$$v_b = \frac{p}{\frac{p}{v_{b,0}} + \delta_c \nabla \frac{p}{l} + \Delta T_d} \tag{4.28}$$

We assume that the maximum travel production of buses Π_b is achieved for the case of full public transport priority ($\nabla = 0$) with no delay at intersections. Π_b will be lower in case of similar delays at intersections for buses and cars ($\nabla = 1$). Similar to Π_c in Eqn. 4.27 for car traffic, we then define Π_b for the bus system in Eqn. 4.29 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 delays $\delta_c \nabla$ at p/l crossing between two stops and delays due to dwelling ΔT_d.

$$\Pi_b = s_b L \left(1 - \eta_c\right) \left(\frac{\frac{p}{v_{b,0}}}{\frac{p}{v_{b,0}} + \delta_c \nabla \frac{p}{l} + \Delta T_d} \right) \tag{4.29}$$

Consequently, points P_7 and P_8 are defined by:

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

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

We then define two additional points P_9 and P_{10}. They describe the influence of dedicated bus lanes on the multimodal 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:

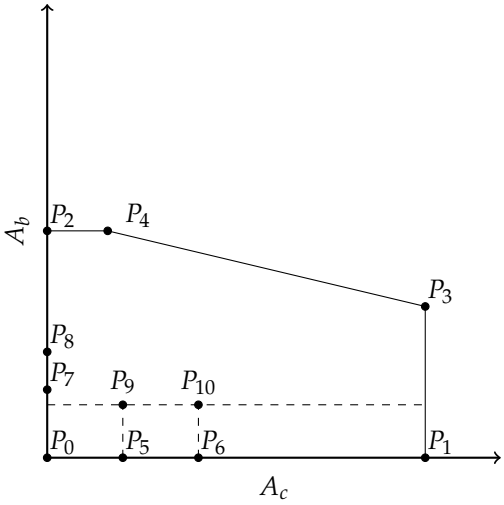


FIGURE 4.13: Additional points for the capacity in the 3D-MFD when buses on their dedicated lanes increase travel production from saturated car traffic.

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

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

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. 4.18 and 4.21, while points P_5 to P_{10} satisfy Eqn. 4.22. However, there are many operational features in multimodal networks that might further limit their productivity. These operational aspects can describe conflicts in mixed traffic, bus bunching, dwelling behav-

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 4.4: Construction of the seven planes from the eleven points.

ior, effects of the built environment, bus stop design (curb side or bus bay), public transport network design, and routing in the network.

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 4.4 numbers the planes and shows which combination of points defines each plane. Figure 4.14 shows the resulting planes, as well as the resulting shape of the 3D-MFD for an artificial network[3]. 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. 4.19 and 4.20. It should be clear, however, that this set of eleven points and the set

3 In Figure 4.14 and in Section 4.4.2.5, 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.

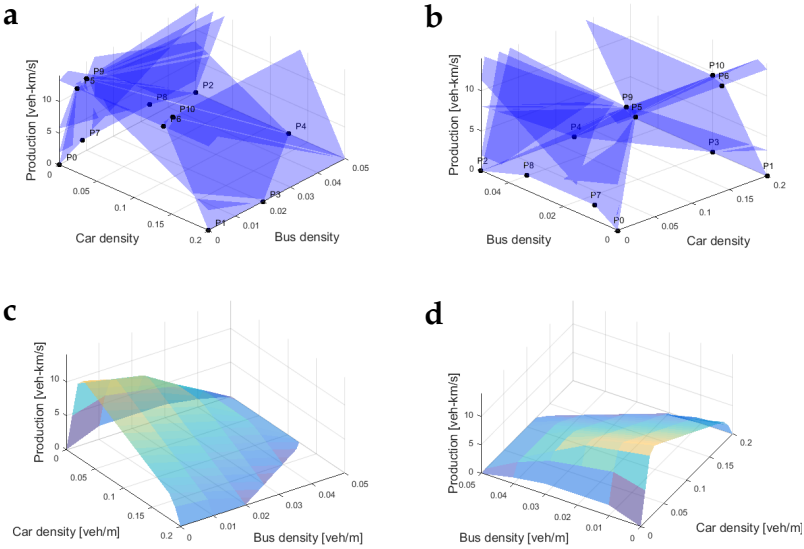


FIGURE 4.14: (**a**-**b**) illustrate with two different angles the location of the points and planes. (**c**-**d**) show with two different angles the shape of the resulting 3D-MFD.

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. 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 multimodal 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 Zürich 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 4.14**c** and 4.14**d** we then use the set of planes from Table 4.4 and Eqn. 4.23 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.* (2018a)). However, as Eqn. 4.23 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.

4.4.2.3 *Derivation of speed functions*

Unfortunately, in the definition of the 3D-MFD as given in Eqn. 4.23, 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 $v_{MFD} := \Pi / (A_c + A_b)$. Thus, to solve for each mode's 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. 4.30.

$$v_b \cong \theta v_c + \beta \tag{4.30}$$

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. 4.30, the speed of buses v_b and of cars v_c then follow according to Eqns. 4.31 and 4.32, respectively.

$$v_b = \frac{\Pi - \beta A_b}{A_c + \theta A_b} \theta + \beta \tag{4.31}$$

$$v_c = \frac{\Pi - \beta A_b}{A_c + \theta A_b} \tag{4.32}$$

For our analysis, we propose to approximate θ and β of Eqn. 4.30 with the parameters of the 3D-MFD shape as provided by Table 4.3 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. 4.33 and 4.34, respectively.

$$\beta = v_b \frac{\eta_b}{1 - \eta_c} \tag{4.33}$$

$$\theta = \frac{v_b}{v_c} \left(1 - \frac{\eta_b}{1 - \eta_c} \right) \tag{4.34}$$

Nevertheless, we consider that Eqn. 4.30 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. 4.31 and cars from Eqn. 4.32 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. 4.35 and car speeds in Eqn. 4.36. Here, we consider a trapezoidal fundamental diagram (Daganzo, 1994) with parameters from Table 4.3. The speeds are obtained by evaluating the fundamental diagram at the respective density (dividing vehicle accumulations by network length) of each mode, for buses $\vartheta_b \left(A_b / (1 - \eta_c) \right)$ and for cars $\vartheta_c \left(A_c / (1 - \eta_b) \right)$. Additionally in Eqn. 4.35, we require that the bus speed equals the MFD speed in cases when the car network is approaching gridlock to avoid negative bus speeds.

$$v_{b,MFD}\left(A_b, A_c\right) = \min\left(\vartheta_b\left(A_b\right); \frac{\Pi - \beta A_b}{A_c + \theta A_b}\theta + \beta; v_{MFD}\left(A_b, A_c\right) \right) \tag{4.35}$$

$$v_{c,MFD}\left(A_b, A_c\right) = \min\left(\vartheta_c\left(A_c\right); \frac{\Pi - \beta A_b}{A_c + \theta A_b} \right) \tag{4.36}$$

4.4.2.4 *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 4.3. 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 ΔT_d. The strategy parameter ∇ 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 4.4.2.5 using the Zürich 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 B.4 scaled by network length (see Section 4.4.2.5).

4.4.2.5 *Validation of the 3D-MFD*

Using planes *I-VII* for the 3D-MFD, we now compare Eqn. 4.23 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 4.15a); and (ii) a heterogeneous network with three levels of road hierarchy denoted as L1, L2, and L3 (Figure 4.15b). 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 Zürich. Figures 4.15c to d show the experimental sites in London and Zürich. 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 Zürich 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 B.4 in the Appendix, where those related to the network topology of London and Zürich are obtained from OpenStreetMap.

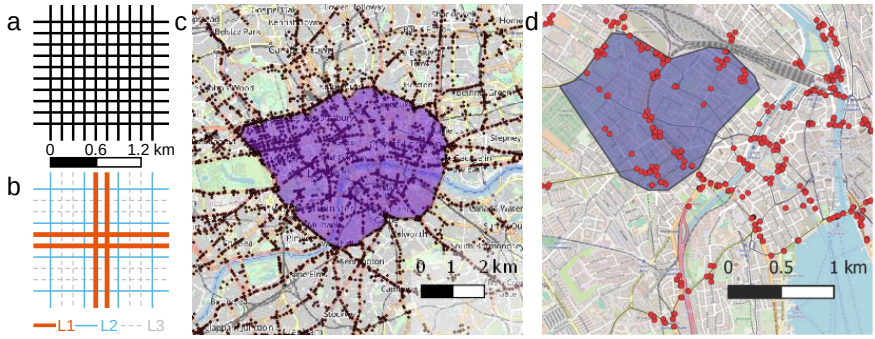


FIGURE 4.15: Network exhibits for the estimation of the 3D-MFDs. (**a**) shows the grid network in a homogeneous road configuration, while (**b**) shows a heterogeneous network with road hierarchy L1>L2>L3. (**c**) shows the neighborhood in London (UK) and (**d**) the neighborhood selected in Zürich (CH).

For all four data sets, we estimate λ with non-linear least squares. The estimated value for each 3D-MFD is given in Table B.4 in the Appendix. We show in Figure 4.16a 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 4.16 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 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 form's single parameter λ is difficult and consequently, the fit of the functional form in Figure 4.16 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.

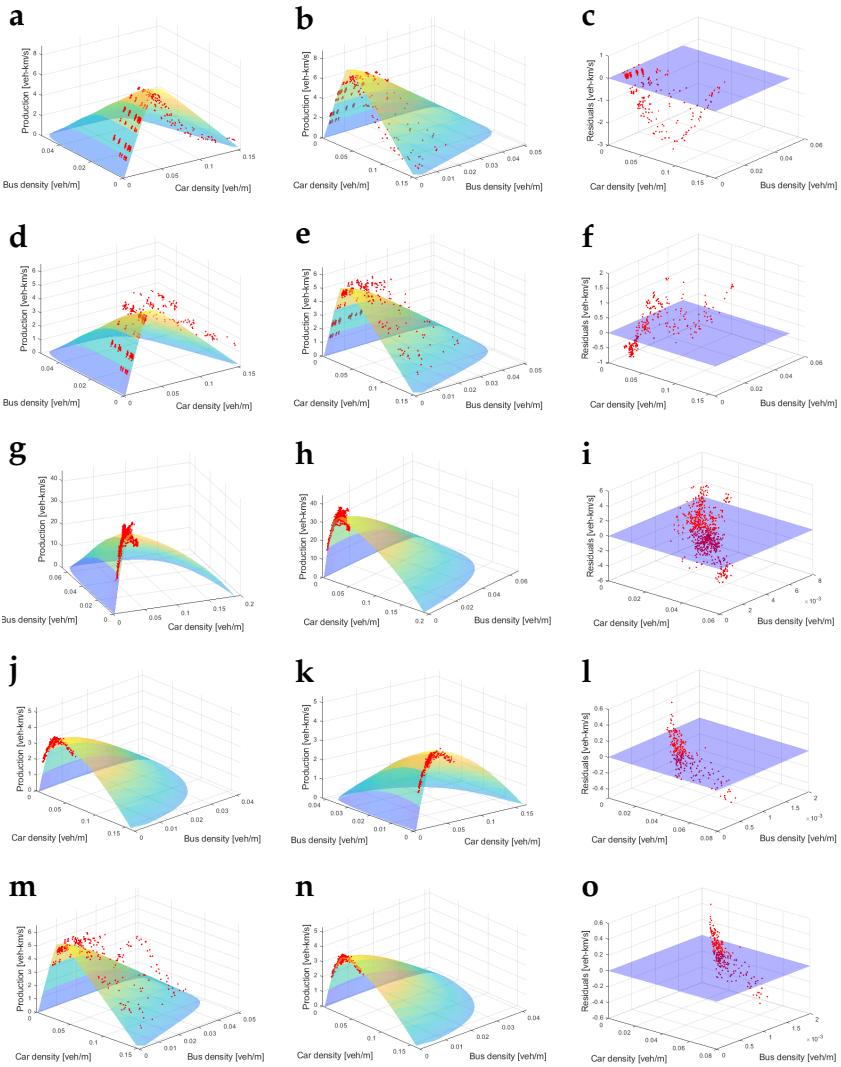


FIGURE 4.16: Validating the functional form for the 3D-MFD: homogeneous network (**a** to **c**), heterogeneous network (**d** to **f**), London (**g** to **i**), and Zürich (**j** to **o**). Here, (**j** to **l**) show the case if bus data is available and (**m** to **o**) if bus data is only approximated using Eqns. 4.37 and 4.38. Column one and two provide different angles of the fit, while the third column provides the residuals of the λ estimation.

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.* (2018a), 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.* (2018a) 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 4.15b. That being said, if the heterogeneity increases significantly, the network should be further partitioned to guarantee that the 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 Zürich, are available. To do so, we follow Daganzo's 2010 idea and we use the Zürich 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 Zürich 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 B.4 in the Appendix as many routes overlap. According to the time table of Zürich, 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. 4.37.

$$\pi_b = \frac{B}{H} = 436 \text{ veh-km h}^{-1} \tag{4.37}$$

With our observed data from Zürich, 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. 4.37 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 A_b with Eqn. 4.38 as used by Daganzo (2010).

$$A_b = \frac{\pi_b}{v_b} \tag{4.38}$$

Here, v_b is the commercial speed of buses, which can be either derived using Eqn. 4.28 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 $A_b = 436/11 \approx 40$ veh, which is close to the observed $A_b \approx 43$ veh during peak hour (with an error of about 7 %). Adding π_b and A_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 B.4 in the Appendix). Consequently, the resulting estimated 3D-MFD in Figures 4.16 m to o looks very similar to the 3D-MFD estimated using all empirical data in Figures 4.16 j to l.

We then compare our proposed functional form with Drake's 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 Drake's 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 Drake's 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 Drake's generalized exponential function and 2.28 for our proposed functional form, although MATLAB's fmincon solver exited for Drake's generalized exponential function with the status "local minimum possible" (see the resulting issue in Figure 4.17). Third, for Zürich we obtain an RMSE of 0.110 for Drake's generalized exponential function and 0.183 for our proposed functional form. However, Figure 4.17 emphasizes that it is difficult to obtain a satisfying 3D-MFD shape with Drake's generalized exponential function: Only with a full range of densities 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

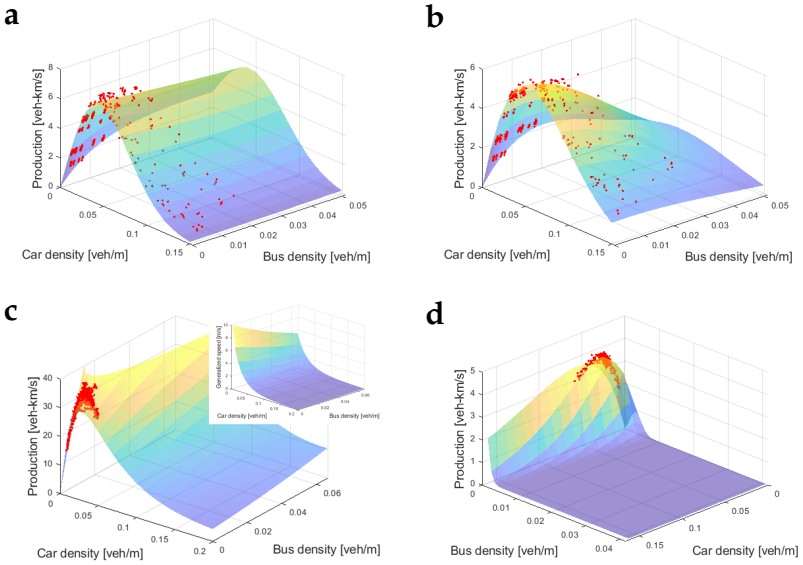


FIGURE 4.17: Estimation of Drake's exponential function as proposed by Geroliminis *et al.* (2014): (**a**) results from the simulation with the homogeneous network configuration, (**b**) results from the simulation with the heterogeneous network configuration, (**c**) shows the function estimated for London with the inset showing that the speed constraint is satisfied, and (**d**) shows the function estimated for Zürich.

optimum, the familiar 3D-MFD shape is clearly not recovered, although the proposed constraints are satisfied. For Zürich, 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.

4.4.2.6 *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. 4.30-4.32 as

well as our proposed speed model in Eqns. 4.35 and 4.36. Figure 4.18 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 the proposed linear relationship by Geroliminis *et al.* (2014) as well as our proposed speed model approximates the observed behavior relatively well.

4.4.3 *A Bose-Einstein condensate approach*

This sections proposes a novel and general methodology to capture analytically the effects of local and microscopic interactions between different transport modes in the MFD. We use a recently formulated functional form for the MFD (Ambühl *et al.*, 2018a) in conjunction with the two-fluid theory of town traffic by Herman and Prigogine (1979), based on the Bose-Einstein condensation, to bring the effects of multimodal microscopic interactions to the network level.

4.4.3.1 *Linking delays to the MFD*

The proposed interaction model applies to the interaction of m modes. Here we focus on the interactions between cars (subscript c), buses (subscript b), and bicycles (subscript v, for velo - french bicycle). We define $\mathbf{k}$ as the three-dimensional state vector of the system with elements car density, k_c, bus density, k_b, and bicycle density, k_v. Each mode has a well-defined MFD. Generally, the MFD is an upper envelope to all possible states in the relationships between network's average flow, q, and density, k (Daganzo, 2007; Daganzo and Geroliminis, 2008). Originally defined for car traffic only, we transfer the idea to buses and bicycles as the three modes only differ in propulsion and some operational characteristics, e.g. speeds, passenger occupancy, vehicle size. Here, we denote this upper envelope as the upper MFD (uMFD) and assume that it is known a priori (Ambühl *et al.*, 2018a; Daganzo *et al.*, 2018). Then, all observed traffic states will always be located *below* this uMFD due to traffic heterogeneity (e.g. Mazloumian *et al.*, 2010; Geroliminis and Sun, 2011; Gayah and Daganzo, 2011; Daganzo

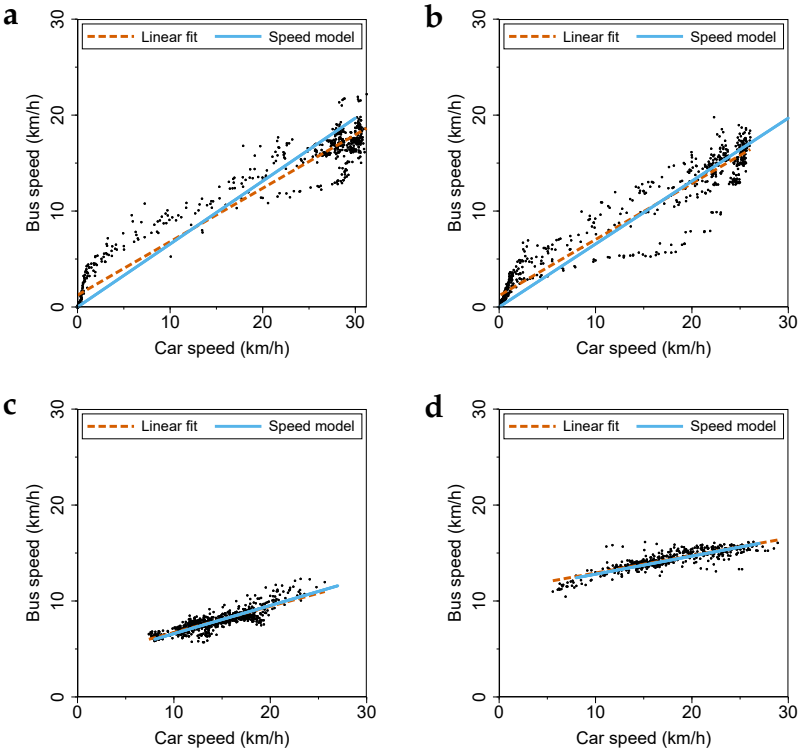
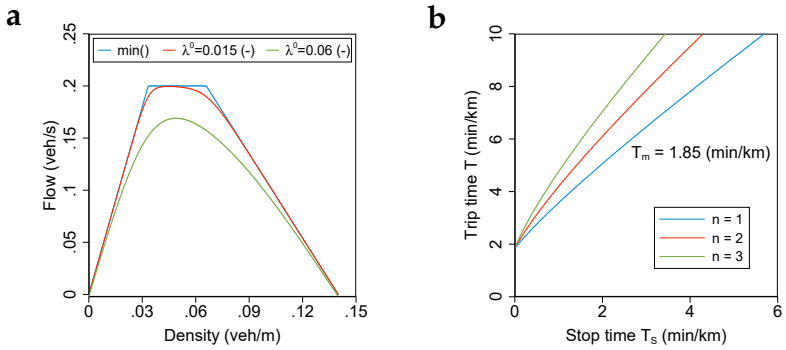


FIGURE 4.18: Validation of the proposed speed functions: (**a**) results from the simulation with the homogeneous network configuration, (**b**) results from the simulation with the heterogeneous network configuration, (**c**) shows the relationship for London, and (**d**) shows the relationship for Zürich.



(a) λ MFD function behavior

(b) Fundamental two-fluid relationships for different values of its parameter n.

FIGURE 4.19: Fundamental relationships of urban-scale traffic. (**a**) illustrates the λ MFD function with a trapezoidal uMFD, and (**b**) the fundamental relationship of the two-fluid theory of urban traffic.

et al., 2011) and network dynamics (e.g. Mariotte *et al.*, 2017). Here, we use the functional form for the MFD proposed by Ambühl *et al.* (2018a) that captures the gap between the a priori known uMFD and the observed MFD with just a single parameter, λ^0. This parameter can be seen as a quantification of network heterogeneity or the between-vehicle interactions. Eqn. 4.39 shows this functional form for a trapezoidal uMFD. Such uMFD shape has been used, for example, by Daganzo *et al.* (2018). For the reader's convenience we omit the subscript m for the mode as it appears for every equation for all variables and parameters.

$$q(k) = -\lambda^0 \ln\left(\exp\left(-\frac{v^f k}{\lambda^0}\right) + \exp\left(-\frac{Q}{\lambda^0}\right) + \exp\left(-\frac{(\kappa - k)\, w}{\lambda^0}\right)\right) \tag{4.39}$$

Here, v^f is the free flow speed in the network, Q is the network's capacity as defined by the most constraining intersection (Daganzo and Geroliminis, 2008), κ is the jam spacing in the network, and w is the backward wave speed. Arguably, each mode m has its characteristic values. In Figure 4.19**a** we illustrate the behavior of this MFD function for different values of λ^0

in comparison with the uMFD as defined by the minimum operator with the trapezoidal shape (Eqn. 4.39). Here, each of the trapezoidal shape's three segments enters as an argument to the minimum operator. With λ^0 approaching zero, the resulting curve approaches the uMFD, in this case the trapezoidal shape. When λ^0 increases, the curve moves further down but still describes the familiar MFD shape.

We also define the pace (travel time per unit length) T_m of mode m. It consists of two parts as given by Eqn. 4.40. The first term, $T_m^0(k_m)$, denotes the undisturbed pace of mode m given its current accumulation levels, k_m, i.e. without any interactions with other modes. The second term, $\Gamma_m(\mathbf{k})$, describes the additional delays caused by the interactions across modes. In other words, the undisturbed pace of m is given by each modes' own MFD, while the additional interaction delays Γ_m jointly depend on all modes' accumulations.

$$T_m(\mathbf{k}) = T_m^0(k_m) + \Gamma_m(\mathbf{k}) \tag{4.40}$$

The additional interaction delays $\Gamma_m(\mathbf{k})$ will decrease the flow of all vehicles as interactions always negatively effect each other. This means, that an increase in $\Gamma_m(\mathbf{k})$ will increase λ in Eqn. 4.39. In other words, we want to find $\mu(\mathbf{k})$ as defined in Eqn. 4.41, which increases λ^0 to $\breve{\lambda}$ due to the interactions between modes at a given state vector $\mathbf{k}$. Here, $\mu(\mathbf{k})$ is a function with $\mathbb{R}^m \to \mathbb{R}$. The model proposed in this paper allows to calculate $\mu(\mathbf{k})$, and then either $\breve{\lambda}$ or λ^0 if one of the latter two values is measured.

$$\breve{\lambda} = \lambda^0 + \mu(\mathbf{k}) \tag{4.41}$$

We will now establish the link between the additional interaction delays Γ and λ with the well-established *two-fluid theory* of urban traffic (Herman and Prigogine, 1979). The two-fluid theory of urban traffic is analogous to the Bose-Einstein condensation of particles at low temperatures (Ardekani and Herman, 1982; Dixit, 2013). In this theory, traffic consists of running vehicles (subscript r) and stopped vehicles (subscript s), where the running speed of vehicles v_r is related to the fraction of running vehicles f_r by Eqn. 4.42 and by definition Eqn. 4.43. Here, n is a network-wide constant and assumed to result from driving behavior, network topology, and signal settings, v^f is the free-flow speed, and v is the average space-mean speed in the network. The inverse of v is equal to the pace or travel time per unit distance T. As there are stopping and moving vehicles, T equals the sum of stopping time per unit distance T_s and the running time per unit distance

T_r, i.e. $T \equiv T_r + T_s$. We assume that the two-fluid theory of urban traffic applies to all considered modes m in the same fashion. For each mode, the parameter n is derived from λ as we show later.

$$v_r = v^f \left(f_r\right)^n \tag{4.42}$$

$$v = v_r f_r \tag{4.43}$$

By definition the fraction of vehicles stopped f_s and the fraction of running vehicles f_r always add up to one: $f_s + f_r \equiv 1$. Then, the space-mean speed in the network v results from Eqn. 4.44.

$$v = v^f \left(1 - f_s\right)^n f_r = v^f \left(1 - f_s\right)^{n+1} \tag{4.44}$$

Importantly, Herman and Prigogine (1979) pointed out that f_s is proportional to a power law with exponent h of the density to jam density ratio. In an empirical study, Lu *et al.* (2018) reported that $h \approx 1$, making computation easier. Here, however, we carry the h further as it can be context specific.

$$f_s = \left(\frac{T_s}{T}\right) \propto \left(\frac{k}{\kappa}\right)^h \tag{4.45}$$

The fundamental equation of the two-fluid theory results from Eqns. 4.42 and 4.44 and is given by Eqn. 4.46 (see for derivation Herman and Prigogine (1979)). This equation establishes a relationship between the total travel time per unit distance T, the running time per unit distance T_r, the free flow speed v^f as well as the network parameter n. We illustrate the functional behavior of the two fluid theory in Figure 4.19**b** for different values of n. For higher values of n, the fraction of stop time out of the total trip time decreases.

$$\log T_r = \frac{n}{n+1} \log T + \frac{1}{n+1} \log \left(\frac{1}{v^f}\right) \tag{4.46}$$

With the MFD expressed by Eqn. 4.39, we can algebraically derive formulae for T, T_r, T_s. The total trip time per kilometer or pace, T, is simply obtained by the inverse of the space-mean speed in the MFD as shown by Eqn. 4.47.

$$T\left(k\right) = \frac{k}{-\lambda^0 \ln\left(\exp\left(-\frac{v^f k}{\lambda^0}\right) + \exp\left(-\frac{Q}{\lambda^0}\right) + \exp\left(-\frac{(\kappa - k)w}{\lambda^0}\right)\right)} \tag{4.47}$$

Then, we obtain the running time per kilometer, T_r, by using Eqns. 4.43, 4.45 and $T(k)$ from Eqn. 4.47.

$$T_r(k) = \left(1 - \left(\frac{k}{\kappa}\right)^h\right) T(k) \tag{4.48}$$

Last, we obtain the stopping time per kilometer, T_s, by subtracting the running time from the total trip time as given by Eqn. 4.49.

$$T_s(k) = T(k) - T_r(k) \tag{4.49}$$

In their empirical work, Herman and Prigogine (1979), Ardekani and Herman (1985) and Ardekani *et al.* (1992) estimated n econometrically from measurements of T_r and T. As we obtained formulae for T and T_r, we can derive n analytically. Thus, after some algebra, we can solve Eqn. 4.46 for n, resulting in Eqn. 4.50.

$$n = \frac{1/\log v^f - \log T_r}{\log T_r - \log T} \tag{4.50}$$

In the following, we define λ^0, n^0, T^0, T_r^0 and T_s^0 with superscript 0 to denote the case of a steady-state traffic stream without any cross-modal disturbances, i.e. interaction with other modes. Let us then consider that the interactions with other transport modes create additional delays Γ. We assume that these delays either affect only the stopping time with $\Gamma_s(\mathbf{k})$ or the running time $\Gamma_r(\mathbf{k})$. We discuss these functions in detail later in this section. Here, Γ is a scalar function with $\mathbb{R}^m \to \mathbb{R}$. Accordingly, the two-fluid travel time variables can be rewritten for the case with additional delays with Eqns. 4.51-4.53.

$$T_r(\mathbf{k}) = T_r^0(k) + \Gamma_r(\mathbf{k}) \tag{4.51}$$

$$T_s(\mathbf{k}) = T_s^0(k) + \Gamma_s(\mathbf{k}) \tag{4.52}$$

$$T(\mathbf{k}) = T_r^0(k) + \Gamma_r(\mathbf{k}) + T_s^0(k) + \Gamma_s(\mathbf{k}) \tag{4.53}$$

With the additional interaction delays Γ on pace or travel times quantified (for a given λ^0 and $\mathbf{k}$), we can identify the network performance measures $\breve{n}$ and $\breve{\lambda}$ in the presence of cross-modal interactions. We use the inverse hat to indicate that the measure includes cross-modal delays. Then, for $\breve{n}$, we use Eqn. 4.50 and Eqn. 4.51-4.53 to obtain Eqn. 4.54.

$$\check{n}\left(\mathbf{k}\right) = \frac{1/\log v^f - \log\left(T_r^0\left(k\right) + \Gamma_r\left(\mathbf{k}\right)\right)}{\log\left(T_r^0\left(k\right) + \Gamma_r\left(\mathbf{k}\right)\right) - \log\left(T_r^0\left(k\right) + \Gamma_r\left(\mathbf{k}\right) + T_s^0\left(k\right) + \Gamma_s\left(\mathbf{k}\right)\right)} \tag{4.54}$$

Then, to obtain $\check{\lambda}$, we equate in Eqn. 4.55 the space-mean speed of the λ trapezoidal function from Eqn. 4.39 and the speed of the two-fluid theory from Eqn. 4.44. Note that the information of $\mathbf{k}$ is now carried along with $\check{n}$ and that the right-hand side of Eqn. 4.55 is similar to the inverse of Eqn. 4.47, but the λ^0 is replaced by $\check{n}$ to calculate the interaction effects.

$$v^f\left(1 - \left(\frac{k}{\kappa}\right)^h\right)^{\check{n}(\mathbf{k})+1} = -\check{\lambda}\frac{\ln\left(\exp\left(-\frac{v^f k}{\check{\lambda}}\right) + \exp\left(-\frac{Q}{\check{\lambda}}\right) + \exp\left(-\frac{(\kappa-k)w}{\check{\lambda}}\right)\right)}{k} \tag{4.55}$$

Eqn. 4.55 can simply then be solved as a root problem using a mathematical software when transformed into Eqn. 4.56. The only unknown is $\check{\lambda}$.

$$0 = -\check{\lambda}\frac{\ln\left(\exp\left(-\frac{v^f k}{\check{\lambda}}\right) + \exp\left(-\frac{Q}{\check{\lambda}}\right) + \exp\left(-\frac{(\kappa-k)w}{\check{\lambda}}\right)\right)}{k} - v^f\left(1 - \left(\frac{k}{\kappa}\right)^h\right)^{\check{n}(\mathbf{k})+1} \tag{4.56}$$

The problem formulated in Eqn. 4.56 must be solved for each mode m separately and because of the high non-linearity of model, we propose to solve Eqn. 4.56 for each demand situation separately, i.e. for all possible values of $\mathbf{k}$, instead of assuming constant n or λ values over all densities.

DELAY FUNCTIONS

In the following, we focus on identifying delay functions, i.e. $\Gamma\left(\mathbf{k}\right)$ for an urban network with given MFDs for each mode. Here, we firstly consider that the interactions between modes are uniformly distributed over time

and space (we discuss later spatial heterogeneity of delays). The methodology presented above is generic and allows to use any formulation of $\Gamma_s(\mathbf{k})$ and $\Gamma_r(\mathbf{k})$ functions for all modes that can be represented with an MFD, but here we focus on bicycles, buses, and cars.

We use the following notation: $\Gamma_s^{c\to b}$ describes the interaction stopping delays caused by cars on buses. We use the $\to$ operator to indicate which mode affects which other mode. Where we do not provide the $\to$ operator, Γ corresponds to the total interaction delay caused by all other modes combined. Intuitively, the interaction delay functions depend on the network topology, i.e. in case all modes run on dedicated infrastructures the interaction delays are zero, while they are non-zero when their infrastructure is (partially) overlapping.

As mentioned before, $\Gamma(\mathbf{k})$ has *two* mechanisms: stopping delays $\Gamma_s(\mathbf{k})$ and running delays $\Gamma_r(\mathbf{k})$. We assume additivity of delays within each mechanism as formulated in Eqn. 4.57 for the additional stopping delays for cars $\Gamma_s^c(\mathbf{k})$, i.e. we calculate additional delays pairwise, and their sum is then the total interaction delay. In other words, this assumes no combined or second order effects, e.g. from bicycles and buses on cars.

$$\Gamma_s^c(\mathbf{k}) = \Gamma_s^{b\to c}(k_\mathrm{c}, k_\mathrm{b}) + \Gamma_s^{v\to c}(k_\mathrm{c}, k_\mathrm{v}) \tag{4.57}$$

In Table 4.5 we summarize for each of the two mechanisms the delay model used for quantifying the interactions between all considered modes. In total, we use four different delay models: A continuous multiclass fundamental diagram (FD) taken from Bliemer (2001), a discrete multiclass FD proposed by Wierbos *et al.* (2018), a platoon dispersion model as proposed by Robertson (1969), and a bus dwelling behavior model based on Daganzo (2010). We do not use a continuous multiclass model for the interactions that involve bicycles as we assume that in the congested case of cars and buses, bicycle speeds do not converge to that of the other modes, i.e. bicycles can sneak through the vehicle queues. Therefore, we provide a separate discrete multiclass FD. We generally expect no analytical mechanism for additional interaction stopping delays caused by cars on buses and bicycles, but that, for example, in congested traffic states, additional interaction stopping delays are function of additional interaction running delays, i.e., $\Gamma_s = f(\Gamma_r)$.

For the additional running delays, the general modeling idea is to express the additional delay as the difference in pace between the pace from the multimodal FDs and the unimodal FDs. In other words, the difference between the case where both modes are interacting and the case without

Interaction case	Running delay	Stopping delay
$c \rightarrow b$	Continuous multiclass FD	$f(\Gamma_r)$
$b \rightarrow c$	Continuous multiclass FD	Bus dwelling behavior
$b \rightarrow v$	Discrete multiclass FD	Bus dwelling behavior
$v \rightarrow b$	Discrete multiclass FD	Bicycle platoon dispersion
$c \rightarrow v$	Discrete multiclass FD	$f(\Gamma_r)$
$v \rightarrow c$	Discrete multiclass FD	Bicycle platoon dispersion

TABLE 4.5: Assignment of delay models to the mechanisms of running and stopping delays.

interaction. For the interaction stopping delays, the general modeling idea is to quantify the sources of additional stopping delays caused by the interactions. In this case, we do not need to subtract anything from these delays as they are fully additional, i.e. in the uninterrupted case, no such delays are to be expected.

For the formulation of the delay functions, we consider an urban network, where the interactions are continuously distributed. The features of the urban network are as follows. Intersections are spaced at distance l, each intersection has a width of c, the number of lanes per driving direction is d, and bus stops are placed at distance p. Following Daganzo's 2010 formulation of bus network design, the design of the bus network is α. For α, $0 < \alpha \leq 1$ must hold. Close to its lower bound, α describes a hub-and-spoke network, while at its upper bound it describes a grid network. Values in between are hybrid networks where one can see α as the fraction of network exhibiting a grid structure. Each mode has a given MFD as formulated in Eqn. 4.39. To distinguish between the speeds obtained from the MFD and those from the fundamental diagram, we define $\vartheta(k)$ as the speed from the fundamental diagram and $v(k)$ as the speed from the MFD. Following the idea of the two fluid theory, we assume that the MFD running speed corresponds to the FD speed, i.e. $v_r(k) = \vartheta(k)$. We further define that $\vartheta_m(k_m)$ always refers to the unimodal speed-density relationship, while $\vartheta_m(\mathbf{k})$ refers to the speeds obtained from the multiclass FDs that we introduce in the following.

Continuous multiclass fundamental diagram

The continuous multiclass fundamental diagram used in this delay model has been proposed by Bliemer (2001). This model enters the additional running delays between cars and buses $\Gamma_r^{c\to b}$ and $\Gamma_r^{b\to c}$. Consequently, we express the equations provided by Bliemer (2001) for this specific interaction type with two vehicle classes. Similar to Bliemer (2001), we consider class 1 to be the standard passenger and class 2 to be buses that are moving bottlenecks (Muñoz and Daganzo, 2002; Eichler and Daganzo, 2006; Castrillon and Laval, 2018), similar as trucks on a motorway as studied by Bliemer (2001).

The fundamental diagram for the two mode case consists primary of three equations. The first equation is given by Eqn. 4.58 and is the inverse of the speed-density relationship for cars. In Eqn. 4.58, μ_1, μ_2, and μ_3 are parameters to be estimated and v_c^f is the free-flow speed for cars. Note that in Eqn. 4.58 we omit the density argument for ϑ_c as the speed-density relationship is given as the inverse.

$$k_c\left(\vartheta_c\right) = \frac{1}{\mu_1 + \dfrac{\mu_2}{v_c^f - \vartheta_c} + \mu_3 \vartheta_c} \tag{4.58}$$

The second equation makes the simplifying assumption that the speed of buses ϑ_b is a function of the (faster) car speed ϑ_c, i.e. $\vartheta_b = f\left(\vartheta_c\right)$. There are several functional forms possible but Bliemer (2001) proposes Eqn. 4.59, which we for simplicity use here as well. Importantly, the differences in free flow speeds can be substantial in case agencies have a desired bus speed to which bus drivers have to stick in order to maintain the schedule. In Eqn. 4.59, parameter ν_b has to be estimated from data.

$$\vartheta_b\left(k_b, k_c\right) = \frac{\vartheta_c\left(k_b, k_c\right)}{1 + \nu_b\left(\vartheta_c\left(k_b, k_c\right)\right)^{\left(\log\left(v_c^f / v_b^f - 1\right) - \log(\nu_b) / \log\left(v_c^f\right)\right)}} \tag{4.59}$$

The third equation quantifies the effective density $\overline{k}$, measured in the size of the standard passenger car size of class 1 as formulated in Eqn. 4.60. Here, $\varphi\left(\vartheta_c\right)$ is the passenger car equivalent of a bus that we assume to be a function of car speed $\vartheta_c\left(k_b, k_c\right)$ and $\psi_{bc}\left(\vartheta_c\right)$ is the interaction parameter, also assumed to be a function of $\vartheta_c\left(k_b, k_c\right)$.

$$\overline{k} = k_c + \varphi\left(\vartheta_c\right) k_b + \psi_{bc}\left(\vartheta_c\right) k_b k_c \tag{4.60}$$

Consequently, we need to estimate the parameters of Eqns. 4.58, 4.59 and 4.60 with observations of speed and densities by both modes. To account for the effective density $\bar{k}$ in the unimodal fundamental diagram, we substitute Eqn. 4.60 for $k_c(\vartheta_c)$ from Eqn. 4.58 to obtain Eqn. 4.61.

$$\frac{1}{\mu_1 + \dfrac{\mu_2}{u_c^f - \vartheta_c(k_b, k_c)} + \mu_3 \vartheta_c(k_b, k_c)} = k_c + \varphi(\vartheta_c) k_b + \psi_{bc}(\vartheta_c) k_b k_c \quad (4.61)$$

For convenience, we assume that we can express $\varphi(\vartheta_c)$ as a linear function as given in Eqn. 4.62, and we assume for $\psi_{bc}(\vartheta_c)$ a linear function as well as given in Eqn. 4.63.

$$\varphi(\vartheta_c) = \varphi_0 - \varphi' \vartheta_c(k_b, k_c) \quad (4.62)$$

$$\psi_{bc}(\vartheta_c) = \psi_0\ \vartheta_c(k_b, k_c) \quad (4.63)$$

Discrete multiclass fundamental diagram

The discrete multiclass fundamental diagram used here has been recently introduced for bicycles and cars by Wierbos *et al.* (2018). This model enters all running delay formulations where bicycles are involved. Namely, $\Gamma_r^{v\to c}$, $\Gamma_r^{v\to b}$, $\Gamma_r^{b\to v}$, and $\Gamma_r^{c\to v}$. Consequently, we define for this section a subset $m' \in \{b, c\}$ to generalize the equations.

The multiclass traffic flow is illustrated in Figure 4.20. Here, let us define two threshold density values for the bicycle density k_v. The first value, k'_v, defines an upper limit up to which vehicles experience no obstruction by cyclists as they are only loosely spaced as shown in Figure 4.20**a**. The second threshold value, k''_v, where $k''_v > k'_v$, describes the situation where vehicles can still overtake bicycles at a reduced speed of $v_{m',red}$ as presented in Figure 4.20**b**. If the number of bicycles exceeds k''_v vehicles need to adopt the speed of bicycles as we illustrate in Figure 4.20**c**.

We define the speed of bicycles $\vartheta_v(k_{m'}, k_v)$ as outlined in Eqn. 4.64. In case vehicles (cars or buses) are loosely spaced, i.e. $k_{m'} < k'_{m'}$, the speed of bicycles is $\vartheta_v(k_v)$, but if traffic gets denser, bicycles reduce their speed to δ of $\vartheta_v(k_v)$.

$$\vartheta_v(k_{m'}, k_v) = \begin{cases} \vartheta_v(k_v), & \text{if } k_{m'} < k'_{m'} \\ \delta\ \vartheta_v(k_v) & \text{if } k_{m'} \geq k'_{m'} \end{cases} \quad (4.64)$$

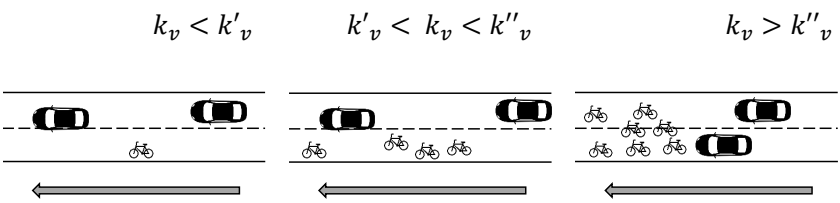


FIGURE 4.20: Illustration of bicycle and car interactions within the link causing additional running time delays for cars, $\Gamma_r^{v \to c}$. In case bicycles are only loosely spaced in the network with $k_v < k'_v$ as shown in (**a**) then cars experience no additional delay at all. When the bicycle density increases, but it holds $k'_v < k_v < k''_b$, then cars overtake bicycles with a reduced speed for safety reasons as shown in (**b**). Once the bicycle density exceeds $k_v > k''_v$ the speed of cars is determined by the speed of bicycles as illustrated in (**c**). The car by Pablo Rozenberg from the Noun Project and the bicycle by Jens Tärning from the Noun Project.

As aforementioned and shown in Figure 4.20, we define in Eqn. 4.65 that the speed of vehicles, $\vartheta_{m'}(k_{m'}, k_v)$ is equal to the unimodal speed of vehicles $\vartheta_{m'}(k_{m'})$ in case the bicycles are only loosely spaced, i.e. $k_v < k'_v$. In the case of $k'_v < k_v < k''_v$, we expect that vehicles drive either at a reduced speed for overtaking cyclists or their unimodal speed $\vartheta_{m'}(k_{m'})$ if it is less than $v_{m',red}$. Then, when $k_v > k''_v$, we expect that vehicles need to adopt the speed of cyclists $\vartheta_v(k_{m'}, k_v)$ or their unimodal speed if it is less than the speed of cyclists.

$$\vartheta_{m'}(k_{m'}, k_v) = \begin{cases} \vartheta_{m'}(k_{m'}) & \text{if } k_v < k'_v. \\ \min\left(v_{m',red}, \vartheta_{m'}(k_{m'})\right) & \text{if } k'_v < k_v < k''_v. \\ \min\left(\vartheta_v(k_{m'}, k_v), \vartheta_{m'}(k_{m'})\right) & \text{if } k_v > k''_v \end{cases} \tag{4.65}$$

Bicycle platoon dispersion

This model's output enters the additional stopping delays for cars caused by bicycles, $\Gamma_s^{v \to c}$, and for buses caused by bicycles, $\Gamma_s^{v \to b}$. The key element

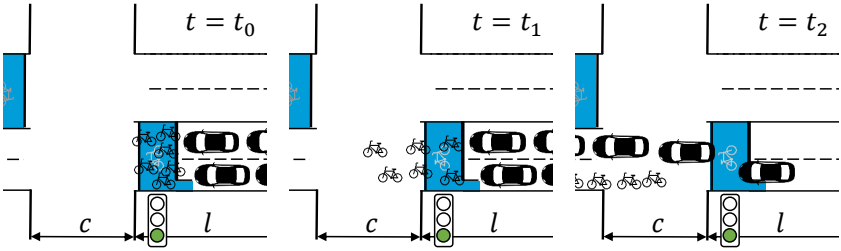


FIGURE 4.21: Illustration of the bicycle platoon dispersion at a traffic signal. At $t = t_0$ all bicycles are located upfront of the cars and the signal turns green. Then, first the bicycles move but cars cannot at $t = t_1$ as the bicycle platoon is still blocking the intersection. Cars can clear the intersection and overtake the bicycles once the bicycle platoon has dispersed and cleared most intersection space at $t = t_2$. The car by Pablo Rozenberg from the Noun Project and the bicycle by Jens Tärning from the Noun Project.

of this delay model is the situation-adopted Robertson's platoon dispersion model (Robertson, 1969; Paul *et al.*, 2016; Axhausen and Körling, 1987).

First, consider the situation in the panels of Figure 4.21. During a red signal phase of length R, b_1 bicycles accumulate in a so called *pocket* in front of the stop line of cars. This kind of intersection design is becoming more and more common in cities like London and Zürich. At time $t = t_0$ the signal turns green. Now, the bicycle platoon starts to move, but the cars cannot accelerate as all lanes are blocked with bicycles at time $t = t_1$. At time $t = t_2$, the bicycle platoon cleared the intersection and has dispersed to that extent that cars can now discharge and drive into the next link. In this model, we generally assume that with more bicyclists in the *pocket*, cars have to wait longer before discharging.

At the beginning of the discharge phase, $t = 0$, the bicycle platoon is of size b_1 that is calculated by Eqn. 4.66 as the number of bicycles that arrive during red phase R at most of density k_v and the maximum possible speed $\vartheta_v\,(k_v)$
v_v^f.

$$b_1 = R\, k_v\, v_v^f \tag{4.66}$$

To model this platoon dispersion, we adopt Robertson's 1969 layout definition and consider the intersection length c as the context-specific dis-

persion corridor as shown in Figure 4.21. Further, we know the average time, T_a, and standard deviation thereof, σ, of a cyclist to cover distance c. We define the additional delay for stopping as the time until the fraction ψ of the initial platoon b_1 traversed the dispersion corridor. In other words, we need to find the time duration τ, where the cumulative sum of downstream arrivals of bicycles b_t^d, is equal to or greater than $\psi\ b_1$. The downstream arrivals b_t^d is expressed in Eqn. 4.67 (Seddon, 1972) and T is expressed in units of g, which is the length of the modeling interval in g seconds.

$$b_t^d = \begin{cases} 0 & \text{if } t \leq T \\ \sum_{i=1}^{t-T} b_i F\,(1-F)^{t-T-i} & \text{if } t > T \end{cases} \tag{4.67}$$

Where the factor F is calculated by Eqn. 4.68.

$$F = g\frac{\sqrt{g^2+4\sigma^2}-g}{2\sigma^2} \tag{4.68}$$

We obtain the minimum time to travel along c using Eqn. 4.69.

$$T = \beta_n T_a \tag{4.69}$$

Where we calculate β_g using Eqn. 4.70.

$$\beta_g = \frac{2T_a + g - \sqrt{g^2+4\sigma^2}}{2T_a} \tag{4.70}$$

For simplicity of the model, we assume that the bicycle platoon upstream is only non-zero at $t = 1$, i.e. b_1 is non-zero. This then simplifies Eqn. 4.67 into Eqn. 4.71.

$$b_t^d = \begin{cases} 0 & \text{if } t \leq T \\ b_1 F\,(1-F)^{t-T-1} & \text{if } t > T \end{cases} \tag{4.71}$$

Then, we define Eqn. 4.72 which determines how many modelling intervals or time steps are required until γ of the initial bicycle platoon b_1 cleared the way for cars.

$$\gamma\ b_1 \geq \sum_{t=1}^{\tau} b_t^d \tag{4.72}$$

We expect that γ depends on the initial platoon size b_1. Arguably, at low densities there will be close to no additional difficulties in passing the

dispersing platoon, while at very high bicycle densities we expect that car drivers have to wait until almost the entire platoon has cleared the intersection. Thus, for simplicity, we assume a linear relationship as formulated in Eqn. 4.73, where ψ is the platoon clearing factor and $\overline{b}$ is the bicycle clearing limit. Both together mean that cars have to wait when $\overline{b}$ bicycles are waiting until ψ have cleared the bicycle stopping box before start moving. Both parameters need to be estimated from data.

$$\gamma = \psi \frac{b_1}{\overline{b}} \tag{4.73}$$

Consequently, the time until the vehicles beyond the bicycle platoon can pass the intersection is given by $g\ \tau$.

Bus dwelling behavior

For quantifying additional stopping delays caused by buses on cars and bicycles, we concentrate on the feature that buses stop on the curbside and thus force other vehicles to stop behind them. A network perspective to this has been provided by Daganzo (2010) with the average commercial speed of buses v_{com} as formulated in Eqn. 4.74. The commercial speed combines the running speed $\vartheta_b(\mathbf{k})$ with the average network-wide dwelling behavior of buses that is a function of several network parameters: constant dwell time ζ at every stop placed at distance p and demand-depending dwell time ζ' as a function of the demand Λ, the headway H, the structure of the bus network α, and the diameter of the regional network D.

$$\frac{1}{v_{com}} = \frac{1}{\vartheta_b(\mathbf{k})} + \frac{\zeta}{p} + \frac{0.5\ (1+e_T)\,\zeta'\Lambda\ p\ H/D^2}{3\alpha - \alpha^2} \tag{4.74}$$

Clearly, Eqn. 4.74 illustrates the running behavior, which has already been quantified in Section 4.4.3.1, and the stopping behavior with the second and third term of the right-hand side. As in this paper we are working with vehicle density k_b instead of headway H, we need require the analytical link between both variables in order to use the stopping behavior of Eqn. 4.74 for our delay formulations. For this, we divide Eqn. 3 (fleet size) by Eqn. 1 (bus network length) of Daganzo (2010), which gives the bus density as formulated by Eqn. 4.75.

$$k_b = \frac{2\frac{\left(3\alpha - \alpha^2\right)}{H}\left(\frac{1}{\vartheta_b(\mathbf{k})} + \frac{\zeta}{p}\right) + (1+e_T)\,\zeta'\Lambda\ p/D^2}{1+\alpha^2} \tag{4.75}$$

If the dwelling dynamics as a function of demand are omitted, i.e. $\zeta' = 0$, then the third term of Eqn. 4.74 as well as the second term in the numerator of Eqn. 4.75 can be neglected. Further, if the perspective is only on an urban corridor, then one can assume $\alpha = 1$ and $e_T = 0$.

Formulation of additional delays

With the discussed delay models, we can now formulate the additional running and stopping delays. The formulation of the additional running delays Γ_r is straightforward as it is only the difference in pace between the interacting case and the case where the interaction is not considered. Recall that we define $\vartheta_m(k_m)$ as the unimodal speed-density FD and $\vartheta_m(k_m, k_{m'})$ as the multiclass speed-density FD. Then, we can formulate all required Γ_r functions as provided in Eqns. 4.76 to 4.81.

$$\Gamma_r^{v\to c}(k_c, k_v) = \frac{1}{\vartheta_c(k_c, k_v)} - \frac{1}{\vartheta_c(k_c)} \tag{4.76}$$

$$\Gamma_r^{c\to v}(k_c, k_v) = \frac{1}{\vartheta_v(k_c, k_v)} - \frac{1}{\vartheta_v(k_v)} \tag{4.77}$$

$$\Gamma_r^{c\to b}(k_b, k_c) = \frac{1}{\vartheta_b(k_b, k_c)} - \frac{1}{\vartheta_b(k_b)} \tag{4.78}$$

$$\Gamma_r^{b\to c}(k_b, k_c) = \frac{1}{\vartheta_c(k_b, k_c)} - \frac{1}{\vartheta_c(k_c)} \tag{4.79}$$

$$\Gamma_r^{v\to b}(k_b, k_v) = \frac{1}{\vartheta_b(k_b, k_v)} - \frac{1}{\vartheta_b(k_b)} \tag{4.80}$$

$$\Gamma_r^{b\to v}(k_b, k_v) = \frac{1}{\vartheta_v(k_b, k_v)} - \frac{1}{\vartheta_v(k_v)} \tag{4.81}$$

Secondly, we formulate the additional stopping delays. For the additional stopping delays caused by bicycles and cars and buses, we use the duration until vehicles stopped beyond bicyclists at intersections can pass the intersection $g\tau$ and scale the delay per intersection spacing and the likelihood of encountering a red phase at a signal. Recall that the τ equals the number of intervals and g the interval length. Then the additional delays for cars and buses are given by Eqns. 4.82 and 4.83, respectively.

$$\Gamma_s^{v\to c}(k_c, k_v) = g\tau\frac{R}{C}\frac{1}{l} \tag{4.82}$$

$$\Gamma_s^{v \to b}(k_b, k_v) = g\tau \frac{R}{C} \frac{1}{l} \tag{4.83}$$

Regarding the stopping delays caused by buses, we use the dwelling component of Eqn. 4.74. Accordingly, the additional stopping delay $\Gamma_s^{b \to c}$ is defined as given by Eqn. 4.84 and the additional stopping delay $\Gamma_s^{b \to v}$ is given by Eqn. 4.85.

$$\Gamma_s^{b \to c}(k_b, k_c) = \frac{\zeta}{p} + \frac{0.5\ (1 + e_T)\, \zeta' \Lambda\ p\ H/D^2}{3\alpha - \alpha^2} \tag{4.84}$$

$$\Gamma_s^{b \to v}(k_b, k_v) = \frac{\zeta}{p} + \frac{0.5\ (1 + e_T)\, \zeta' \Lambda\ p\ H/D^2}{3\alpha - \alpha^2} \tag{4.85}$$

ACCOUNTING FOR HETEROGENEITY OF DELAYS

In reality, the interaction delays are not spatially and temporally homogeneous as previously assumed: In an urban network with d lanes, buses and bicycles usually only use a single lane, the curbside lane. Further, buses have a given headway and thus *their* bottleneck nature is only activated when a bus is present. Thus, we need to scale interaction delays by the probability of interaction to obtain the expected value of interaction delays. Importantly, the required scalars are case specific and need to be defined for each interaction situation separately. For the tri-modal case of bicycles, buses and cars, we consider ρ_l to capture the spatial probability (or heterogeneity) induced by the number of lanes and ρ_b the temporal probability (or heterogeneity) of bus interactions. Accordingly, both are scalars to the respective delays.

First, for the spatial heterogeneity of interaction delays, we assume that $d - 1$ lanes of the network are dedicated for cars while only one lane is mixed. Consequently, the interaction effects will fully affect the mixed lane, while the other lanes are less affected. However, due to spillover effects, the $d - 1$ lanes are not completely unaffected. Thus, we assume that the network-wide probability of interaction is a convex function, decreasing with the number of lanes. We propose Eqn. 4.86 as a possible functional representation of this behavior by using the square root. However, future research is required to validate this assumption. For the case of bicycle-bus interactions, where both use the curbside lane, we consider here only the mixed lane, i.e. $\rho_l = 1$.

$$\rho_l = \frac{1}{\sqrt{d}} \tag{4.86}$$

Second, we define the temporal probability of interaction, ρ_b, on a link where vehicles of mode m are affected by stopping buses. We assume a distance p between bus stops in the network and define $\hat{p} = p^{-1}$ as the number of bus stops per kilometer, the headway of buses is H that is a function of bus density k_b, ΔT_d is the total bus dwell time, calculated using the commercial bus speed equation by Daganzo (2010), and α is the design of the bus network as proposed by Daganzo (2010), for which $0 < \alpha \leq 1$ holds. At its lower bound it describes a hub-and-spoke network, while at its upper bound it describes a perfect grid network. Values in-between indicate hybrid structures. Here, we consider that the temporal probability of encountering a stopping bus at a bus stop follows the binomial distribution as given in Eqn. 4.87 as there is a chance of encountering a bus at one of p stops per kilometer, at two of p stops or at every p stops. Note that the first expression in parentheses in Eqn. 4.87 is the binomial coefficient. The probability of encountering a bus at a bus stop is given by $\Delta T_d / H$. The minimum number of bus stops a car will encounter per kilometer in a network is given by $\lfloor \alpha / \hat{p} \rfloor$. Importantly, this is a strict simplification of multimodal urban traffic as it as assumes independence of interactions.

$$\rho_b\left(k_b\right) = \sum_{j=1}^{\lfloor \alpha/\hat{p} \rfloor} \binom{\lfloor \alpha/\hat{p} \rfloor}{j} \left(\frac{\Delta T_d}{H\left(k_b\right)}\right)^j \left(1 - \frac{\Delta T_d}{H\left(k_b\right)}\right)^{\lfloor \alpha/\hat{p} \rfloor - j} \tag{4.87}$$

CALIBRATION

The presented multimodal interaction model for the tri-modal MFD requires a calibration of all parameters. These need either to be measured, estimated or approximated. Section B.2 summarizes the calibration routine. In Figure 4.22 we then show the resulting car MFDs where we illustrate the bi-modal interaction effects with buses (see panels **(a)** and **(b)**) and with bicycles (see panels **(c)** and **(d)**). Note that each presented pair of figures is based on the parameters given in Table B.2 and thus only the view point differs. In all four figures we clearly find the intuitive and expected effect that more interactions decrease the performance of car travel.

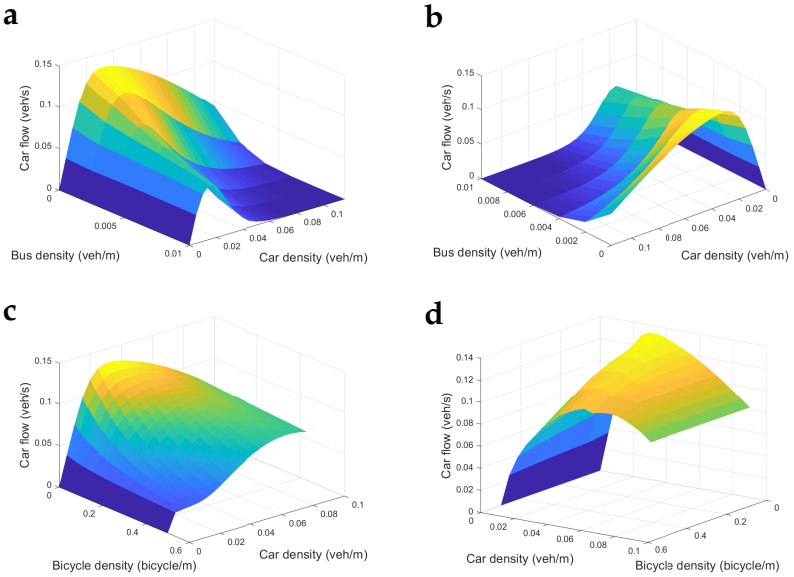


FIGURE 4.22: Car MFDs with interaction effects. **(a)** and **(b)** show the bus interaction effects on car flow; **(c)** and **(d)** show the bicycle interaction effects on car flow.

4.5 SUMMARY

This chapter contributes fourfold to the existing literature. It not only provides the first extensive comparative empirical evidence on the existence of the MFD and show that its shape is indeed a function of network topology, but also proposes three different approaches to model the network effects of local interactions between modes (cars, buses and bicycles) in the MFD: An empirical data driven approach (see Section 4.4.1), a geometric approach (see Section 4.4.2) and Bose-Einstein condensate based approach (see Section 4.4.3).

The proposed approaches are all formulated from a vehicle perspective, but they can be used to investigate the effects on passengers' choices, the so called 3D-passenger-MFD (3D-pMFD). Thus, with known preferences of travelers, choices and their distribution across modes, one can investigate the passenger throughput of entire urban networks and subsequently optimize it. Aside using the MFD for network assignment modeling (see Chapter 5), this passenger perspective is very useful for policy making as we show in two applications build around the 3D-MFD in Chapters 6 and 7.

5

3D-MFD NETWORK ASSIGNMENT PROBLEM

> *The journey times on all the routes actually used are equal, and less than those which would be experienced by a single vehicle on any unused route.*
>
> — Wardrop (1952)

In transport modeling and planning, the well-known four-step model is the workhorse underlying most likely most analyses (Ortúzar and Willumsen, 2011). In detail, the four steps are the trip generation at the origin, trip distribution to the destinations, mode choice and then the network assignment. The network assignment is a mathematical procedure that allocates a given origin-destination demand to a specific transportation system, e.g. the road network or transit network (Patriksson, 1994). Consequently, the traffic assignment takes as an input an origin-destination matrix and computes the vehicle or passenger flows as well as speeds in the network as the output.

In this chapter, we formulate a static stochastic user equilibrium based on the MFD and formulate the problem with regional paths as a mixed complementarity problem. We organize this Chapter as follows. First we review the basic ideas of the traffic assignment in Section 5.1. Second, we introduce to the idea of a network of regions in Section 5.2. Third, in Section 5.3 we formulate the 3D-MFD network assignment problem. Last, in Section 5.4 we discuss further model extensions and the calibration procedure in Section 5.5.

5.1 BACKGROUND

Historically, many transport planners were first focusing on heuristic or ad hoc algorithms for the network assignment, but as Patriksson (1994) formulates "*a lack of a rigorous scientific approach to problem formulations*" by the transport planners prevented that they became aware of the development of algorithms by the mathematicians' community with lead finally to a decade of "*lost opportunities*" (Boyce, 1984). However, this gap is closed now and transport network problems have become part of wider and generalized network problems (Nagurney, 1993; Dafermos and Sparrow, 1969).

Without loosing generality, there are three pairs of adjectives that are widely recognized as key features of traffic assignment problems. First, assignments are considered to be either dynamic or static. Dynamic traffic assignments analyze the evolution of flows over time, while static traffic assignments consider only one time instant. The second pair distinguishes whether assignments are either deterministic (full information) or stochastic (incomplete information, or heterogeneity). In the deterministic case, the route choice rationale is entirely a function of fully known travel time, while in the stochastic case, stochastic effects account for incomplete information on path costs or heterogeneity of preferences. In other words, in the deterministic assignment, the actual path costs are considered, while in the stochastic assignment the *perceived* path costs (Daganzo and Sheffi, 1977; Fisk, 1980). The third pair relates to the quote by Wardrop (1952) at the beginning of this chapter: traffic distributes in the network according to rules, where Wardrop's (1952) rules became standard. He distinguishes between the user equilibrium (UE) in which every driver is minimizing her or his journey time, while the system optimum (SO) is the choice of routes in which total travel time is minimized. In transport networks, usually one assumes the UE, where the SO offers comparability to assess the efficiency of policies. There is further a fourth pair of relevance, but usually less prominent. The assignment problem can be formulated as a link-based or route based problem (Patriksson, 1994).

Starting from mathematical problem formulations in the 1950s (e.g. Prager, 1954; Beckmann *et al.*, 1956), today, Wardrop's standard formulation of the optimization problem from the introductory quote can be formulated as a variational inequality and as complementary problem (Dafermos, 1980; Pang and Harker, 1990; Ferris *et al.*, 1999), see Appendix D for an overview. In the mathematical formulation, two issues, especially in the urban context must be discussed. First, assignment problems usually consider capacity constraints of the system, e.g. a certain link can only accommodate a maximum flow or a transit vehicle only a maximum number of riders. Here, the problem arises as summarized by Bliemer *et al.* (2014) that "*although adding the capacity constraints seems natural, it is not consistent with the link travel time functions [...] such that 'tricks' with Lagrange multipliers are needed*". Second, link-based volume-delay functions have an unrealistic asymptotic behavior near capacity (Patriksson, 1994) leading to unrealistic travel times (Boyce *et al.*, 1981) and computation issues (Daganzo, 1977).

Model-wise, the MFD is similar to a volume-delay function, but in this particular case, it contains in its own formulation already a capacity con-

straint and thus has less asymptotic issues. As the MFD has only been recently formulated, mathematical formulations of the MFD traffic assignments problem are not well established and still subject to research. So far, the literature shows only dynamic MFD traffic assignments (e.g. Leclercq and Geroliminis, 2013; Yildirimoglu and Geroliminis, 2014; Aghamohammadi and Laval, 2018), which is intuitive as this follows a general trend in research (Bliemer *et al.*, 2014) and uses the full potential the MFD offers. Here, we have seen two general approaches (Mariotte *et al.*, 2017): The accumulation based approach (Daganzo, 2007) as well as the trip-based approach (Arnott, 2013; Daganzo and Lehe, 2015; Lamotte and Geroliminis, 2018), which differ most notably in the rule of how the speed information propagates through the network over time. However, the much simpler static assignment seems to have been ignored, although its simplicity offers for many other disciplines already a very powerful tool in strategic transport planning.

5.2 REGIONAL MODEL

For the 3D-MFD network assignment problem, we transform the urban road network into a network of regions. The primary motivation here is that the MFD, defined for a regional network and not single streets, is usually only well-defined in homogeneously loaded in smaller partitions of the road networks (see Section 4.2.1). Consequently, reasonable regional networks must be chosen, e.g. with network partitioning (e.g. Ji and Geroliminis, 2012; Saeedmanesh and Geroliminis, 2016, 2017). Figure 5.1 illustrates this idea. Figure 5.1**a** shows how the urban road network is partitioned into several regional networks, where we can define in each region, or as Daganzo (2007) puts it "*neighborhoods*", a 3D-MFD. Then, we abstract this perspective into a network of regions as shown in Figure 5.1**b** where each region is connected to other zones with interchanging flows of vehicles as well as flows within each zone.

In this chapter and the two following ones, we follow the notion of regional paths as shown in Figure 5.2. Demand is aggregated into macronodes, where travelers originate at i and arrive at j. In each region k, there can be no node at all, a single node or several nodes, but for simplicity, we consider here that each region has only a single node. Travelers choose from i to j their transportation mode m (bus or car) along a regional path r through (several) regions as illustrated in Figure 5.2. In this model, the regional paths are not explicitly mapped to roads as only the macroscopic

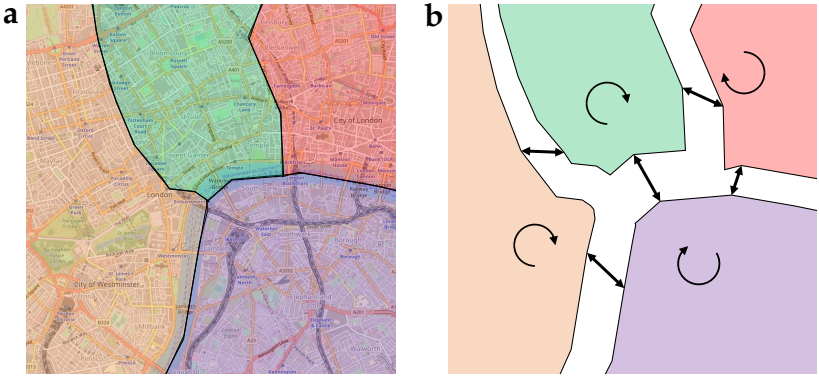


FIGURE 5.1: Illustration of the regional network idea. The background map of **(a)** is provided by OpenStreetMap.

trip distance d_{ijmr} is important to obtain travel times T_{ijmr} (Aghamohammadi and Laval, 2018). Here, we define that θ_{ijkmr} maps regional paths to regions. If its value is zero, the regional path is not traversing through region k, if its value is non-zero, it gives the fraction of the regional path's length d_{ijmr} that is located within k.

The benefit of using regional paths, i.e. a route-based assignment, is that we do not to define a separate node model, as all interactions between vehicles vehicles that lead to delays are already accounted for in the MFD formulation. Contrary, in a route-based assignment, we have to enumerate all regional paths, but given the macroscopic nature of the model, arguably, the number of route alternatives each traveler considers are limited.

5.3 MATHEMATICAL PROBLEM FORMULATION

All parameters and variables introduced in this Chapter are summarized in Table C.4 and Table C.5. We adopt a multimodal traffic assignment, formulated as a stochastic user equilibrium (SUE) following Wardrop's first principle of the user equilibrium based on perceived travel cost. We assume that agents choose mode m and route r with the lowest perceived costs, $\check{C}_{ijmr}$. Thus, a particular route and mode between origin i and destination j is only chosen if its perceived path costs $\check{C}_{ijmr}$ along that route are equal to the minimum path costs, i.e. $M_{ij} \equiv \min_{mr} \check{C}_{ijmr}$. In other words, $N_{ijmr} > 0$ only if its path costs are equal to the minimum cost M_{ij}. If costs

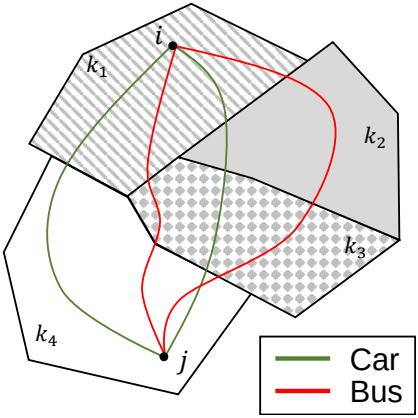


FIGURE 5.2: Illustration of regional paths.

exceed M_{ij}, the route is not used, i.e. $N_{ijmr} = 0$. This feature is captured in the complementary condition of Eqn. 5.1.

$$\check{C}_{ijmr} - M_{ij} \geq 0 \quad \perp \quad N_{ijmr} \geq 0 \tag{5.1}$$

The perceived paths costs are defined according to Eqn. 5.2, where C_{ijmr} are the actual path costs as defined by Eqn. 5.3 and μ_R is the associated scale parameter. Note that Eqn. 5.2 is adopted from (Chen, 1999) and describes a simultaneous route and mode choice multinomial logit model.

$$\check{C}_{ijmr} = C_{ijmr} + \frac{1}{\mu^R} \log\left(N_{ijmr}\right) \tag{5.2}$$

The path costs C_{ijmr} are the generalized cost of travel and combine as given by Eqn. 5.3 the in-vehicle travel time T_{ijmr} for both modes and shadow prices resulting from the capacity constraints, i.e. for parking ρ^P, car ownership ρ^C, bus passenger capacity ρ^B, and season ticket ownership ρ^T. The associated constraints are introduced in Section 5.3.1. Note that the monetary costs for using a particular mode and mobility tools are separately considered in the mobility tool ownership constraints in Section 5.3.2. Further, the path costs contain for buses a general waiting time defined as half the headway H_i in departure zone i. Preferences of agents and other factors that influence the choice as well are captured in $\overline{\varphi}_{ij}$ that requires calibration from data.

$$\begin{aligned} C_{ij,\text{car},r} &= T_{ij,\text{car},r} + \rho_i^C + \rho_j^P \\ C_{ij,\text{bus},r} &= T_{ij,\text{bus},r} + \frac{H_i}{2} + \rho_i^B + \rho_i^T + \overline{\varphi}_{ij} \end{aligned} \tag{5.3}$$

Travel times T_{ijmr} are calculated with Eqn. 5.4 that is the sum of travel times within each region along each route r. Here, V_{km} is the journey speed of mode m in sub-region k, $\theta_{ijkmr} d_{ijmr}$ equals the trip distance in region k. For buses, the travel time contains the in-vehicle time including dwelling behavior.

$$T_{ijmr} = \sum_k \theta_{ijkmr} \frac{d_{ijmr}}{V_{km}} \tag{5.4}$$

We derive the journey speeds from the 3D-MFD (Geroliminis *et al.*, 2014; Loder *et al.*, 2019). The 3D-MFD links the current accumulation of cars, $A_{k,car}$, and buses, $A_{k,bus}$, to the average speed of mode m in region k as formulated by Eqn. 5.5. The shape of the 3D-MFD results from the features and topology of the road and bus networks. Consequently, when changing the network design variables of the 3D-MFD-NDP, the 3D-MFD will change and thus affect the speeds in the network.

$$V_{km} = \text{3D-MFD}_{km} \left(A_{k,car}, A_{k,bus} \right) \tag{5.5}$$

In our traffic assignment model, we cannot use Edie's (1963) definition to calculate the accumulation or density of vehicles as there is no distinct time interval. Therefore, we calculate each modes' vehicle accumulation differently. For the accumulation of cars, $A_{k,\text{car}}$, we assume a vehicle occupancy of one passenger and obtain the accumulation by Eqn. 5.6, where $\theta_{ijk,\text{car},r}$ is the fraction of trip length of that particular route going through region k.

$$A_{k,\text{car}} = \sum_{ijr} \theta_{ijk,\text{car},r} N_{ij,\text{car},r} \tag{5.6}$$

We derive the accumulation of buses, $A_{k,\text{bus}}$, from the structure and design of the bus network (adopted from Daganzo (2010)) as given by Eqn. 5.7. Here, α_k is an exogenous parameter describing the design of the bus network in each region for which $0 < \alpha_k \leq 1$ holds. Close to its lower bound, α_k describes a hub-and-spoke network, while at its upper bound it describes a grid network. Values in between are hybrid networks where

one can see α_k as the fraction of network exhibiting a grid network. As in many cities bus lines are partially overlapping, we introduce z_k that quantifies how many bus lanes are overlapping on the bus infrastructure B_k. In case no bus lines are overlapping, $z_k = 1$, if two bus lines are overlapping on the entire network, then $z_k = 2$ and so on. Last, $V_{c,\text{bus}}$ is the design commercial speed of buses in the network.

$$A_{k,\text{bus}} = z_k \frac{2B_k}{H_k} \frac{3\alpha - \alpha^2}{1 + \alpha^2} / V_{c,\text{bus}} \tag{5.7}$$

Lastly, Eqn. 5.8 then provides the conservation of passenger flows for each origin and destination pair. M_{ij} is the associated complementary variable. Importantly, as we are formulating the 3D-MFD-NDP based on regional paths, there is no requirement to explicitly model the in- and outflows at each node as in a link based assignment.

$$n_{ij} - \sum_{mr} N_{ijmr} = 0 \quad \perp \quad M_{ij} \geq 0 \tag{5.8}$$

5.3.1 *Physical system (capacity) constraints*

The static traffic assignment model has a set of four inequalities that describe the physical constraints of the system and which have associated shadow price variables that factor into the path costs. The first constraint describes the parking supply in each zone as given by Eqn. 5.9. Here, $\overline{P}_j$ is the parking supply in zone j, an exogenous parameter. Consequently, the total arrival car passenger flow cannot exceed the parking supply. If the parking demand exceeds the parking supply, then a non-zero shadow price, ρ_j^P, will ensure that the number of arriving cars is restricted to the parking supply.

$$\overline{P}_j - \sum_{ir} N_{ij,\text{car},r} \geq 0 \quad \perp \quad \rho_j^P \geq 0 \tag{5.9}$$

The second and third inequality captures the model's property that all departing trips of a zone of a certain mode must not exceed the availability of mobility tools in that zone. In other words, the number of car trips starting in i cannot be greater than the number of available cars in i as given by Eqn. 5.10. Similarly, the number of outbound bus passenger trips cannot exceed the number of public transport season-tickets, or in brief *abos*[1],

1 Abbreviation of the German word Abonnement.

in that zone as formulated in Eqn. 5.11. In these equations, Q_{ijt} describe the shares of mobility tool ownership, where the elements of set t are having only a car, an abo or having both. The calculation of Q_{ijt} is discussed along with Eqn. 5.15. When the inequality becomes binding, the respective shadow prices ρ_i^C and ρ_i^T become non-zero.

$$\sum_j \left(Q_{ij,\text{car}} + Q_{ij,\text{both}}\right) n_{ij} - \sum_{jr} N_{ij,\text{car},r} \geq 0 \quad \perp \quad \rho_i^C \geq 0 \tag{5.10}$$

$$\sum_j \left(Q_{ij,\text{abo}} + Q_{ij,\text{both}}\right) n_{ij} - \sum_{jr} N_{ij,\text{bus},r} \geq 0 \quad \perp \quad \rho_i^T \geq 0 \tag{5.11}$$

The fourth inequality describes that the passenger capacity of the bus system is limited to a total passenger accumulation of Z_k as formulated in Eqn. 5.12. The total bus passenger flows have to be always less or equal to that capacity. When supply equals demand, public transport users experience additional waiting time ρ_k^B in their departing zone as all arriving buses are full.

$$Z_k - \sum_{ijr} \theta_{ijk,\text{bus},r} N_{ij,\text{bus},r} \geq 0 \quad \perp \quad \rho_k^B \geq 0 \tag{5.12}$$

The total bus passenger capacity in a region Z_k results from the accumulation of buses in each region according to Eqn. 5.13. Note that overline variables denote observed calibration values.

$$Z_k = \overline{Z}_k \frac{A_{k,\text{bus}}}{\overline{A}_{k,\text{bus}}} \tag{5.13}$$

5.3.2 *Mobility tool ownership constraints*

The shares of mobility tool ownership Q_{ijt} change with the prices of the chosen portfolio t, π_{ijt}^{total}, in a two stage logit-based choice. In this choice environment, individuals have three mobility tool portfolio options to choose from: only a car, only an *abo* (recall that this is a public transport season-ticket), or both. This simplifies the complexity of choices for mobility tools typically available to individuals, as shown, for example, in Switzerland (Becker *et al.*, 2017c; Loder and Axhausen, 2018). Note that for readability we omit the set indices ij for Q and π in the following two equations.

We define the utility functions for the logit model as given in Eqn. 5.14. The alternative specific constant (ASC) is related to the calibrated market

share $\overline{Q}_{ijt}$. Utility changes with changes relative to the calibration prices with scale parameter μ^M that captures the price elasticity of mobility tool ownership. Eqn. 5.15 then presents the two-stage logit-based model to obtain the shares of mobility tool ownership. The first stage determines the shares of having both or not both mobility tools (car and abo), while the second stage determines the shares between car and abo owners along those not having both mobility tools. Note that the formulation of a logit model ensures that the shares always add up to one.

$$\begin{aligned} u_{\text{both}} &= \log\left(\overline{Q}_{\text{both}}\right) + \left(\pi^{\text{total}}_{\text{both}} / \overline{\pi}^{total}_{\text{both}} - 1\right) / \mu^M \\ u_{\text{not both}} &= \log\left(1 - \overline{Q}_{\text{both}}\right) \\ u_{\text{car}} &= \log\left(\overline{Q}_{\text{car}}\right) + \left(\pi^{\text{total}}_{\text{car}} / \overline{\pi}^{\text{total}}_{\text{car}} - 1\right) / \mu^M \\ u_{\text{abo}} &= \log\left(\overline{Q}_{\text{abo}}\right) + \left(\pi^{\text{total}}_{\text{abo}} / \overline{\pi}^{\text{total}}_{\text{abo}} - 1\right) / \mu^M \end{aligned} \tag{5.14}$$

$$\begin{aligned} Q_{\text{both}} &= \frac{\exp\left(u_{\text{both}}\right)}{\exp\left(u_{\text{both}}\right) + \exp\left(u_{\text{not both}}\right)} \\ Q_{\text{car}} &= \left(1 - Q_{\text{both}}\right) \frac{\exp\left(u_{\text{car}}\right)}{\exp\left(u_{\text{abo}}\right) + \exp\left(u_{\text{car}}\right)} \\ Q_{\text{abo}} &= \left(1 - Q_{\text{both}}\right) \frac{\exp\left(u_{\text{abo}}\right)}{\exp\left(u_{\text{abo}}\right) + \exp\left(u_{\text{car}}\right)} \end{aligned} \tag{5.15}$$

The average price π^{total}_{ijt} or total cost of ownership for mobility tool portfolio t between i and j is calculated with Eqn. 5.16. Here, F_{ijtm} gives the fraction of using mode m with mobility tool set t (defined below in Eqn. 5.17) and the last term in parentheses gives simply the average trip distance. Recall that the fix costs or price per mode m with mobility tool portfolio t is π^{fix}_t, while the variable and distance-depending prices for mode m and mobility tool portfolio t is π^{var}_{tm}. For simplicity in this analysis, we do not distinguish between different price or cost components and subsume all taxes, fares etc. under the term costs. Further, for cars we do not account for the variable operating costs (as we do not study agents' income), but only the fixed costs, e.g. car registration and fees for parking at the destination.

$$\pi^{total}_{ijt} = \pi^{\text{fix}}_t + \sum_m \pi^{\text{var}}_{tm} F_{ijtm} \left(\sum_r \frac{d_{ijmr}}{|\mathcal{R}|} \right) \tag{5.16}$$

The fraction F_{ijtm} using mode m with mobility tool set t is defined according to Eqn. 5.17. I_{tm} is an indicator function that equals one if $t =$ abo & $m =$ bus or $t =$ car & $m =$ car, and equals zero otherwise. In other words, when owning only either an abo or a car, only the respective mode can be used, i.e. $F_{ijtm} \equiv 1$, and the other mode cannot be used, i.e. $F_{ijtm} \equiv 0$. Only when having both mobility tools, F_{ijtm} can be different from zero or one as shown in Eqn. 5.17. Then, F_{ijtm} is simply the number of mode m agents over all agents for each origin-destination pair having both mobility tools.

$$F_{ijtm} = \begin{cases} \dfrac{\sum_r N_{ijmr} - n_{ij} \sum_{t'} I_{t'm} Q_{ijt'}}{Q_{ijt}\, n_{ij}}, & \text{if } t = \text{both} \\ I_{tm}, & \text{otherwise} \end{cases} \tag{5.17}$$

5.4 FUTURE MODEL EXTENSIONS

The introduced 3D-MFD network assignment is static, while the MFD itself is suitable for modeling the dynamics of macroscopic congestion (Mariotte *et al.*, 2017) and first approaches to the dynamic assignment have already been proposed for multi-regional models (Laval *et al.*, 2018; Yildirimoglu and Geroliminis, 2014). However, the literature just started to explore how to model dynamic regional models with stochasticity in the path lengths due to detours and the resulting distribution of speeds (e.g. Batista and Leclercq, 2018a,b).

In this regard, we have assumed that the number of regional paths or routes r is a priori known and their length determined as well as that all alternatives are available in the choice model. However, as in other models of this kind, the generation of suitable regional paths based on available shortest paths - and their variability - is something to be considered in further model expansions.

Certainly, the modeling framework with the constraints of mobility tool ownership and including the prices in this choice is a different approach to existing models where usually the full paths costs are used. Alternatively, the mobility tool ownership model can be ignored if all prices are included in the actual path costs C_{ijmr}, but then the effects of having both mobility tools available cannot be reflected anymore.

Last, in larger metropolises, typically larger elevated and underground networks such as motorways and subways exist that operate segregated from other transport modes. As these modes consequently do not physi-

cally interact with cars and buses, they do not require a representation in the 3D-MFD. However, the 3D-MFD network assignment can be extended to accommodate trips that partially use elevated or underground networks. The key idea is to define alternative routes, where a fraction of d_{ijmr} is not assigned to a zone with a 3D-MFD, but the flow of passengers is assigned in a conventional procedure to the elevated or underground networks, where the delays are returned and added to the travel times. For public transport services, transfer and walking times have to be included as well.

5.5 MODEL CALIBRATION

The present model requires calibration in order to represent an observed equilibrium. For this, we require the following reference information.

1. Origin and destination matrix, n_{ij}.
2. Network topology for bus and road networks.
3. Speeds V_{km} and accumulations A_{km}, as well as regional public transport capacity Z_k.
4. Mode shares between i and j.
5. Mobility tool ownership levels, Q_{ijt}, and price levels, π_{ijt}^{total}.
6. behavioral choice parameters μ^R and μ^M.

In the following, we discuss each of these calibration steps. At first, the reference number of agents between i and j, n_{ij} is important in the node balance as formulated in 5.8. There are various ways of obtaining theses matrices from a variety of data sources with methods well documented in literature (e.g. Willumsen, 1978; Friedrich *et al.*, 2010). Therefore, we consider here that n_{ij} is given and can be aggregated to the regional node model as introduced in Section 5.2.

Second, the 3D-MFDs - as mathematically formulated in Section 4.4.2 - need to be estimated based on the topology of road and bus networks as well as the operational features. Here, the present zoning as used in the regional model together with the spatial information from, e.g., OpenStreetMap, can be used to calculate all required parameters from Table B.4. For the operational features, time-table information as well as measurements from traffic signals need either to be measured or obtained from private communication, e.g. with the city's traffic engineers.

Third, for the benchmark equilibrium we require observed vehicle accumulation of both modes as well as speeds to obtain the travel times during the morning peak. The MFD analyst needs to investigate at which time instant - on average - the network is loaded most so that the average network loading and travel time from that time period is used for the calibration. For cars, the space mean speed at peak hour can be obtained from floating car data. We propose to estimate in each region the space mean speed along several shortest paths and then take the weighted mean thereof. For public transport, this data can be readily provided by the agency as they usually record trajectories or past arrival and departure times of all vehicles (see Loder *et al.* (2017) for further details). The regional public transport capacity Z_k can be calibrated in the benchmark equilibrium by either multiplying the number of vehicles in each region by an approximation of the passenger capacity of each vehicle. A robustness check in the calibrated model is then to check whether the final passenger flows are below or equal to the calibrated passenger numbers. If the gap between both figures is too wide, the analyst should then calibrate the model to a capacity that is tighter to the observed flows than to an estimated total passenger capacity. Importantly, one can expect that the road network is not fully utilized, e.g. at intersections etc, and consequently we need to scale the network length to the observed speeds and flows in the MFD to account for the inefficiency in the use of infrastructure.

Fourth, each origin and destination pair has in addition to observed travel times a variety of unobserved factors that influence mode choice. These factors are summarized in φ_{ij} as introduced in Eqn. 5.3. Here, we propose to use observed mode shares for each origin and destination pair or in the accumulation in each region to calibrate the cost functions with a non-linear program by finding the optimal φ_{ij} values that result in the observed outcomes using the idea by van Nieuwkoop (2014).

Fifth, as we propose to model changes in mobility tool ownership around the benchmark as a function of prices as formulated in Eqn. 5.15, we need to identify $\overline{Q}_{ijt}$. This can be done with using data from a travel behavior survey, either for origin and destination pairs or if that is not possible only for the trip originating zone. If that kind of data is not accessible, the shares of car ownership from a census could be used to approximate the shares. We also need to calibrate the benchmark price levels as defined in Eqn. 5.16 with fixed and variable components. Both components should be provided on a daily basis as all monetary flows here are referenced on a per day basis.

Last, we need to calibrate the behavioral parameters or choice elasticities, namely the price elasticity of mobility tool ownership μ^M as well the scale parameter for the combined route and mode choice model μ^R. While the price elasticity μ^M can be typically obtained from stated preference experiments or similar panel surveys for single elements of costs, e.g. fuel or fares (e.g. Goodwin, 1992; Litman, 2012). As we consider the full costs P_{ijt} we have to define a composite elasticity. Regarding the scale parameter μ^R, additional information regarding the perception of route costs is required, e.g. from experiments or surveys. If this is not available, we propose to fix the value to unity or any other reasonable number.

5.6 SUMMARY

This chapter introduced a static traffic assignment problem for a multi-region 3D-MFD model with a stochastic user equilibrium and explicit mobility tool ownership model in mixed complementarity problem formulation. The simplicity of the model formulation allows fast computation and thus integration in applications as the following two chapters will show.

6

OPTIMAL INFRASTRUCTURE IN CITIES

> *In theory, the most recent computer models can predict almost anything on a multimodal transportation network in minute detail, but not in practice.*
>
> — Daganzo (2007)

For cities to provide optimal mobility, infrastructure investment and pricing choices are difficult. This is not only because of high costs or strong public opposition, but also but also because population and income growth as well as changes in preferences will inevitable change the expected demand patterns and desired outcomes. Although it might be intuitive to argue that we most likely have too many cars in our cities, and we know that we can reduce with collective efforts our car dependence (Newman and Kenworthy, 1991; Buehler *et al.*, 2016), it is unclear by how much we need to change. Consequently, urban and transport planners not only require guidance to understand how their infrastructure investments and pricing policies affect travel behavior, but also to derive optimal investment and pricing strategies to "*enhance social wellbeing - the social side of life that is all too often ignored*" (Fisk, 2000) or increase productivity of cities.

The geometric approach to estimate the 3D-MFD from Section 4.4.2 and the 3D-MFD network assignment presented in Chapter 5 provide an opportunity to investigate the question of which infrastructure and pricing decisions improve mobility (or productivity) at urban scale. Such macroscopic perspectives help to understand global mechanisms that explain why cities exist (Bettencourt, 2013), but they do not provide information on where to exactly build a road or place a bus stop. Thus, to analyze the question of optimal infrastructure and prices for a city we formulate in this chapter the 3D-MFD network design problem (3D-MFD-NDP) as a mathematical problem with equilibrium constraints.

We organize this chapter as follows. In Section 6.1, we first review the ideas of road network and bus network design problem, respectively. Second, in Section 6.2 we then provide the mathematical formulation of the 3D-MFD-NDP. Third, in Section 6.3 we apply the 3D-MFD-NDP to the city of Zürich to illustrate the mechanisms of the model.

6.1 BACKGROUND

The 3D-MFD-NDP originates - as emphasized by its name - from the MFD and its multimodal extension to the 3D-MFD, but also from the classic (continuous) network design problem and the bus network design problem. The question of optimal network design for a city, i.e. given distribution of demand, is a key question in urban and transport planning. Consequently, it has been intensively studied. Thus, for in-depth reviews on the network design problems we refer to Boyce (1984), Magnanti and Wong (1984), Friesz (1985), Migdalas (1995), Yang and Bell (1998), Guihaire and Hao (2008), Kepaptsoglou and Karlaftis (2009) and Farahani *et al.* (2013). In this section, we then concentrate on the main ideas and definitions of the road network design problem in Section 6.1.1 and the transit network design problem in Section 6.1.2.

6.1.1 *Road networks*

The common network design problem (NDP) has the objective of finding the optimal network configuration for a given (elastic) demand, where the objective function is typically related to maximize social welfare or minimize total travel time (Yang and Bell, 1998). In literature, three different groups for the NDP have been established: the discrete-NDP, e.g. adding single links to the network, the continuous-NDP, e.g. capacity improvements to existing infrastructure (e.g. Adbulaal and LeBlanc, 1979; Marcotte, 1986), then obviously, also a mixed NDP, combining discrete and continuous ideas, can be applied as well.

In non-heuristic approaches, the NDP typically encompasses a bi-level optimization problem. At the lower level, given demand is assigned to the network, which physical characteristics are subject to the upper level investment and management decisions that satisfy the objective of the network design problem. Consequently, there are two sets of constraints to the system. The first set of constraints limits the solution space to all (physically) feasible solutions of the system. The second set of constraints is the lower level as only solutions that satisfy these constraints can be considered for the upper level optimization (Yang and Bell, 1998).

6.1.2 *Transit networks*

For the optimal design of transit networks, Ceder and Wilson (1986) propose three levels: (i) strategic planning in terms of network design, (i) tactical planning, e.g. frequency setting, and (ii) operational planning, e.g. vehicle scheduling. The question of optimal transit networks started with heuristic methods (e.g. Patz, 1925; Sonntag, 1977). Later, mathematical problems were defined to identify optimal network and frequency configurations (e.g. Salzborn, 1972; Schéele, 1980). The recent literature reviews by Guihaire and Hao (2008) and Ibarra-Rojas *et al.* (2015) provide a comprehensive overview on all three mentioned levels of the transit network design problems. A common strategy to improve bus services is the provision of dedicated bus lanes such as cars and buses encounter less conflicts. Consequently, the identification of optimal allocation of dedicated bus lanes is a further relevant dimension to improve mobility (e.g. Hadas and Ceder, 2018; Yu *et al.*, 2015; Zhang *et al.*, 2018; Dantsuji *et al.*, 2017; Zheng *et al.*, 2017).

Recently, the 3D-MFD also has already been used for discussing the optimal design of transit networks, the space distribution among both modes and the placement of dedicated bus lanes (e.g. Amirgholy *et al.*, 2017; Zheng *et al.*, 2017; Zheng and Geroliminis, 2013; Zhang *et al.*, 2018; Dantsuji *et al.*, 2017). However, neither of the 3D-MFD approaches obtained the shape of the 3D-MFD directly, i.e. endogenously, from road and bus network topology, but obtained its shape exogenously from running simulations.

6.2 THE 3D-MFD NETWORK DESIGN PROBLEM

In this 3D-MFD-NDP, we are looking for a network configuration and mobility pricing strategy that minimizes total system cost, subject to the constraints that the existing demand is assigned to the network, that monetary expenditures for infrastructure equal revenue, and that the system's physical constraints are satisfied. Consequently, we formulate the 3D-MFD network design problem as a mathematical program with equilibrium constraints (MPEC) (Luo *et al.*, 1996), because the upper level objective is to minimize total travel time costs and subsidy in the network and the lower level objective is a multimodal user equilibrium traffic assignment (see Appendix D.3 for details). The design (free) variables of the 3D-MFD-NDP are (i) network length L, (ii) share of dedicated bus lanes η_b, (iii) the bus

service frequency or headway H, (iv) the fixed price of mobility tools π^{fix}, i.e for the car or public transport season-ticket, (v) the variable costs per unit distance π^{var}, (vi) and the system's (public) transport subsidy S.

The mathematical formulation of the 3D-MFD-NDP uses Eqns. 5.5 - 5.8 from Chapter 5 for the traffic assignment, but requires the definition of the upper level objective function (Section 6.2.1) and of further economic constraints (Section 6.2.2). With these functions, the mathematical problem is formulated in Section 6.2.3.

6.2.1 *Objective function*

The upper level objective y of the 3D-MFD-NDP is defined in Eqn. 6.1. Here, the first part is the total travel time of all passengers. The second part of the objective function equals to the total system subsidy S that is used to fund the city's mobility assets: road infrastructure and buses. To value both parts with the same units, we transform travel time to monetary terms using the city's value of time VOT.

$$y = VOT \sum_{ijmr} N_{ijmr} T_{ijmr} + S \tag{6.1}$$

The costs for mode use will be accounted for in the income balance in Section 6.2.2. They are not accounted for in Eqn. 6.1 for two main reasons. First, they have to be paid by agents anyway to cover the infrastructure expenses and thus only the subsidy matters. Second, this model does not include a mechanism for the agents to generate income to reflect their trade-offs in time and money.

6.2.2 *Economic constraints*

The constraints introduced in this section restrict solutions to the 3D-MFD-NDP where the revenue from mobility (ownership π_t^{fix} and use π_{tm}^{var}) and the subsidy S equals the operational costs O for the city's mobility assets. This operational costs O for the city's mobility assets, i.e. the costs for the provision of roads and bus operations, are calculated with Eqn. 6.2. Here, c_k^{road} and c_k^{bus} are the unit prices for the provision of infrastructure and buses, respectively. Consequently, the totals then depend on the size of the network L_k and number of buses $A_{k,\text{bus}}$.

$$O = \sum_k c_k^{bus} A_{k,\text{bus}} + c_k^{road} L_k \tag{6.2}$$

Then, the income balance of costs, revenue and subsidy is mathematically expressed in Eqn. 6.3, where Q_{ijt} corresponds to the shares of mobility tool ownership for that particular origin-destination pair (see derivation in Section 5.3.2). Here, we assume that the total revenue counts towards the available budget for infrastructure spending, although in reality this is too simplistic as public funding and budgets are usually more complex (see our discussion in the calibration of the model in Section 6.3.1).

$$\sum_{ijt} \pi_t^{\text{fix}}\, n_{ij} Q_{ijt} + \sum_{ijmr} \pi_{tm}^{\text{var}} N_{ijmr} d_{ijmr} + S = O \tag{6.3}$$

The costs for buses c_k^{bus} and roads c_k^{road} can be subject to (dis-) economies of scale. Therefore, we account for this by the formulations for both costs in Eqn. 6.4 and 6.5, respectively. The cost functions are calibrated to market values that are indicated by an overline to the cost function's elasticity ε. In case $\varepsilon = 0$, the costs are constant, while for $\varepsilon > 0$ diseconomies of scale, and for $\varepsilon < 0$ economies of scale result.

$$c_k^{road}\left(L_k\right) = \overline{c}_k^{road} \exp\left(\varepsilon^{road} \log\left(L_k / \overline{L}_k\right)\right) \tag{6.4}$$

$$c_k^{bus}\left(A_{\text{bus},k}\right) = \overline{c}_k^{bus} \exp\left(\varepsilon^{bus} \log\left(A_{\text{bus},k} / \overline{A}_{\text{bus},k}\right)\right) \tag{6.5}$$

6.2.3 *Mathematical problem formulation*

The objective function y in Eqn. 6.1, the traffic assignment defined in Eqns. 5.5 - 5.8 and the economic constraints defined in Eqns. 6.2-6.5 allow to formulate the 3D-MFD-NDP as formulated in Eqn. 6.6: The 3D-MFD-NDP is looking for the solution of the network design and pricing variables that reduces the total system costs (travel time and subsidy), subject to the constraints that the existing demand is assigned to the network, that the physical capacity constraints are satisfied and that monetary expenditures equal revenue.

$$
\begin{aligned}
&\text{minimize} && y \\
&\text{subject to} && \text{(5.1) - (5.8)} && \text{solving MCP} \\
&\text{and} && \text{(5.9) - (5.13)} && \text{solving capacity constraints} \\
&\text{and} && \text{(5.15) - (5.17)} && \text{solving ownership constraints} \\
&\text{and} && \text{(6.2) - (6.5)} && \text{solving economic constraints}
\end{aligned}
\tag{6.6}
$$

The model formulation in Eqn. 6.6 provides an opportunity to understand the interaction of multimodal vehicle physics and behavioral response.

6.3 APPLICATION TO GREATER ZÜRICH

6.3.1 *Calibration*

The 3D-MFD-NDP requires calibration to an observed point, not only because of $\overline{Q}_{ijt}$, $\overline{\pi}^{total}_{ijt}$, and $\overline{\varphi}_{ij}$, but also because to provide meaningful starting values and to determine a meaningful solution space.

In this paper, we calibrate the 3D-MFD-NDP to the morning commute in the greater region of Zürich, Switzerland. Figure 6.1 shows the extent of the case study area. We partition the network into two regions: Zones 1-12 denote the city where the commuters live and work and zones 101-111 are zones where commuters live and have to commute into the city, i.e. zones 1-12 to work. We investigate pricing effects for all commuters, i.e. living in all zones, but investigate investment effects only to the infrastructure in the city, i.e. zones 1-12. In this model, zones 1-12 each exhibit a 3D-MFD as formulated by Loder *et al.* (2019) to obtain the speeds as given in Eqn. 5.5. The speeds for zones 101-111 are considered to be fixed, i.e. independent of demand. The zones 101-111 are added to the model to capture the influence of suburban commuting into the city.

For the origin and destination matrix n_{ij} we use the commuting matrix of a synthetic Swiss population for the agent-based simulation MATSim (Bösch *et al.*, 2016). We re-scale the total arrivals in each of the twelve inner zones of MATSim's commuter matrix to correspond to the work place totals used by the national transport model (NPVM).

We obtained spatial information on the regional paths for both modes from the Google directions API: For each origin and destination pair, we

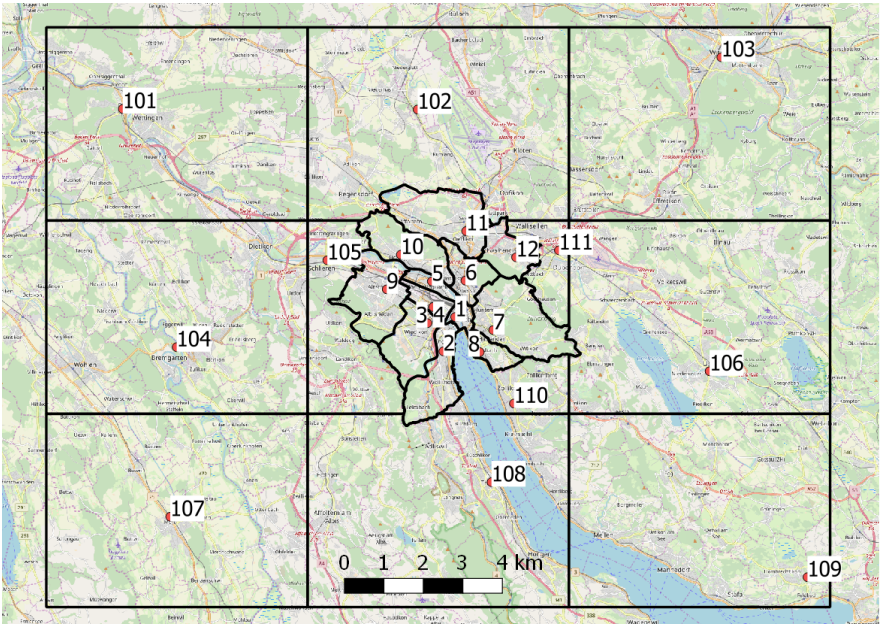


FIGURE 6.1: Zonal system for the case study.

requested the shortest route including alternative routes without using the motorway and calculated θ_{ijkmr} thereof. We calibrated the 3D-MFD speed functions based on OpenStreetMap data and measurements based on the data used by Loder *et al.* (2017). From the 2015 Swiss travel survey we obtained the mode shares of outbound trips of each region (Swiss Federal Statistical Office and Swiss Federal Office for Spatial Development, 2017). With all this information, we solved for each origin and destination pair a nonlinear programming problem minimizing the squared difference between the observed mode share and resulting mode choice from Eqn. 5.8 to calculate $\overline{\varphi}_{ij}$ of the path costs.

For Zürich, we consider the following mobility tool portfolio and pricing situation: When having a car, commuters face fix costs as well as variable costs per unit distance. When having a bus season-ticket (or as defined here "*abo*"), commuters face only fix costs, but no variable costs. In other words, it is not possible to purchase single ride or distance-dependent bus tickets. This situation reflects the situation in Switzerland where most public transport commuters own a season-ticket. When having both mobility tools, commuters have to pay the costs of both single mobility tools. We obtain from the 2015 Swiss travel survey the shares of mobility tool ownership $\overline{Q}_{ijt}$ based on all commuters living in the case study zones. Note that we assign all commuters without a car or abo (season-ticket) to the abo category. For most origin-destination pairs, this share was less than ten percent of the total origin-destination demand.

In Table 6.1, we summarize the economic and behavioral parameters used for the model and its benchmark. Note that in solving the 3D-MFD-NDP MPEC, we will solve it once with economies of scale of investment in roads and buses as well as assuming no economies of scale. We further comment in Table 6.1 on where or how the values are obtained.

In Table 6.2 we summarize the benchmark model's most relevant performance indicators. For the situation in Zürich as shown in Figure 6.1 it is important to explain how the mobility tool revenue relates to the public budget. We define for inbound commuters that they only pay a fraction of their mobility tool expenses to the city's transport budget. The fraction is determined by the fraction of their trip length in the city. We further define that 60 % of the ticket revenue goes to the budget, while we consider the remaining part is going to the railway operators (not considered in this analysis) and to the cantonal transport agency. Similarly, we assume for the car that only 25 % of the revenue is directed to the city's budget, while

Parameter or observed value	Unit	Value
$\bar{c}^{bus}$	(CHF/day bus)	3100[1]
$\bar{c}^{road}$	(CHF/day km)	1900[2]
ε^{bus}	(-)	$\{0; -0.2\}$[7]
ε^{road}	(-)	$\{0; 0.2\}$
$\overline{\pi}^{\text{fix}}_{abo}$	(CHF/day)	3[3]
$\overline{\pi}^{\text{var}}_{abo}$	(CHF/km)	0[4]
$\overline{\pi}^{\text{fix}}_{car}$	(CHF/day)	5[5]
$\overline{\pi}^{\text{var}}_{car}$	(CHF/km)	0.1[6]
μ^R	(-)	10
μ^M	(-)	-0.5[9]
VOT	(CHF/h)	25[8]

[1] Zürich's VBZ has annual expenditures of 600 million CHF for their operation of 470 vehicles in 2016.
[2] Swiss average of annual expenditures of cities `https://opendata.swiss/de/dataset/statistik-der-schweizer-stadte-strassenrechnung`
[3] Costs for annual ZVV pass (season-ticket) for three zones approx CHF 1200, divided by 365 working days.
[4] Set to zero as we focus on the option of season-tickets.
[5] Assuming fix annual taxes, fees etc. of CHF 800, and annual parking costs of CHF 1.200, in total CHF 2.000, divided by 365 working days.
[6] For urban traffic, assuming 8 litre per 100km, a fuel price of CHF 1.30 per litre.
[7] Bösch *et al.* (2017) reported average discounts of 20% for fleet operators.
[8] Estimated by Schmid (2019).
[9] Average fuel price elasticity estimated by Erath and Axhausen (2010).

TABLE 6.1: Price and cost information for the calibrated 3D-MFD-NDP.

Observed value	Unit	Value
$\sum \overline{L}$	(km)	1400[1]
$\sum \overline{A}_{\text{bus}}$	(bus)	1040[2]
$\overline{H}$	(h)	0.125
$\overline{S}$	(CHF/day)	48'000[3]
Share of car trips	(-)	0.33[4]
$\overline{Q}_{\text{car}}$	(-)	0.18
$\overline{Q}_{\text{both}}$	(-)	0.35
$\overline{Q}_{\text{abo}}$	(-)	0.47

[1] Includes the city's and the canton's road network, the infrastructure of public transport and the space on the high capacity roads. The calibration of the 3D-MFD to observed speeds, L is overestimated by around 200 km.
[2] This vehicle number is twice as much as the bus operator owns, but this results from the problem that we cut bus routes at the zones from Figure 6.1 and consequently count vehicles several times.
[3] Calculated as the difference between infrastructure costs and revenue from mobility tools.
[4] Calculated based on travel kilometers.

TABLE 6.2: Transport network calibration values for the 3D-MFD-NDP.

the remaining revenue goes to private companies (e.g. petrol stations) and other tax purposes (e.g. the CO_2 tax).

Lastly, we have to calibrate or define meaningful upper and lower bounds for all design variables. Arguably, the model is calibrated to the current situation and solutions far off this situation, e.g. building twice as many roads, are physically not feasible. Therefore, we allow only for changes in the length of the road network of $\pm 10\,\%$ of its length. We further limit the length of dedicated bus infrastructure to 90 % of the actual network length of buses because some interactions with cars, e.g. at intersections are unavoidable. For the prices, we set the following upper values $\pi^{\text{fix}}_{abo} = 20$ (CHF / day), $\pi^{\text{fix}}_{car} = 50$ (CHF / day), and $\pi^{\text{var}}_{car} = 0.1$ (CHF / km). We further bound the number of bus services with a headway between 1 and 12 minutes.

6.3.2 *Results*

In this section, we solve the MPEC for the 3D-MFD-NDP to illustrate feasibility of the model. We investigate the optimal system configuration and its costs compared to the calibration point, introduced in the previous section.

In this scenario analysis, we set all design variables, i.e. network length, dedicated bus lanes, bus frequency, fixed costs for car and bus season-ticket (the abo option), variable car costs as well as the subsidy) as free variables. In Figure 6.2a we show the relative changes of the design variables compared to the calibration point and in Figure 6.2b several global indicators of the total system performance. In general, we find that having constant prices and economies of scale leads to a similar response pattern with only small differences between both cases. However, the existence of scale economies lead to a larger reduction in travel time.

In Figure 6.2a we find that costs for the car increase sharply, so are the costs for the season-tickets. These increases compensate for the subsidy reduction to its lower bound, but also to shift demand from cars to the bus system as seen in the changed shares of mobility tool ownership as seen in Figure 6.2b. The substantial increase for the variable car costs can also been seen as congestion pricing. For the bus system, the optimal solution suggests to increase the share of dedicated bus lanes as well as the bus frequency. In presence of scale economies, both effects are strengthened. Notably, the total network length experiences almost no changes, but as the share of dedicated bus lanes is increasing, less space for cars is available. The subsidy is in both scenarios driven to its lower bound because this variable can be reduced without consequence of any other budget at the expense of higher prices for mobility tools. However, note that in reality the subsidy is paid by the overall public budget with results from tax revenue paid by agents, but this loop is omitted in the 3D-MFD-NDP.

The consequences of the optimal solution are savings in travel time of almost 20 % in presence of scale economies, a more than 70 % reduction in car kilometers travelled, and increased costs for the mobility assets (road infrastructure and the operations of buses) and user costs. With scale economies present, the cost increase is intuitively smaller.

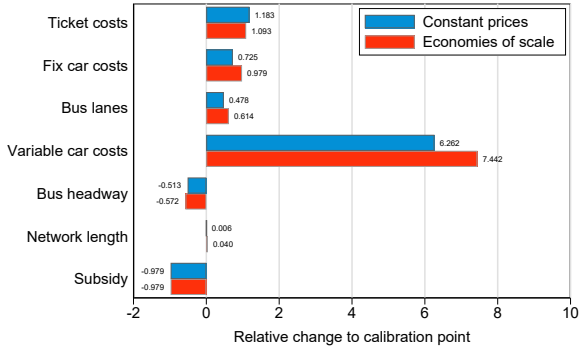
Lastly, to illustrate the system changes on the shape of the 3D-MFD we plot the calibrated and resulting 3D-MFD for zone 1 (see Figure 6.1) in Figure 6.3. It can be clearly seen that the reduction of 15 % in road infrastructure decreased the total capacity of the zone's road network. Further, we find that separating both modes in the 3D-MFD-NDP solution reduces

the impact of production losses due to cross-modal interactions, especially in the congested branch.

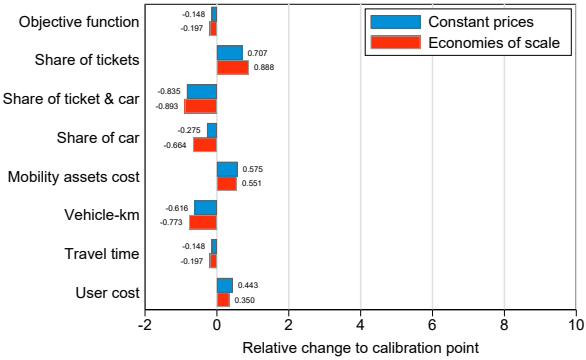
6.4 SUMMARY

The 3D-MFD-NDP provides cities a macroscopic tool to study how they can improve the performance of their multimodal transportation system by changing prices and infrastructure. Although macroscopic, this model generates fast intuitive insights into which strategies for improving mobility should be formulated. Further, the 3D-MFD-NDP is a starting point for identifying the solution space for microscopic road or bus network design problems. The results from the application to Greater Zürich emphasize that a substantial improvement in the overall system performance can only be achieved by adapting car costs, not necessarily public transport costs.

This model can be improved at various instances. For example, the model currently assumes that road infrastructure and public transport vehicles can easily be substituted, which is not the case in cities. Consequently, revising the the model with a suitable elasticity of substitution can improve the model's ability to obtain policy relevant implications. Further, the model does not address the question of optimal bus network design. Here, Daganzo's 2010 model would be a starting point to include this question in the presented 3D-MFD-NDP framework.

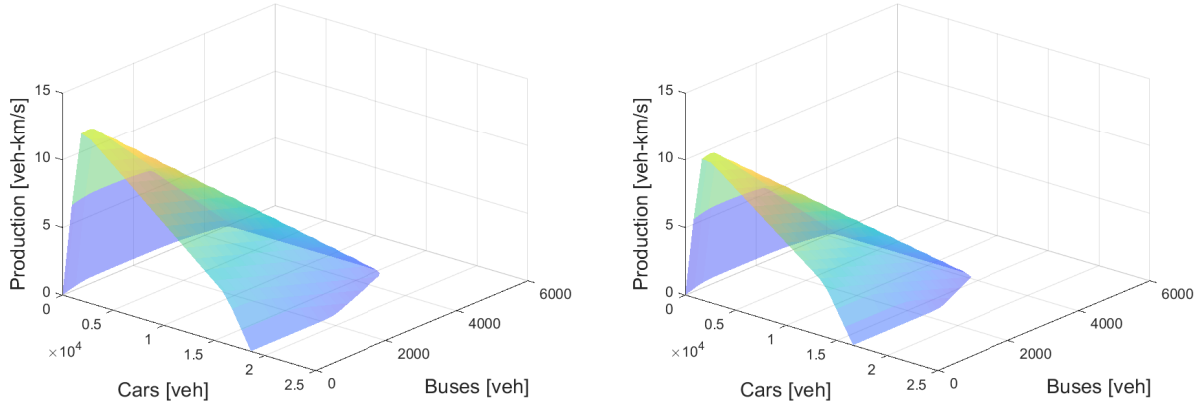


(a) Design variables.



(b) System response.

FIGURE 6.2: Solution to the 3D-MFD-NDP with all design variables as free variables. (**a**) shows the changes of the design variables relative to the calibration point. (**b**) shows the changes of the system response relative to the calibration point.



(a) Calibrated 3D-MFD.

(b) Solution 3D-MFD

FIGURE 6.3: Comparison of resulting 3D-MFDs for zone 1 of the case study. (**a**) shows the 3D-MFD for the calibration point. (**b**) shows the 3D-MFD resulting from the 3D-MFD-NDP without presence of economies of scale.

7

AGGLOMERATION AND THE 3D-MFD

> *As impressive as the role of cities in generating new innovations may be, the primary informational role of cities may not be in creating cutting edge technologies, but rather in creating learning opportunities for everyday people.*
>
> — Glaeser (1999)

The last step of this thesis is then to analyze how infrastructure investments and pricing strategies affect behavioral choices in a wider-economic context. In this context, it is important to consider that the movement of people to activities is not evenly distributed across space, but concentrated to urban centers. Such concentrations result from agglomeration (see Section 2.2.1) and dispersion forces and their understanding has implications for urban policy making and economic development as appropriate land-use policies provide opportunities to increase inhabitants' welfare (Ahlfeldt *et al.*, 2015; Allen and Arkolakis, 2019).

Recalling on the system perspective in Figure 1.1**a**, cities may wish not only to improve mobility for everyone, but also to improve the agglomeration process to realize productivity gains. However, appropriate policy making requires an understanding of the interaction of the physics of multimodal congestion and the agglomeration forces because all larger cities rely not only on cars to transport people.

Not only has agglomeration been well understood and studied (e.g., Duranton and Puga, 2004; Rosenthal and Strange, 2001, 2004; Ciccone, 2002; Melo *et al.*, 2009, 2017; Redding and Sturm, 2008; Ahlfeldt *et al.*, 2015), but also the interaction between congestion and agglomeration (e.g., Solow, 1972; Anas and Kim, 1996; Wheaton, 1998; Tabuchi, 1998; Anas and Xu, 1999; Wheaton, 2004; Arnott, 2007; Brinkman, 2016; Zhang and Kockelman, 2016; Graham, 2007b). However, as Chatman and Noland (2011) summarized "There is little discussion of public transport." So far, accounting for physical multimodal interactions in congestion or the operational features of public transport in economic models has been a rather complex and difficult task, as it usually requires extensive amounts of data and separate microscopic simulations. However, advances in research on the multi-

modal macroscopic fundamental diagram (MFD) by Daganzo (2007), Daganzo and Geroliminis (2008), Geroliminis and Daganzo (2008), Geroliminis *et al.* (2014) and in Chapter 4 of this thesis provide the first opportunity to account for physical multimodal interactions simultaneously within economic models without iterative model linking or simplistic assumptions. The MFD is an empirical framework that treats urban road networks as a system or a factory that produces trips, and where the system's internal function explicitly accounts for network topology and traffic operations, e.g., signals and buses.

This chapter provides a quantitative urban spatial equilibrium model with simple urban structure that incorporates agglomeration following Venables (2007) and the multimodal MFD framework to describe endogenously multimodal congestion and the effects of transport policies on congestion and productivity. In this model of a closed economy without external trade and urban growth, commuters choose where to live and work, their portfolio of mobility tools (car and/or bus season ticket) and their preferred mode of transport (car or bus). We use a measure of accessibility to economic mass based on Hansen (1959) as the link between changes in the transportation system and productivity. The model simultaneously solves for an economic and traffic equilibrium following exogenous changes in the prices for mobility tools or the stock of public infrastructure. Here, we illustrate the model's applicability for the greater Zürich metropolitan area, where we investigate the effects of changes to the attractiveness of the public transport system by varying ticket and car costs. We find that improvements to economic output and wages in Zürich are realized by increasing the costs for cars and not by changing public transport pricing, while improvements also result from bus frequency improvements and more dedicated bus lanes.

This chapter is organized as follows. Section 7.1 introduces the economic equilibrium model and Section 7.2.2 the illustrates the applicability of the model for the greater Zürich metropolitan area.

7.1 THE MODEL

The integrated model combines a spatially disaggregated economic equilibrium model of housing and labour markets with a transport equilibrium model of an urban network that includes mode (car and public transport for season ticket holders) and route choice.[1] In this integrated model we

1 An early iteration of this type of model is found in Schreiber *et al.* (2016).

divide the city into several zones which can be thought of as neighbourhoods, whereby we use the terms "zones" and "nodes" interchangeably. The model is calibrated to reflect reference morning and evening commute times, as we want to investigate the effects of agglomeration on wages and productivity. All travellers in the network move from their home location to their work location. We denote the home location by subscript i and the work location by subscript j. All zones that a representative traveller must pass through from i to j are denoted by subscript k. Further, as real cities usually have a substantial number of inbound commuters from the hinterland, we account for this by adding zones around the city without economic activity, i.e., without workplaces, which are strictly residential areas with a fixed population.

The economic model is composed of profit-maximizing, zone-specific aggregate production sectors all producing a homogeneous good, and utility-maximizing representative agents, denominated by where they live and work. This component of the model follows the Walrasian-Arrow-Debreu paradigm and is expressed in Harberger units. Representative agents (RA_{ij}) living in zone i provide labour to productive activities in zone j. Agents must choose where to live and work based on wages, housing prices and commuting times. Agents also choose their mobility tool portfolio t based on exogenous price levels. They further choose commuting modes and routes based on travel times that are endogenously computed in the transport equilibrium model. Following traditional urban economic models, a representative absentee landlord LL owns the capital and housing stocks in all zones and accrues rental payments from production sectors and households (Fujita and Thisse, 2013). The methodology used for representing agglomeration effects due to concentration in the workforce follows Venables (2007) with a gravity-based accessibility index. Empirical evidence for this link's validity was provided by Axhausen *et al.* (2015).

The problem is formulated as a single mixed complementarity problem (MCP) implemented in the MPSGE (Mathematical Programming System for General Equilibrium) framework in GAMS (General Algebraic Modeling System). We first introduce the key ideas of MCP in the next subsection and subsequently discuss the economic sub-model, the transport sub-model and the sorting conditions.

In Table 7.1 we list all sets used in the notation of variables and parameters. Table C.4 lists all model variables, and Table C.5 lists all model parameters. Benchmark or reference values are represented by an overline.

Index	Description
i,j,k	Zone identifier
m	Mode identifier with values *b* for bus and *c* for car
r	Route identifier
t	Mobility tool portfolio: just car, *c*; just season ticket, *s*; all tools, *a*

TABLE 7.1: Model sets

7.1.1 *Economic sub-model*

In this integrated model we consider a macroscopic perspective of a city with several separate zones for people to live and work. The model requires standard general equilibrium assumptions (firms maximize their profits and households maximize their utility) while accounting for agglomeration effects on wages and productivity.

The economic sub-model assumes perfect competition and is characterized through three types of conditions: (i) activities (households and firms must make zero profits); (ii) supply must be greater than or equal to demand; and (iii) incomes must balance with expenditures. In the following, we introduce each set of equations in that order.

7.1.1.1 *Zero-profit condition*

We denote Π_j^Y as the unit profit function for production in zone j (Y_j) and Π_{ij}^U as the unit profit function for the utility of living in i and working in j (U_{ij})[2]. Using Shepard's Lemma, input coefficients are calculated by differentiating the associated unit profit function with respect to input and output prices. Activity levels are the complementarity variables associated with zero-profit conditions. If costs exceed revenues, then the level of the activity must be zero.

In the production sector, a macro-output good is produced in each zone j according to Eq. (7.1). Unit revenue is characterized by the homogeneous price P that is set to the numeraire[3]. The unit cost function assumes Cobb-Douglas technologies, combining labour in zone j (with the wage rate W_j), capital (at the rental rate R_j) and a specific factor (with the price PSF_j)

2 Note that Π without superscript is used in Chapter 4 to denote total travel production.

3 Note that P is also used for parking capacity in Chapter 5, which interpretation is not used in this Chapter.

to account for differences in production technologies and processes across the zones.

We calculate the benchmark level of zone-specific factors as the difference between each zone's gross product and the value added from labour and capital expenditures. The exponents in the cost function in Eq. (7.1) denote the value shares for each production factor in a particular zone j.

$$-\Pi_j^Y = -P + \left(\frac{W_j}{\overline{W}_j}\right)^{\theta_j^l} \left(\frac{R_j}{\overline{R}_j}\right)^{\theta_j^k} \left(\frac{PSF_j}{\overline{PSF}_j}\right)^{\theta_j^s} \geq 0 \quad \perp \quad Y_j \geq 0 \tag{7.1}$$

In the household sector, the utility U_{ij} for the representative agent RA_{ij} living in zone i and working in zone j is composed of demand for leisure (with the price PLS_{ij}), the macro good at price P and demand for housing in zone i at price PH_i. We define utilities as given by Eq. (7.2) with a nested constant elasticity of substitution (CES) form. At the top level, demand splits between leisure and other goods combined in the composite good CG using Cobb-Douglas preferences.

$$-\Pi_{ij}^U = -PU_{ij} + \left(\frac{PLS_{ij}}{\overline{PLS}}\right)^{\theta^{ls}} CG_i^{1-\theta^{ls}} \geq 0 \quad \perp \quad U_{ij} \geq 0 \tag{7.2}$$

In the nested layer, CG_i combines the demand for the macro good and housing as given by Eq. (7.3). Preferences for the composite good follow an exogenously set elasticity of substitution, σ^c. The aggregate price of a unit of utility is represented as PU_{ij}.

$$CG_i = \left(\theta^c P^{1-\sigma^c} + (1-\theta^c)\left(\frac{PH_i}{\overline{PH}}\right)^{1-\sigma^c}\right)^{1/1-\sigma^c} \tag{7.3}$$

We assume in the benchmark that reference prices do not vary across agents choosing different residence and work zones, implying identical preferences. We capture idiosyncratic preferences by solving the model for reference prices consistent with observed sorting behaviours.

7.1.1.2 *Market-clearing conditions*

Market-clearing conditions are needed for each endogenous price in the model. First, the market-clearing condition for the homogeneous macro good is given by Eq. (7.4). As the good is not zone-specific, the total supply is the sum of all zones' supply. The total demand for the macro good is the

sum of the demand by the absentee landlord LL, who demands only the macro good, and the demand by each representative agent RA_{ij}.

$$\sum_j \frac{\partial \Pi_j^Y}{\partial P} \geq \frac{LL}{P} + \sum_{ij} \frac{\partial \Pi_{ij}^U}{\partial P} \quad \perp \quad P \geq 0 \tag{7.4}$$

Second, the market-clearing condition for the labour market as given by Eq. (7.5) includes the representation of agglomeration effects with the productivity index X_j. Thus, we scale the labour force in j by the productivity index that represents productivity gains from increased worker density in a particular location. The effective wage therefore changes for the representative agent to $W_j\ X_j$.

$$\sum_i NLW_{ij} X_j \geq \frac{\partial \Pi_j^Y}{\partial W_j} \quad \perp \quad W_j \geq 0 \tag{7.5}$$

Third, we assume that the capital stock in j, $\overline{K}_j$, is fixed to a zone. Then the market-clearing condition for capital becomes Eq. (7.6).

$$\overline{K}_j \geq \frac{\partial \Pi_j^Y}{\partial R_j} \quad \perp \quad R_j \geq 0 \tag{7.6}$$

Fourth, similar to the case of capital, we further assume that specific factors $\overline{SF}_j$ are also immobile across zones and thus fixed. Then the market-clearing condition for the specific factor becomes Eq. (7.7):

$$\overline{SF}_j \geq \frac{\partial \Pi_j^Y}{\partial PSF_j} \quad \perp \quad PSF_j \geq 0 \tag{7.7}$$

Fifth, again similar to the cases of capital and the specific factors, we assume a fixed housing stock in each zone $\overline{H}_i$. With this, the market-clearing condition for the housing market becomes Eq. (7.8):

$$\overline{H}_i \geq \sum_j \frac{\partial \Pi_{ij}^U}{\partial PH_i} \quad \perp \quad PH_i \geq 0 \tag{7.8}$$

Sixth, the market-clearing conditions require a definition of leisure. We define total leisure NLS_{ij} as given by Eq. (7.9). The total supply of leisure NLS_{ij} depends on the sorting of representative agents N_{ijmr}. We assume that all agents have a total of 3 hours in the morning and in the evening that can either be used for leisure or for commuting, with travel time T_{ijmr}

between i and j. Consequently, agents with longer commutes have less time for leisure (i.e., a lower NLS_{ij}).

$$NLS_{ij} = 2\sum_{mr} N_{ijmr}\left(3 - T_{ijmr}\right) \tag{7.9}$$

The market-clearing condition for leisure is then given by Eq. (7.10), where we implicitly assume that the value of leisure is not the same as the wage rate:

$$NLS_{ij} \geq \frac{\partial \Pi_{ij}^{U}}{\partial PLS_{ij}} \quad \perp \quad PLS_{ij} \geq 0 \tag{7.10}$$

Last, the market-clearing condition for utility is given by Eq. (7.11):

$$\frac{\partial \Pi_{ij}^{U}}{\partial PU_{ij}} \geq \frac{RA_{ij}}{PU_{ij}} \quad \perp \quad PU_{ij} \geq 0 \tag{7.11}$$

7.1.1.3 *Income balance*

For representative agents RA_{ij}, income consists of the wage income net of productivity gains and the value of their time. The expenditures of representative agents for mobility tools are the total flow of agents along m and r multiplied by the average price between origin and destination for a single traveller obtained with price Π_{ijt}^{total} for mobility tool portfolio t and its share of ownership Q_{ijt}. The income balance for representative agents thus becomes Eq. (7.12).

$$RA_{ij} = W_j\ NLW_{ij}\ X_j + PLS_{ij}\ NLS_{ij} - \left(\sum_{mr} N_{ijmr}\right)\left(\sum_{t} \Pi_{ijt}^{\text{total}}\ Q_{ijt}\right) \tag{7.12}$$

We define Q_{ijt} and Π_{ijt}^{total} in Eq. (5.15) and Eq. (5.16) respectively. This analysis does not differentiate between different price or cost components and subsume all taxes, fares etc. under the term costs.

Second, the absentee landlord, LL, is assumed to own all the capital, housing and specific factor stocks. Thus, the income balance becomes as given by Eq. (7.13). We assume that the revenue from mobility tools and mode use feeds into an external budget.

$$LL = \sum_{j}\left(R_j\overline{K}_j + PSF_j\overline{SF}_j\right) + \sum_{i} PH_i\overline{H}_i \tag{7.13}$$

7.1.2 *Transportation sub-model*

The transportation sub-model follows the 3D-MFD network assignment as introduced in Chapter 5. Here, the analyst can decide whether to include the mobility tool ownership model or simply fix ownership levels at the benchmark values $\overline{Q}_{ijt}$. The ownership model can also be replaced by adding the prices and fare costs to the actual path costs C_{ijmr}, but in this case, the model would not capture anymore the possibility that some commuters have a choice between car and public transport, while some have not.

7.1.3 *Agglomeration effects*

Agglomeration benefits are allowed to accrue from concentrating the workforce. Productivity in zone j increases if nearby areas employ more people. Put differently, productivity is a decreasing function of distance between zones due to a decrease in the ability to engage in knowledge spillovers, labour market pooling, etc. We capture the agglomeration effects following Venables (2007) by using an agglomeration index X_j for each node j, which increases as the employment density and accessibility to other workplaces ACC_j around zone j increase. Though we assume perfect competition in production, X_j represents external economies of scale.

We define the agglomeration index X_j as given by Eq. (7.14). In the benchmark equilibrium, X_j equals unity. The reference level of employment density in zone j is denoted by $\overline{ACC}_j$. The parameters μ^A and β quantify the impact of accessibility to employment (i.e., economic mass) on productivity. Figuratively speaking, μ^A determines the slope of the relationship while β simply scales changes in economic mass to productivity.

$$X_j = 1 + \beta\left(ACC_j^{\mu^A} - \overline{ACC}_j^{\mu^A}\right) \tag{7.14}$$

We define the employment density and accessibility to workplaces around j as ACC_j according to Eq. (7.15).

$$ACC_j = \sum_i \frac{LD_i}{M_{ij}} \tag{7.15}$$

Eq. (7.15) provides a measure of economic mass in zone j with effective labour demand LD_j that combines the productivity gains represented by X_j multiplied by the total labour force at j, which is given by Eq. (7.16).

$$LD_j = \sum_i N_{ij} X_j \tag{7.16}$$

7.1.4 *Sorting*

The representative agents sort across residential and work locations following a two-step, logit-based assignment. In our two-step logit formulation we first have to compute a utility index for living at node i as given by Eq. (7.17). The constraint requires that if the utility from living in i and working in j, denoted as U_{ij}, changes relative to its benchmark $\overline{u}_{ij}$, then UL_i can fluctuate to enforce that probability's sum to one. Here θ^u_{ij} describes the share of all residents living in zone i and working in j. For example, if the utility levels across all people living in i but working elsewhere in the city increase, then the utility index also increases. However, if the utility increases for some work locations but decreases for others, the relative difference determines how the utility index changes. Further, μ^L reflects an elasticity of housing location choice.

$$\sum_j \theta^u_{ij} \exp\left(\frac{1}{\mu^L}\left(\frac{U_{ij}}{\overline{u}_{ij} U_i^L} - 1\right)\right) = 1 \tag{7.17}$$

The initial number of commuters choosing to live in i follows from Eq. (7.18). Changes in utility as a result of system changes scale the benchmark number of people living in a particular zone $\overline{NL}_i$.

$$NL_i = \overline{NL}_i \left(\frac{\exp\left((U_i^L - 1)/\mu^L\right)}{\sum_j \theta^{li}_j \exp\left((U_j^L - 1)/\mu^L\right)}\right) \tag{7.18}$$

The number of travellers living at i and working at j, denoted as NLW_{ij}, follows from Eq. (7.19). Here $\theta^u_{ij} NL_i$ represents the total number of residents of i working in zone j. Changes in the utility from living in i and working in j and in the living utility index UL_i scale the benchmark total following a working locational choice elasticity μ^W.

$$NLW_{ij} = NL_i \; \theta^u_{ij} \left(\frac{\exp\left(\frac{1}{\mu^W}\left(\frac{U_{ij}}{\overline{u}_{ij} U_i^L} - 1\right)\right)}{\sum_{j'} \theta^u_{ij'} \exp\left(\frac{1}{\mu^W}\left(\frac{U_{ij'}}{\overline{u}_{ij'} U_i^L} - 1\right)\right)}\right) \tag{7.19}$$

7.2 APPLICATION TO GREATER ZÜRICH

We now apply the integrated model to the Greater Zürich. Again we want to understand the mechanisms of the model by studying the effects on the economic sorting and economic output resulting from pricing and investment schemes. In particular, we study the influence of (i) public transport service frequency, (ii) dedicated public transport infrastructure, (iii) season ticket pricing, and (iv) variable car costs. We discuss the calibration of the integrated model in Section 7.2.1 before presenting the results in Section 7.2.2.

7.2.1 *Calibration*

The zoning and traffic assignment calibration is identical to the calibration of the 3D-MFD-NDP from Chapter 6. Information on the calibration of the economic sub-model is provided in Appendix C. The endogenous variable NLW_{ij} in this Chapter correspond to the exogenous variable n_{ij} in Chapter 5. The benchmark value of $\overline{NL}_i$ is consequently calculated by the sum over all destinations j of n_{ij} of the benchmark origin-destination matrix.

Table 7.2 summarizes the relevant behavioral parameters of the model. Given a yearly moving rate of around 10 % (van Nieuwkoop, 2014), we assume a rather inelastic choice of the residential location and consequently set $\mu^W = 0.3$. We consider that the labor market is slightly more elastic that the choice of the residential location and thus set $\mu^W = 0.5$. For the value of the price elasticity of mobility tool ownership, μ^M, we consult as guidance the long-term fuel price elasticity, which has been reported to be in the interval $[-0.5; -0.1]$ (Erath and Axhausen, 2010). As we expect in the city a higher willingness to substitute mobility tools, we set $\mu^M = -0.5$. We set the scale parameter of the combined mode and route choice to $\mu^R = 10$ to maintain deterministic aspects in route and mode choice. Last, regarding the impact of economic mass on productivity, we follow the conservative values by Venables (2007) for μ^A and β. In our analysis, this results in an agglomeration elasticity of around 0.1, which is chosen to align with a 7/6 scaling relationship as reported by Bettencourt *et al.* (2007).

7.2.2 *Results*

We use the calibrated integrated model to study the effects of changes to the public transport system on urban productivity. We first simulate how

Symbol	Description	Value
μ^L	Elasticity of residential location choice	0.3
μ^W	Elasticity of workplace choice	0.5
μ^M	Mobility tool ownership price elasticity	-0.5
μ^R	Scale parameter for mode and route choice	10
μ^A	Impact of economic mass on productivity	0.2
β	Scaling of economic mass' impact on output	0.2

TABLE 7.2: Behavioural parameters

frequency changes to public transport (measured as headway) affect output (section 7.2.2.1). Second, we quantify the effects on agglomeration of reallocating dedicated space to or away from buses (section 7.2.2.2). Finally, we study how the agglomeration process is influenced by season ticket pricing (section 7.2.2.3) and by variable car costs (section 7.2.2.4). To test the sensitivity of the model with respect to the impact of agglomeration and to see whether agglomeration is important to consider, we simulate the models with the agglomeration impact parameter $\beta \in \{0; 0.2; 0.4\}$.

Recall that changes in mobility tool ownership Q_{ijt} only result from exogenous price changes, as discussed in sections 7.2.2.3 and 7.2.2.4. However, they do not result from or respond to changes in network topology, as discussed in sections 7.2.2.1 and 7.2.2.2.

7.2.2.1 *Public transport service frequency*

Our first scenario describes the outcomes of changing headway, or the time between arriving public transport vehicles. We model this shock by altering h_i on all zones i which imposes additional (or lessened) time costs for public transport. The prices of the mobility tools remain unchanged and therefore, given our construction of the model, the shares of mobility tool ownership Q_{ijt} also remain unchanged. However, because a certain share of the population has access to both mobility tools, changing the headway influences their mode choices.

In Figure 7.1 we summarize the macroscopic effects observed from varying the headway from its benchmark value 0.125h in the interval from 0.03 to 0.15. Figures 7.1a-b show the changes for city totals, and Figure 7.1c shows the changes for the entire city averages. Figures 7.1d-f provide the

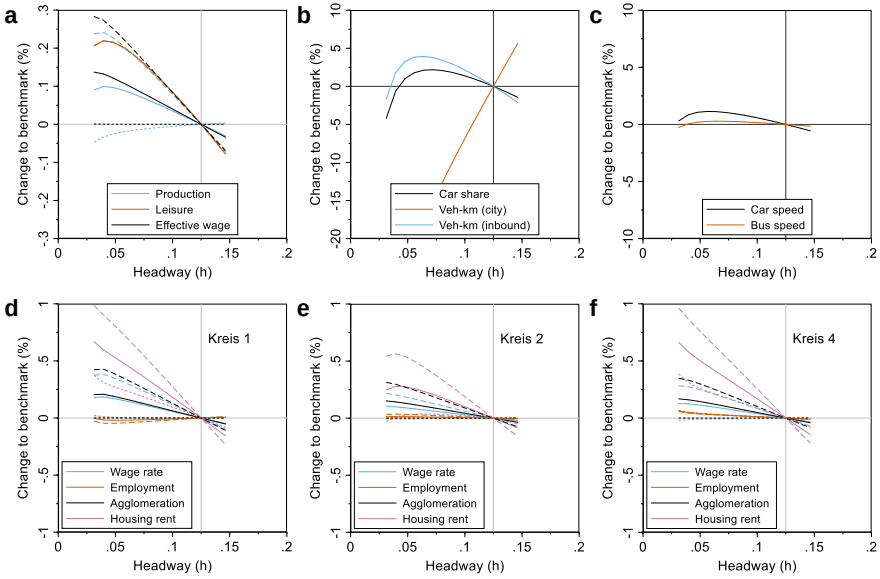


FIGURE 7.1: Transportation system and economic market responses to changes in public transport service frequency. In panels a and d to f, the solid line corresponds to $\beta = 0.2$, while the short dashed line corresponds to $\beta = 0$ and the long dashed line corresponds to $\beta = 0.4$.

results for zones (Kreise) 1,2, and 4. Figures 7.1a and 7.1d-f also provide the results for different impact levels of agglomeration as a sensitivity analysis, with a short dashed line corresponding to $\beta = 0$, a solid line corresponding to $\beta = 0.2$ and a long dashed line corresponding to $\beta = 0.4$. The figures in the following subsections are identically structured.

In Figure 7.1a we see that decreasing the headway increases total output and leisure and raises the effective wage for $\beta > 0$. In the case of $\beta = 0$, the effective wage changes only marginally due to changes in the labour markets. Better access by bus results in more commuters choosing workplaces in the central city zones, as wages there are higher. Increasing the supply of workers decreases wages, consequently resulting in a decrease of output (see Eq. (7.5)).[4] Contrary to intuition, we find in Figure 7.1b that the modal split increases with decreasing headway, but this behaviour can be

4 Given our parameterization, increasing headway beyond 0.15h leads to a breakdown of the transportation system, as the existing infrastructure is not able to accommodate all potential passengers.

explained by differentiating the total vehicle kilometres by zone. City residents drive to work less with a decrease in headway, leaving more space on the streets which is then filled by inbound commuters who had previously used public transport. Because the inbound commuters' trip distances are longer, the modal split increases. Figure 7.1c shows that improving the bus system's headway improves the journey speeds of both modes, despite the effects observed in Figure 7.1b. Comparing spatial effects at the zonal level in Figures 7.1d-f shows that increasing bus frequencies shifts employment from Kreis 1 to the other Kreise. Increased accessibility to other zones increases the productivity of the workforce and results in higher wage levels. Housing prices increase across zones given higher incomes and demands.

7.2.2.2 *Dedicated public transport infrastructure*

Our second simulation scenario proxies for changes to dedicated public transport infrastructure by altering the number of bus lanes available to commuters. Because we do not vary the mobility tool prices there will be no change in Q_{ijt}. However, as in section 7.2.2.1, mode choice and location choice effects can still be observed for the portion of the population with access to both modes.

In Figure 7.2 we summarize the macroscopic effects observed when increasing or decreasing the length of bus infrastructure by -50 to 50 % compared to the benchmark. We see in Figures 7.2a-c that more dedicated bus lanes increase economic output and leisure in the system and raise the effective wages in the case of $\beta > 0$, while changes in the total wages again remain small in the case of $\beta = 0$. Increasing the number of bus lanes increases bus speeds throughout the system, leading to more time spent for leisure and increased productivity through improved ease of access. These gains are obtained by shifting demand towards bus transportation, that is, by improving the journey speeds of both modes. In Figures 7.2d-f we report represenative spatial impacts. Reducing the number of bus lanes reduces access to the city centre and reduces the attractiveness of working there. Because the elasticity of residential location choice (μ^L) is relatively smaller than work choice elasticity, price impacts are more pronounced for housing than for wages. However, the direction of impacts at the zonal level is heterogeneous. For instance, employment increases with decreasing bus lanes in the first and fourth zones, but decreases in zone 2. In the former case, more people are willing to supply labour to these zones given longer commute times, which results in lower wages. Wage changes are particularly pronounced with larger levels of β. While employment does

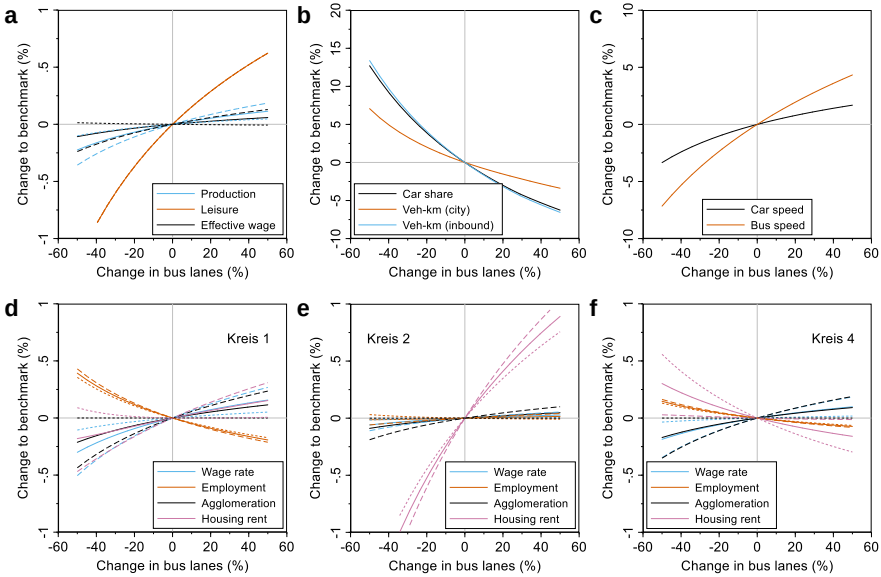


FIGURE 7.2: Transportation system and economic market responses to changes in the shares of dedicated public transport infrastructure. In panels a and d to f, the solid line corresponds to $\beta = 0.2$, while the short dashed line corresponds to $\beta = 0$ and the long dashed line corresponds to $\beta = 0.4$.

increase, our measure of accessibility decreases in the formulation of the productivity index X_j given increases in the perceived cost of travel between nodes. In contrast to Kreis 1, Kreis 4 shows increases in housing rent as a result of fewer bus lanes, indicating that more people are interested in moving there given the close proximity to zones with high wages. The converse is reported for Kreis 2 (a largely residential zone stretching down the coast of the lake), given lessened demand for living further away from high-wage zones.

7.2.2.3 *Public transport season ticket pricing*

In addition to changes in infrastructure, changes in transportation pricing can influence commuting behaviour and economic sorting. In our third scenario, we vary the season ticket price from 0-10 CHF/day (around its benchmark value of 5 CHF/day). The exogenous price change leads to

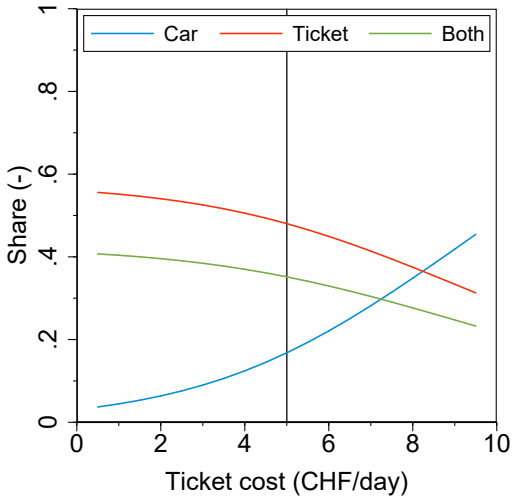


FIGURE 7.3: Shares of mobility tool ownership as a function of season ticket costs

changes in the shares of mobility tool ownership Q_{ijt}, as shown in Figure 7.3. Lower season ticket prices cause the share of car-only commuters to decrease and shares of season ticket holders and ownership of both mobility tools increase. Price changes lead to an uptake or drop in the ownership of a second mobility tool.

In Figure 7.4 we summarize the results of this scenario, showing network-wide effects in Figures 7.4a-c and spatial effects for Kreise 1, 2 and 4 in Figures 7.4d-f. Decreasing or increasing the ticket price did not impact economic output at all, except for changing the agents' income, or the transportation system's performance. This is explained by Figure 7.3, which shows that many car drivers choose to hold a season ticket as a second mobility tool but do not drop their car. Subsequently, this system change does not force them to change their travel behaviour when making route and mode choices. Conversely, when the season ticket price is increased, economic and transport performance decrease as seen in Figure 7.4a-c, leading to an increase in car use (see Figure 7.4b) that consequently decreases both modes' speeds, as shown in Figure 7.4c. Figures 7.4d-f show that increasing ticket costs tends to lead to reductions in employment, housing rent, wages and productivity externalities. Notably, this is modelled as a closed economy, and for reasons similar to the case of reduced bus lanes, employ-

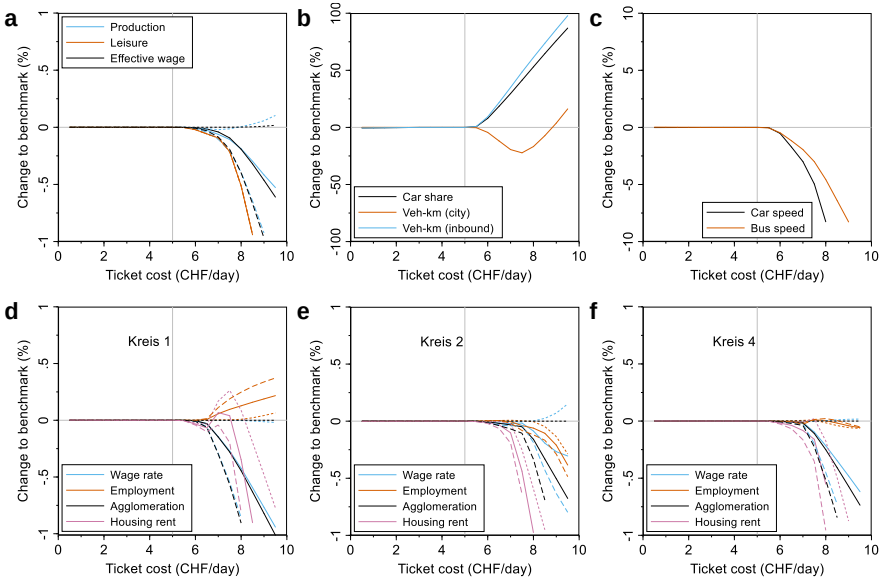


FIGURE 7.4: Transportation system and economic market responses to changes in the price of public transport season tickets. In panels a and d to f, the solid line corresponds to $\beta = 0.2$, while the short dashed line corresponds to $\beta = 0$ and the long dashed line corresponds to $\beta = 0.4$.

ment increases in Kreis 1 when ticket prices rise. We allow individuals to choose housing and employment within the defined spatial boundaries, which leads to increases in commuting distances. Kreis 1 has the largest share of employed people in Zürich in our reference equilibrium and is able to accommodate addition workers post policy change. The results also indicate a "tipping point" regarding housing rent given rising season ticket costs: Demand for housing in Kreis 1 rises until the cost of a daily ticket surpasses roughly 7.5 CHF/day, at which point individuals would rather have longer commutes than pay higher rent.

7.2.2.4 *Variable car costs*

As an alternative price mechanism, the fourth simulation analyses how commuters respond when the variable car cost (e.g., the price of fuel) is changed from almost zero to 2.50 CHF/km (our reference cost is 0.5

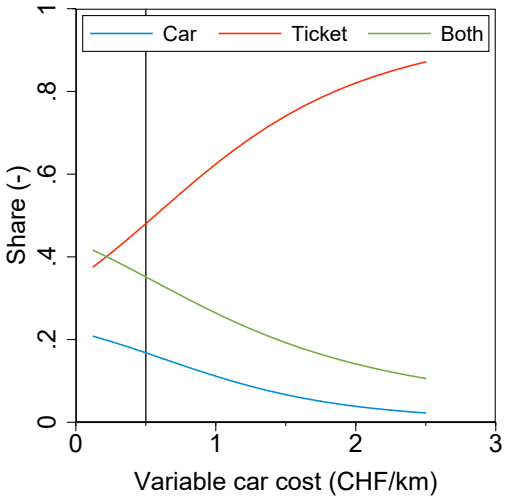


FIGURE 7.5: Shares of mobility tool ownership as a function of variable car costs.

CHF/km). Changes in the price of private vehicle use lead to changes in shares of mobility tool ownership, as shown in Figure 7.5. In contrast to the season ticket case, discussed in section 7.2.2.3, increasing the variable private car costs substantially decreases the total share of commuters who choose to own a car (either as a single mode of transport or along with season ticket ownership) and substantially increases the total share of those who own a season ticket.

Economic responses are shown in Figure 7.6. In Figure 7.6a we find that economic output and the effective wage rate increase in conjunction with variable car costs, with little variation in the interval between 0-1 CHF/km, and stronger growth thereafter. Figures 7.6b-c particulary illustrate the effects on the transportation system when variable car costs increase. When variable car costs increase beyond 1 CHF / km, inbound commuters drive less, as shown in Figure 7.6b, but their space in the network is filled by commuters living in the city, whose increase in costs is less than that of inbound commuters. Since trip distances are shorter for city commuters, the share of car trips decreases. Nevertheless, the journey speeds of both modes improve as shown in Figure 7.6c. As before, Figures 7.6d-f describe representative regional impacts. Time saved given increases in car and bus speeds reduces the cost of commuting. Kreis 1 and Kreis 4 experience per-

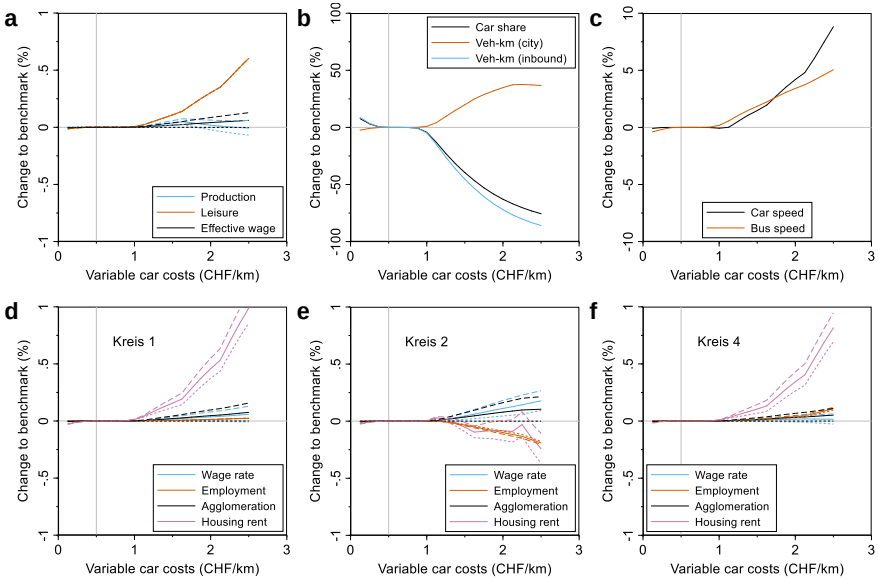


FIGURE 7.6: Transportation system and economic market responses to changes the variable car costs. In panels a and d to f, the solid line corresponds to $\beta = 0.2$, while the short dashed line corresponds to $\beta = 0$ and the long dashed line corresponds to $\beta = 0.4$.

cent increases in their overall level of employment along with wage increases given an increasing productivity externality. Notably, employment declines in Kreis 2, but the productivity index increases, implying that the decrease in the perceived cost of commuting is smaller than the decrease in labour demand. Even in the absence of the agglomeration externality here, wages increase, indicating a reduction in the supply of workers to this zone, which is magnified when accounting for the productivity index.

7.3 SUMMARY

This chapter presented an integrated computational equilibrium model that links multimodal congestion to agglomeration externalities, where the transport sub-model is based on the 3D-MFD traffic assignment introduced in Chapter 5. The link between congestion and agglomeration has been widely studied, but not based on the 3D-MFD. The model is applied to the Greater area of Zürich and to analyze how changes in bus frequencies,

dedicated bus lanes, season ticket price as well as variable car costs affect economic productivity and commuters' salary, leisure and total income.

The model highlights three key issues and conflicts of urban transportation policy making. First, improvements to economic output and individual income can only be realized by increasing car costs, not by public transport pricing. Although this does not surprise, given existing knowledge (e.g. Parry, 2009; Anas and Lindsey, 2011; de Palma and Lindsey, 2011), the simulation results in Figure 7.4 and 7.6 should clearly convince decision makers. Second, this model gives advice whether a bus network expansion suffices or an expensive subway is necessary to improve economic output and individual income, also when having a growing city in mind. For example, as shown in Figure 7.1**a**, improving the headway below 4 minutes decreases economic output, as faster modes are required to improve the access the economic mass. Third, the simulations results clearly emphasize the issue of induced demand (Goodwin, 1996; Cervero and Hansen, 2002; Mogridge, 1997, 1990), e.g., as seen in Figure 7.1**b**: Once a policy measure made road space available by a modal shift, the space is immediately occupied again, if no other measures, e.g., road space removals, are taken. Consequently, the integrated model provides a simple system perspective to analyze and understand this mechanism in macroscopic contexts.

8

DISCUSSION

> *Ich höre ja, der rechte Ring besitzt die Wunderkraft beliebt zu machen; vor Gott und Menschen angenehm.[..] Nun; wen lieben zwei von Euch am meisten?–Macht, sagt an! Ihr schweigt? Die Ringe wirken nur zurück? und nicht nach außen? Jeder liebt sich selber nur am meisten?–Oh, so seid ihr alle drei betrogene Betrüger!*
>
> — Lessing (1779)

Analyzing the complex interactions of human beings and mechanical elements of urban transportation systems in an abstract system as shown in Figure 1.1**a** means to sacrifice many details. Contrary, focusing only on the basic properties of these interactions allows to understand the mechanisms that push and pull humans in their decisions. This is necessary to uncover insights that would have been completely covered by complexity otherwise. As a consequence, I need to discuss first in this chapter the limitations of my thesis in Section 8.1 and then, second, its implications in Section 8.2.

8.1 LIMITATIONS

Throughout this theses, I reduced the complexity of cities to a simplified model. Not surprisingly then, the presented modeling approaches and results have limitations, which I address in the following.

First, the MFD and the proposed 3D-MFDs implicitly assume ordered traffic streams and structured behavior of road users. This is certainly not the case in many cities around the world. In particular, jaywalking and other traffic violations are hard to account for, but can have substantial effects on the network performance. Thus, whatever the proposed optimization problems will result for in terms of pricing and investment strategies, one has to challenge the results with respect to "*is the MFD truly describing what is going on in the network*".

Second, so far, this thesis only focused on congestion as the negative externality but there are two others that are of utmost concern in the 21st century: climate change and health. One could argue that less congestion

is leading to less pollution and consequently fewer health issues, but first there is neither a line of causality nor does this take the propulsion technology into account. Therefore, all implications and suggestions for policy implementations derived from the model should be further enhanced with our knowledge on health and climate change aspects.

Third, accessibility is a key element of transport planning with links to many other fields, e.g. economics (Vickerman *et al.*, 1999; Geurs *et al.*, 2012), where Geurs and van Wee (2004) point out to the many dimensions of accessibility. However, in this thesis, I followed an aggregated travel time perspective, which in turn ignores the microscopic and personal accessibility perspective (e.g. Le Vine *et al.*, 2013; Sarlas and Axhausen, 2019) that may become relevant as in many cities people do not only move between home and workplaces but also to many other locations. Thus, implementing the 3D-MFD network assignment inside an agent-based simulation, e.g., MATSim (Horni *et al.*, 2016), could increase the understanding of investment and pricing schemes. This idea also extends to coupling the 3D-MFD network assignment model to dedicated and microscopic road and bus network design problems in a sequential optimization. This not only allows to speed up problem solving, but also to monitor along the way the urban-scale effects.

Fourth, transport is always influenced by and is influencing politics. Here, I always consider that people respond with an expected and rational behavior, just based on their commute. This is of course simplifying as in reality people commence many more trips, inertia may dominate and people also might oppose certain pricing and investment schemes. These aspects might be very challenging to implement in macroscopic models, however, one has to bear in mind that there is more to consider for the economic activity than just the journey to work.

Last, the applications of the 3D-MFD network assignment do not consider Downs's (1962) fundamental law of congestion. A reason for this is that the current formulation of the applications model only a single ride, which completely ignores that people might be willing to travel more frequently when accessibility improves (see the discussion by Metz (2008) and Weis and Axhausen (2009)). Another reason is that I consider only a fixed population (closed city), but one can expect that more productive cities will attract more inhabitants. Another missing element is that mobility tool ownership choices are only based on price (or generalized cost), but this does not includes the findings from Chapter 3, where the land-

use effects of quality of public transport and spatial typology affect travel behavior as well.

8.2 IMPLICATIONS

The contributions of this thesis have several important implications for the academic literature (see Section 8.2.1), transport policy (see Section 8.2.2), vehicle automation (see Section 8.2.3) as well as developing countries (see Section 8.2.4).

8.2.1 *Academic literature*

This thesis has at least three implications for the academic literature. First, with the formulation of the geometric and Bose-Einstein condensate approach to the MFD, I offer researchers new frameworks to model a variety of interactions between cars and other transportation modes in a closed form. This allows new procedures to be developed in the future, e.g. for the bus network design problem, modeling of multi-layer transportation networks and other existing features in urban transportation networks.

Second, it seems that the academic literature skipped static traffic assignment models for the MFD right away and stepped immediately into the (more complex) dynamic models. I propose a static model that is not only easy to formulate, but can also be implemented with small effort into existing professional traffic assignment software tools. As this kind of software is widely used around the world for planning, further extending the proposed static assignment procedures may boost the overall acceptance of the MFD.

Third, for the economics community, the proposed 3D-MFD framework provides a throughout simple, but physically consistent approach to model multimodal traffic and congestion. As in many instances, economists refrain from using a full traffic assignment step due to its complexity, the presented 3D-MFD network assignment provides a simple tool.

8.2.2 *Transport policy*

The results of this thesis certainly emphasize that there is currently too much space and too many resources allocated to cars in our cities and that one can improve the entire system by using public transport. This is of course not surprising as it has been in the public debate ever since cars

congested streets and has been brought to academia in the 1960s by Smeed (1961, 1968). However, the introduced models in Chapters 6 and 7 follow a long-term planning perspective and allow rather intuitively fast and easy computation of policy relevant question of *how many cars are too many*.

However, the proposed 3D-MFD, the traffic assignment and optimization models allow to obtain insights into how close or how distant cities are to the optimum, given their infrastructure and demand. As high-quality public transport operations is expensive so is the provision of road infrastructure, my models allow to quantify which level of provision is optimal. Nevertheless, the complexity - which is also present in the number of variables in the models - emphasizes that there is no *general remedy* for success. This means that every city has to analyze the effects of pricing and investment schemes for itself.

Nevertheless, the contributions of this thesis show that using an objective such as collective travel time, accessibility (and its impact on productivity) or total passenger throughput can easily be analyzed with the 3D-MFD. Consequently, these measures should be considered in the decision making, and not necessarily only travel time savings as already discussed by (Metz, 2008).

8.2.3 *Vehicle automation*

During the time this thesis was drafted, vehicle automation was considered a soon-to-arrive disruptive technology that is expected to result in radical changes to our transport system (Fagnant and Kockelman, 2015; Bösch, 2018). In light of the substantial changes to the cost-structure caused by automation (Bösch *et al.*, 2017), the following question arises: *how are the contributions of this linked to this innovation*

In particular, the expected cost changes correspond to the production cost. As the two multimodal congestion mechanisms in Sections 4.4.2 and 4.4.3 are flexible to accommodate the expected technology changes, one can investigate the behavioral and system responses with 3D-MFD-NDP from Chapter 6 and the integrated equilibrium model from Chapter 7. Consequently, this thesis can advise policy makers in finding appropriate investment and pricing schemes.

Second, the emergence of new technologies and business models also changes the way one *thinks* about cars. For example, a shared autonomous vehicles does not end its vehicle trip, but only the passenger trip as well as trip length will become a random variable (e.g. demand level etc. will in-

fluence the route choice of the agency), which in turns makes the modeling more difficult. However, the Bose-Einstein based approach to the 3D-MFD is not restricted to the discussed interactions between cars, buses and bicycles, but can be expanded to model such new services as well.

8.2.4 *Developing regions*

We can argue that in our developed world, we already poured that much concrete that our (wrong) investment decisions are manifested. Consequently, we lost many degrees of freedom to implement large scale changes to the system. However, many cities around the world still have plenty of concrete to pour, dollars to spend and - hopefully - are aware of our *mistakes* of the past.

In particular, the optimization problem following Venables (2007) based on the 3D-MFD allows developing cities to analyze under which scenarios they can ignite the full power of agglomeration economies. This can help them to switch from a self-sufficient city to a trading city (Venables, 2017a,b). The 3D-MFD-NDP then will tell planners at a macroscopic scale consequently the optimal strategy between car and public transport choices.

Further, the presented optimization models also allow to see which long-term effects the construction of subways can have not only on the travel time savings, but also in terms of economic output (Loo and Cheng, 2010; Roth *et al.*, 2012). As these projects are usually rather expensive, I provide a simple and fast-to-solve framework to see whether it makes sense to build a subway at all. Then, with a sequential optimization approach, the second layer of the optimization can use the output of the macroscopic optimization as starting values for in-depth network design problem algorithms for each of the considered modes.

9

CONCLUSIONS

Ich wundere mich oft darüber, wie derselbe Mensch, der sich mehr liebt als alle anderen, dennoch mehr Gewicht auf das Urteil anderer über ihn, als auf das eigene legen kann.

— Antonius (1949)

It is easy to ask "*how many cars are too many for a city*", but it is not easy to answer that question. Not only is there no such thing like a universal way to improve traffic, but also are the many physical and political constraints in our cities an incomprehensible complex system. In concluding this thesis, the contributions provide opportunities not only to identify the optimal parameters $\mathbf{X}^*$ from Figure 1.1**b**, but also to approach the question of "*how many cars are too many*". Nevertheless, one should never ignore the underlying assumptions and the externalities and aspects not considered in this thesis.

In this last chapter, I first summarize the most essential parts of my thesis in Section 9.1, make an outlook to future research in Section 9.2 and make concluding remarks in Section 9.3.

9.1 SUMMARY

The key contributions of this thesis are fourfold. First, I propose a multimodal accessibility indicator to analyze travel behavior. Second, I present substantial empirical evidence for the existence and physical properties of the MFD that implies that urban road networks can be simplified into a system as shown in Figure 1.1**a**. Third, I present three multimodal congestion mechanisms for the MFD that capture the interactions between cars and buses. This allows us to model passenger choices and vehicle interactions at the same time in a single system perspective. Fourth, I formulate a simple traffic assignment based on the 3D-MFD and illustrate its applicability in two macroscopic optimization problems to identify optimal investment and pricing schemes.

Throughout this thesis, I used the long-existing concept of accessibility in the Hansen (1959) formulation as well as taking part in the renaissance

of macroscopic traffic modeling with the macroscopic fundamental diagram (MFD). This amalgamation of macroscopic perspectives allowed me to derive two city-scale policy advisory tools: the 3D-MFD network design problem and the integration of the 3D-MFD network assignment in Venables's 2007 agglomeration effects model. In the formulation presented, these two models allow urban and transport planners to study the pricing effects of fixed and variable car and public transport prices, including the effects of season tickets, as well as the investment effects in roads and buses, allocation of dedicated bus lanes, frequency setting of buses as well macroscopic traffic signal strategies.

All presented approaches follow a macroscopic thinking that is a reduction of complexity of urban systems to their basic principles. This perspective allows not only to focus on the overarching effects and strategies, important for first steps in long-term urban-scale policy making, but also a much simpler communication and reasoning about results and implications. Nevertheless, these macroscopic models do not replace cost-benefit appraisals that capture a wider range of local effects as well. However, they provide reasonable limits to the solution space and, consequently, help to identify the most relevant scenarios for the public and political debate.

9.2 OUTLOOK

This thesis provides many possibilities for future research that develop the presented models and ideas further. In particular,

- can the static traffic assignment be replaced with dynamic assignments, either with the accumulation or trip-based model (Mariotte *et al.*, 2017).
- the 3D-MFD considers only two vehicle classes, but many urban networks have more classes (cars, taxis, bicycles, buses, trams etc.), meaning that future research should include these many *layers* of urban road networks in the extensions of the 3D-MFD.
- I have only compared critical points of networks, but future research also needs to compare the shapes of MFDs and the determinants of congestion progression. These insights may allow to derive a global law of congestion in cities.
- the bus network design is so far captured macroscopically by α, but future research can use the 3D-MFD-NDP in a sequential optimiza-

tion problem to tap on conventional network design problems (see Section 6.1).

- last, the presented approaches also provide means to study the optimal city size in terms of agglomeration and congestion externalities (e.g. Zhang and Kockelman, 2016), and to study when and how cities should invest (not any further) in subway networks (Derrible and Kennedy, 2010, 2011, 2009).

9.3 CONCLUDING REMARKS

In essence, this thesis followed the idea by Bettencourt (2013) to reduce cities' complexity to a few basic principles. These basic principles govern the decisions of its inhabitants, but the consequences of such decisions determine the city's success. As individual mobility patterns are difficult to change unless economically incentivized, either in monetary or time costs, urban and transport policy making has to plan and manage the transportation system such that mobility, accessibility and/or productivity is optimized or at least improved in the long-term. The presented methodological advances provide policy makers macroscopic tools to study pricing and investment effects that guide policy making towards to - hopefully - decisions to improve mobility for everyone.

In closing, most cities in the Western hemisphere made many of their transport decisions permanent with heaps of concrete, such that investments to change behavior nowadays become costly, and consequently the *solution space* for improving mobility and the success of a city is limited. Contrary, as pointed out in Section 8.2.4 for the developing world, where cities are still growing and plenty of investment decisions are still years ahead, mindful and balanced choices can be a substantial contributor to the economic success of a city. Consequently, the results of this thesis should provide decision makers around the world a handful of ideas to derive strategies and conclusions for their own city.

A

EMPIRICAL TRAFFIC DATA

This appendix summarizes the collection of empirical traffic data and its processing in relationship with the estimation of MFDs (Geroliminis and Daganzo, 2008; Leclercq *et al.*, 2014). We discuss two primary sources: stationary sensors, e.g. inductive loop detectors (ILD) or ultrasonic detectors, and trajectory data, e.g. from automated vehicle location devices (AVL). See Section 4.2.2 for an overview of MFD estimation for each of the two data sources.

This appendix is organized as follows. Section A.1 introduces to the fundamentals of the data collection for each of the two mentioned data sources as well as its MFD related data processing. Then, in Section A.3 we describe how we calibrate MFD to accurately describe speeds and densities.

A.1 OBSERVING TRAFFIC

Arguably, ILDs are the most common sensor type in cities for stationary traffic observations and thus we concentrate in Section A.1.1 on their operational features. Generally, the obtained measurements are Eulerian observations - thus we refer to Section 4.2.2.2 for more details on the idea of these measurements. Contrary, AVL data that we discuss in Section A.1.2 are Lagrangian observations - thus we refer to Section 4.2.2.3 for more details on the idea of these measurements.

A.1.1 *Observations from inductive loop detectors*

ILDs are mounted into the surface of road networks as illustrated in Figure A.1 on the northern most lane. This section is primarily adopted from Bliemer (2001). Usually, they cover a single lane and are made out of coiled wires, creating an inductance, which detects vehicles passing over it. Here, we consider the two geometric features: The length of the detector in driving direction l_l and the distance of the sensor to the downstream stop line l_s. To provide more accurate speed measurements, two sensors can be mounted consecutively in driving direction as illustrated in the southernmost lane. Then, also the gap between sensors l_g needs to be considered.

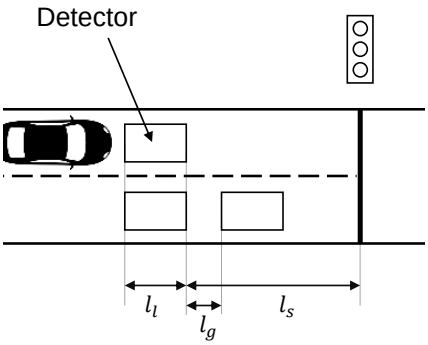


FIGURE A.1: Scheme of loop detector operations, adopted from Bliemer (2001). Car by Pablo Rozenberg from the Noun Project.

ILDs measure traffic at regular intervals, where n is the interval number of the observation period. The duration of the interval is given by the start time at t_{n-1} and end time at t_n. In interval n, a set of vehicles $J(n)$ passed over the detector and has been detected by changes in the magnetic field. Each vehicle $j \in J(n)$ *activates* the detector for duration τ_j. For convenience, we write consider the notation n and t_n to be equal. Then, the control device computes the detector flow rate $q_i(t_n)$ of detector i at time t_n by Eqn. A.1. It is simply the cardinality of set $J(t_n)$, i.e. the number of detected vehicles, divided by the duration of the interval.

$$q_i(t_n) = \frac{1}{t_n - t_{n-1}} \sum_{j \in J(t_n)} I(\tau_j > 0) \tag{A.1}$$

The second measure single ILDs provide is the detector occupancy $o_i(t_n)$. Occupancy is an approximation of traffic density as both can be considered to be proportional and positively related (Bickel *et al.*, 2007). In the above mentioned setting, detector occupancy $o_i(t_n)$ as calculated as given by Eqn. A.2 as the total time vehicles of set $J(t_n)$ spent on detector i divided by the interval length.

$$o_i(t_n) = \frac{\sum_{j \in J_{t_n}} I(\tau_j > 0)}{t_n - t_{n-1}} \tag{A.2}$$

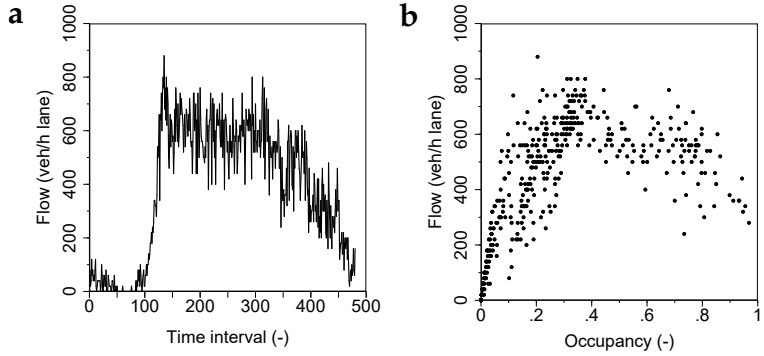


FIGURE A.2: Raw data of a single loop detector in Zürich. **(a)** shows the time-series of flow over the course of a single day, **(b)** shows the scatter plot of the same day of flow versus detector occupancy, that is an approximation of vehicle density.

The usual report of a detector is similar to the output presented below. With a time-stamp denoting the beginning or end of the observation interval t_n, the detector number, the flow rate and the occupancy. Some system may perform a plausibility check on the measured values and report an error if implausible.

timestamp	detector	flow	occ	error
702810000000	1337	34	0.31	0
702810300000	1337	49	0.42	0
702810900000	1337	412	0.9	1
702811200000	1337	42	0.37	0

In Figure A.2 we show for a single selected detector the time-series of flow and occupancy (in Figure A.2**a**) and the scatter plot of flow versus occupancy (in Figure A.2**b**). In the time-series figure we can clearly see the loading and unloading of the network during the morning and evening and peak, while we see in the scatter plot that the observations create the silhouette of a fundamental diagram of a road.

Following the fundamental equation of traffic (as given in Eqn. 4.10), the speed follows from flow divided by density. Thus, for a single ILD, its speed can be approximated by Eqn. A.3 where s_i is a detector specific con-

version scalar. In case double loops are installed as given in the southernmost lane of Figure A.1, the speed can be readily measured as the distance between loops and the time gap between two loops is known.

$$v_i(t_n) = \frac{q_i(t_n)}{o_i(t_n)\ s_i} \tag{A.3}$$

For this, Bickel *et al.* (2007) state that a reasonable estimate of s_i is the effective vehicle length which is equal to the length of the vehicle plus the size of the detector. As vehicle lengths usually differ, they can be considered a random variable with a common mean which should be used as s_i. Importantly, ILDs provide an estimate of the time-mean speed as shown in Figure 4.4 on a link, but not the space-mean speed for which we aim for in the MFD analysis. However, if the single time-mean speed measurements ϑ_j are available, we can obtain the space-mean speed v_i^{space} using the harmonic mean as given by Eqn. A.4. Unfortunately, single speed measurements are rarely reported and stored by detectors, but only aggregated values.

$$v_i^{space} = \frac{1}{\sum_{j \in J(t_n)} \frac{1}{\vartheta_j}} \tag{A.4}$$

A.1.2 *Observations from automated vehicle location devices*

The AVL devices are typically used by public transport agencies to track their vehicles to provide passenger information. As we further use this data source in our analysis only for public transport vehicles, we discuss AVL data from the public transport perspective.

The AVL for public transport vehicles usually record time-stamps at the public transport stops which locations are known as well as the approximate driving distance. The recorded data set usually is structured as listed below.

busid	seq	stopid	arrival	departure
4	1	1482	657985020000	657985045000
4	2	0832	657985143000	657985181000
4	3	0420	657985250000	657985265000
4	4	2013	657985402000	657985429000
4	5	3988	657985542000	657985560000

For each bus during one driving sequence, the actual arrival and departure times are recorded along the location identifier of the stop. Along with the distance between stop information, we can calculate the space-mean speed of buses in the network including the dwelling behavior by using Eqn. 4.11.

In case of public transport vehicles, the data sets contain the measurements of all vehicles so that the full production information is available. However, this kind of data set can also be available from car, e.g. from taxis, where instead of the stopid the coordinates are recorded at regular time intervals. Then, the shortest path distance between two observation points must be calculated to obtain the driven distance in order to calculate the space-mean speed. Further, as usually not all cars would be equipped with AVLs, this data source must be fusioned with detector data to obtain full flow and density information (Ambühl and Menendez, 2016).

A.2 DATA PROCESSING

In this thesis, all empirical MFDs are estimated from stationary traffic sensors. The available raw data was processed in two data bases: The first data base stores the actual measurements of the detectors as provided in the listing in Section A.1.1. The second data base contains the static information of the location of the detectors in a geospatial vector data format (including attributes such as road name, functional road class, distance to traffic signal, length of the monitored lane). Both data sets are linked with a unique detector identifier. The geospatial vector data base allows us to select detectors not only by certain attributes, e.g. only main road level detectors, but also by defining regions of interest with an arbitrarily shaped polygon. Furthermore, it allows us to consider only one detector per lane in case there are several detectors located within a lane. The attribute functional road class of each detector is obtained by mapping the detector to

the nearest OpenStreetMap link and acquiring the functional road class attribute of that link.

Despite the fact that many cities already operate an error detection system, there are still many issues and caveats in measuring traffic with stationary sensors (Bickel *et al.*, 2007). Thus, we further scrutinize the raw data to remove measurements containing errors (e.g. flow but no occupancy, or constant values) and outliers, and we reduce the noise with a moving average technique. Further, as most cities do not have a unbiased loop detector distribution (Leclercq *et al.*, 2014), we finally estimate the MFDs with a correction method (Ambühl *et al.*, 2017).

A.3 MFD CALIBRATION

As mentioned in Section A.1.1, we require information of s_i in order to measure traffic density or speed instead of an approximation with occupancy. However, such information is mostly unavailable and rather unreliable. Therefore, we obtain a conversion scalar from the following two-step calibration procedure.

In the first step, we make single detectors' occupancy measurements robust against variations in detector length within a city. We translate occupancy into an approximation of vehicle density depending on the location of the detector. To do so, we assume that the free-flow speed decreases linearly towards the downstream signal. Here, we find that MFDs do not change substantially with changes in the parameter of this first calibration step, but they do when this step is omitted.

In the second step, we calibrate the estimated MFD by matching the average speed from the fastest observed hour in the MFD to the average speed queried from the Google Directions API for that same hour. In particular, we query 1000 random origin-destination pairs and request the travel time and distance. Note that the trips linking those origins and destinations need to be at least 1km long. From this data set we consider only trips that are entirely in the region and are not longer than 6 km. As we are interested in urban roads only, we exclude freeways in our API request.

For the calibration, we then investigate the distributions of both speed distributions with the mean, the 85^{th}, and the 95^{th} percentile. The calibration scalar s is then obtained by dividing the mean of the Google speed by the mean of the MFD speed. Then, the MFD occupancy is converted to density using that scalar. As a sensitivity analysis of the results and as a weight for comparative analyses, we further define a lower value of the

calibration scalar s_{low} by dividing the 85^{th} percentile of the Google speed by the 95^{th} percentile of the MFD speed, and an upper value of the calibration scalar s_{high} by dividing the 95^{th} percentile of the Google speed by the 85^{th} percentile of the MFD speed.

B

DATA SET OVERVIEW

This Appendix summarizes the information for the traffic data sets used throughout this thesis. In detail, we present in Section B.1 the list of data sets for which bus and car data was available to us. In Section B.2, the measurement of bicycle interactions in Amsterdam and London is presented. Section B.3 presents further data tables for the MFD models.

B.1 3D-MFD DATA

B.1.1 *London*

Transport for London (TfL) organizes mobility throughout the city for almost all modes of transport except swimming and flying. In particular, they care about cars, buses, the underground, cycling, pedestrians, cable cars and river boats.

TfL provides detailed records for car and bus system operations. For car transport, TfL located on almost all major streets in London *SCOOT* (Split Cycle Offset Optimisation Technique) detectors, i.e. ILDs. By system design, these detectors are mostly located far upstream of an intersection, in other words right at the beginning of a link, but due to London's rather historic road networks these links can be considered comparatively short. Figure B.1 shows the area of the city for which we received detector data. The region covers an area that extends way beyond the congestion charging area that we highlighted. We use this region in the empirical MFD analysis in Chapter 4.

TfL operates the famous red double-decker buses on many bus routes, in total, there are more than 8000 buses in the city. Each bus is equipped with an AVL, here named *iBus* system. This system records data as listed in Section A.1.2 and is used to measure the performance of the system and in particular to inform passengers about the next departures. Im the available data set we use all regular bus lines and exclude in particular the river boats from the measurements. We also add to Figure B.1 the bus lines in the city for which we have data available.

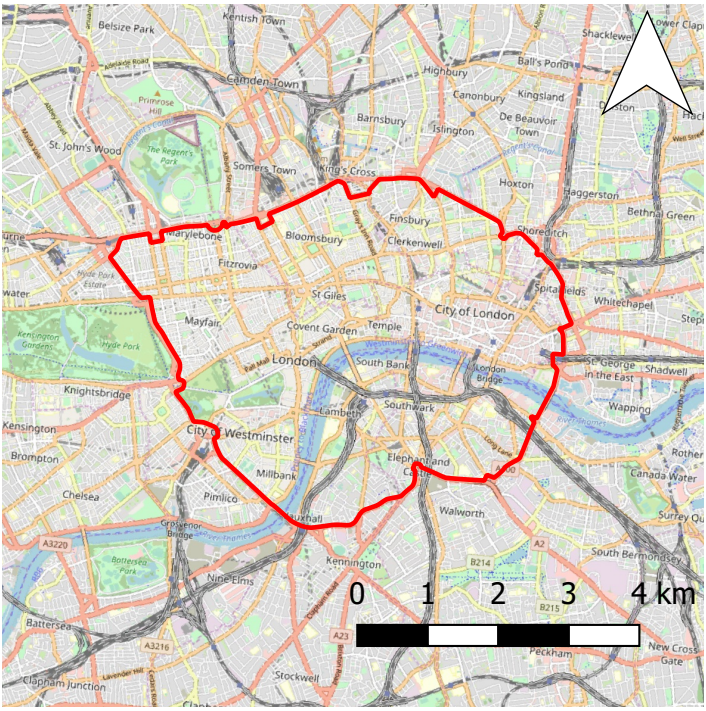


FIGURE B.1: Zoning for 3D-MFD analysis in London. The background map is provided by OpenStreetMap.

In total, we have three weeks of data available of both systems: one week from May 2015, one week from September 2015 and one week from May 2016.

B.1.2 *Zürich*

The traffic management system of Zürich operates an extensive detector system with more than 4500 detectors at 384 intersections for the control of public transport, cars or both (Stadt Zürich, 2015). Their purpose is to give priority to public transport, support traffic signal control, and identify congestion. For traffic measurements, the city operates more than 1000 ILDs (see Appendix A for details).

The geometry of ILDs does not vary much (around 90 % of the loops have a length of 1.5 m and 10 % have a length of 3.5 m). The aggregation

period is three minutes. The city determined that the errors of the detectors were lower than 5 % (Stadt Zürich, 2017). Figure B.2 shows the city of Zürich with the location of the detectors as well as the regions where we analyze MFDs. Importantly, in both highlighted regions, the number of available routes is very limited and car drivers can rarely change their route to adapt to a traffic situation. Thus, we do not expect to observe large differences in the loading and unloading of the network and thus a more or less heterogeneously congested network (Gayah and Daganzo, 2011).

The public transport operator in Zürich, *Verkehrsbetriebe Zürich* (VBZ), have a similar system as TfL's iBus system (Stadt Zürich, 2016). Thus, the available AVL data is the same format as presented in Section A.1.2.

The data used for the analysis in this thesis spans a one week period in October 2015. Here, we focus on two regions within Zürich, each with an area of approximately 2 km^2. Figure B.2 shows both regions shaded in gray. We denote the zone in the west as Wiedikon and the zone in the east as City center. The zones are selected for a couple of reasons. First, the regions differ in their share of dedicated lanes for public transport; in the City center almost 75 % of the lane-km used by public transport are dedicated, whereas in Wiedikon this number is 60 %. Second, we have chosen small areas in order not to violate the homogeneity assumption of the MFD. The size of the zones is about one-fourth of the one analyzed in Yokohama by Geroliminis and Daganzo (2008), and half as large or even magnitudes smaller than the regions which resulted from static and dynamic partitioning in Shenzhen (Ji *et al.*, 2014). Third, our selected regions have a homogeneous topology and similar hierarchy in the road network, e.g. no presence of freeways. Thus, we avoid further partitioning of the network.

B.1.3 *Simulation*

The simulated abstract network was developed using a VISSIM microsimulation platform, as a 10×10 grid, with 180 links and the average block length of 150 m. Each signalized intersection was modeled with a cycle length of 60 sec, and 27 sec of green (plus 3 sec of lost time) for all conflicting signal phases. The tested traffic scenario has 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.

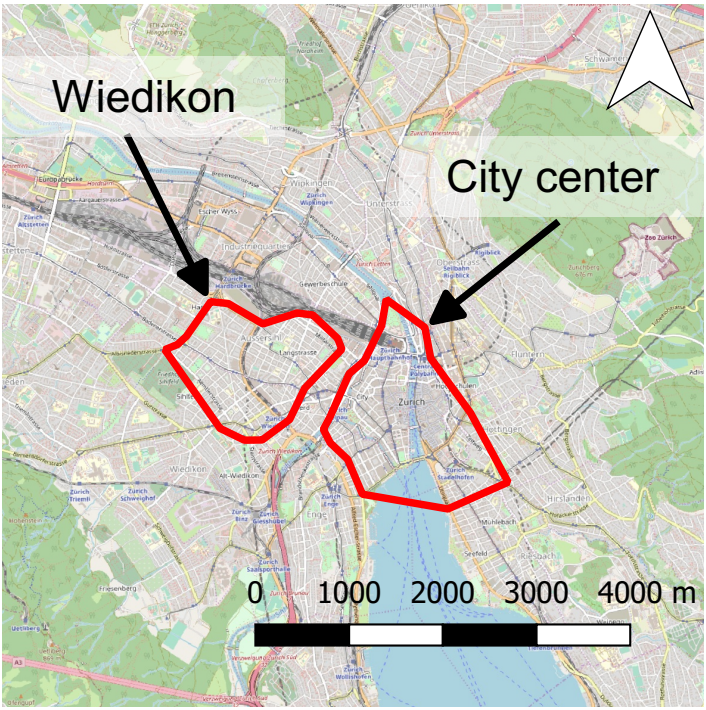


FIGURE B.2: Zones of analysis of the empirical 3D-MFD. The zone on the east is *City center* and the zone in the west is *Wiedikon*. The background map is provided by OpenStreetMap.

As we want to investigate the full shape of the 3D-MFD with this data set, we did simulation runs with a variety of headways.

B.2 MEASURING BICYCLE INTERACTIONS

The calibration of the tri-modal MFD as introduced in Section 4.4.3 to a specific context requires data. While speed measurements of buses and cars are widely available to calibrate the unimodal MFDs for cars and buses as well as the continuous multiclass FD (e.g., Loder *et al.*, 2017, 2019), no data set exists that allows to calibrate the bicycle interaction models. Therefore, we collected suitable video data in Amsterdam (NL) and London (UK). Figure B.3 shows the experimental sites.

For the calibration of the (bi-modal) discrete fundamental diagram, we measured travel times along Sarphatistraat in Amsterdam (see Figure B.3**a**) and Stamford Street in London (see Figure B.3**c**). We had to rely on two different experimental sites because the Amsterdam site did not have enough cars to identify an interaction effect of bicycles on cars, while the London site did not have enough bicycles to identify an interaction effect of cars on bicycles. As we calibrate with the data the proposed delay models and do not do MFD curve fitting, having separate data sources does not lead to issues of combining both data sources. The Sarphatistraat is designed such that cyclists have priority but cars can ride along. Travel time data for cars and bicycles was collected in the westbound direction between Alexanderplein and Weesperplein as shown in B.3**b**. To account for vehicle entries and exits between both intersections, we use only the flows and travel times between locations II and III as shown in Figure B.3**b**. Data has been collected in the morning peak between 8-9am on the 5th of June, 2018. While the total flows at each location were recorded, travel times were only obtained for some vehicles. Along Stamford Street in London, bicycles, buses and cars share the same lane. Flow and travel time data was collected in both directions in the first week of June 2019. Again, we account for vehicle entries and exits. At the marked locations 1 to 5 in London in Figure B.3**c**, we measured as the additional delays to cars and buses the duration until bicycles cleared the space in front of cars and buses so that they can pass the intersection. Figures B.3**d-e** show the bicycle stop boxes at the five locations for measuring the additional delays.

The results of our measurements and the implications on delays are shown in Figures B.1-B.5. In particular, Figure B.1 shows the results for Sarphatistraat, i.e., the effects of car presence on bicycle flows. Figure B.1**a**

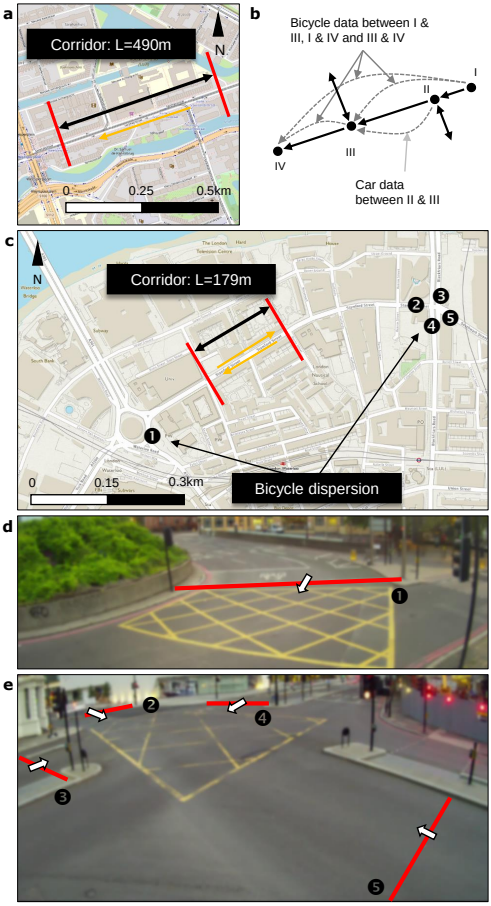


FIGURE B.3: Experimental sites for the measurements of bicycle interactions. (**a**) Travel time corridor along Sarphatistraat in Amsterdam. Background map courtesy of `https://www.openstreetmap.org/` (**b**) Locations between which travel times of cars and bicycles were recorded. There are several locations to account for in and outflow from site streets. The measurements used for the calibration are between locations II and III. (**c**) Measurement locations for the calibration of the bicycle interaction models along Stamford Street in London. Background map courtesy of `https://osmaps.ordnancesurvey.co.uk/`. (**d**) Situation plan for bicycle stopping box number 1 of this analysis for bicycle platoon dispersion. (**e**) Situation plan for bicycle stopping boxes 2-5 of this analysis for bicycle platoon dispersion.

shows the corridor's cumulative inflows and outflows of bicycles. The links between both curves mark those bicycles for which travel times have been recorded. Arguably, travel times are almost uniformly collected during the observation period. Figure B.1**b** shows the intuitive unimodal speed-density relationships for cars and bicycles. The negative marginal effect of density can be clearly seen for both modes. With the available data, we want to see from which car density threshold onwards bicycle speeds are significantly reduced or the pace is significantly increased. We do so by estimating bicycle pace as a function of bicycle density and a binary variable for the car density threshold. We increase the latter incrementally until its effect becomes statically significant. We find that up from $k_c = 0.01$ (cars/m), cyclists' pace increases by around 12 % as shown in Figure B.1**c**, i.e., $\delta = 0.88$.

For the Stamford Street, we show the measurements and delay effects in Figure B.4. In total, we recorded car and bicycle travel times during three different time slots to cover multiple different traffic conditions. Figures B.4**a-f** show the recorded travel times with grey trajectories being cars and red trajectories being bicycles. The different traffic conditions and the effect of the traffic signal in the eastbound direction can be clearly seen as well as cars overtaking bicycles in free flow and vice versa in congestion. In Figure B.4**g** we show each direction's car pace as a function of car density where we find congestion in the eastbound direction, but not in the westbound direction. We then estimate the bicycle interaction effects for two different cases. First, we calculated the average additional delay caused by overtaking 1,2 or 3 bicycles along the corridor. Overtaking means here that the trajectories of a car and bicycle cross. Figure B.4**h** shows the intuitive results as with more overtaken bicycles, cars experience more interaction delays. Second, we calculate for each car the bicycle density in front of the vehicle during its trip along the corridor and compute the average additional interaction delay as a function of bicycle density as shown in Figure B.4**i**. Again, we find the intuitive relationship that a higher bicycle density increases the additional interaction delay for cars. We find that the additional delay in free flow leads to $v_{red} = 4\,\mathrm{m/s}$ for both modes.

The bicycle platoon dispersion measurements at all five sites are shown in Figure B.5. We find the expected relationship that more bicycles in stopping boxes require more time to empty the space for cars. In our data, we find an average additional interaction delay of 0.534 seconds (with standard error 0.022) per bicycle for cars. We tested for an influence of the experimental site in Figure B.5's relationship and found statistically signif-

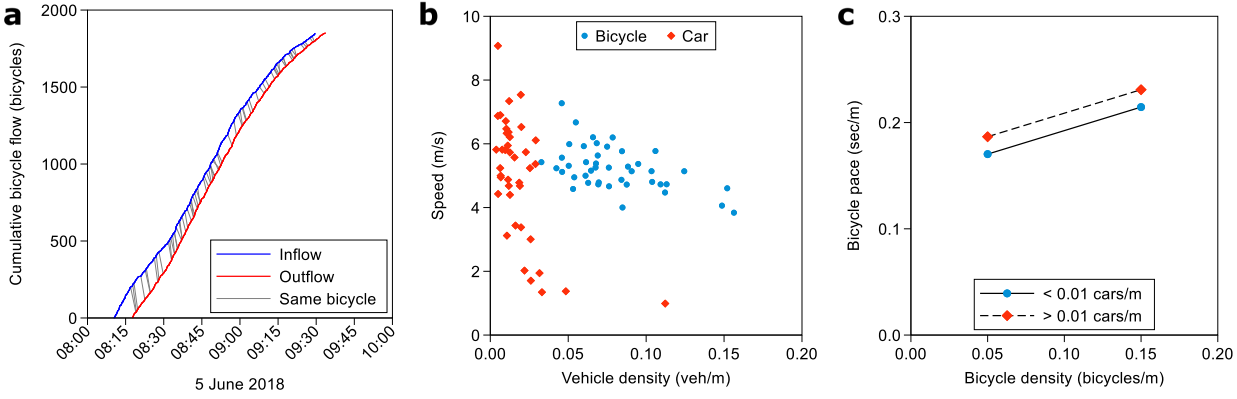


TABLE B.1: Bi-modal measurements from Sarphatistraat in Amsterdam. (**a**) presents the cumulative inflow and outflow curves of bicycles and marks those observations from which the travel time information is drawn. (**b**) shows the speed density relationship for bicycles and cars, aggregated to 60 seconds intervals. (**c**) displays the effect of car presence on bicycle pace. This relationship follows from a linear model of bicycle pace as a function of bicycle density and the car density indicator. The lines (**c**) are the predicted effects of that model.

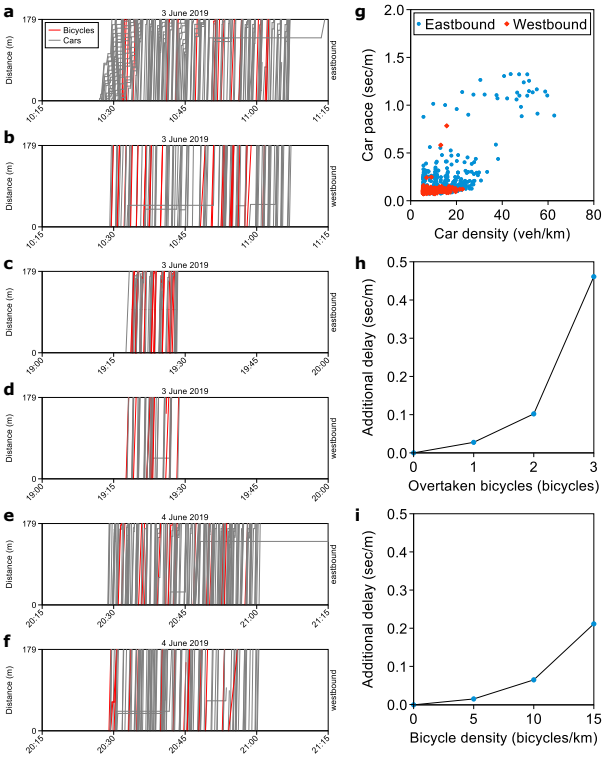


FIGURE B.4: Bi-modal measurements from Stamford Street in London. (**a**-**f**) show the raw trajectory data for cars (grey lines) and bicycles (red lines) for different time slots and two directions. Horizontal lines correspond to vehicles stopping or parking. (**g**) presents the car pace as a function of vehicle density for both corridor directions. (**h**) displays the average additional interaction delay for cars by the number of bicycles overtaken. The number of overtaken bicycles is determined by the number of bicycle trajectories cut by each car. We estimate the additional delay as the effect of previous (ordinal) variable on car pace (controlling for car density, observation day and time and corridor direction) with a linear model. (**i**) shows the average additional interaction delay for cars as a function of bicycle density. The bicycle density is calculated with Edie's 1963 for each vehicle and then binned to 0, 5, 10 and 15 bicycles per kilometer. We estimate the additional delay similar as for (**h**).

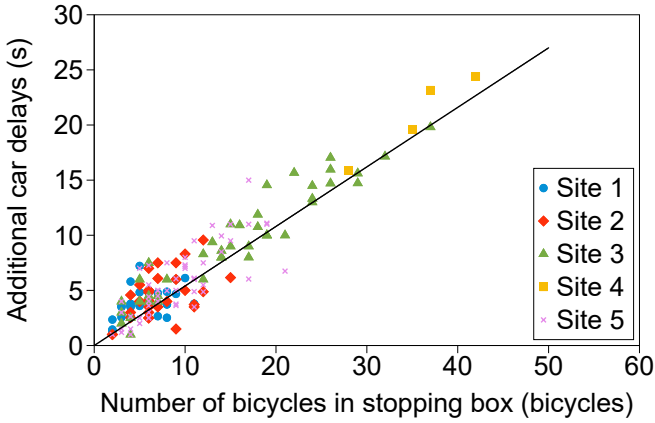


FIGURE B.5: Effects of bicycle platoon dispersion at intersections on car flow. The experimental sites refer to the locations defined in Figure B.3.

icant effect sizes of around one second. We expect that this small effect does not alter the implications of the revealed relationship.

With the existing and collected data, we then calibrate the tri-modal MFD. Unfortunately, with the limited data we were not able to estimate interaction effects along the full range of vehicle densities. However, we expect that the observed variation in the data corresponds to the observations in most cities around the world. We summarize all parameter values in Table B.2. The upper part of the table lists the parameters that define the general network configuration, the center part of the table summarizes the bus network characteristics, and the lower part gives the parameter values of the interaction situations. The parameters in the upper two parts reflect data used by Loder *et al.* (2017), Ambühl *et al.* (2018a) and Loder *et al.* (2019) and corresponds to typical values for European cities such as London or Zürich. In the lower part, we find the parameters for the continuous multiclass FD and the platoon dispersion model by minimizing with nonlinear least squares the squared difference between the observed and fitted dependent variable. All remaining parameters are identified previously directly from the empirical data. Last, we expect that additional stopping delays caused by cars follow a linear function of the additional running delays caused by cars with a slope of 0.25.

Parameters	Symbol	Unit	Global	Mode		
				Car	Bus	Bicycle
Network configuration						
Intersection spacing	l	(m)	250			
Network diameter	D	(m)	5000			
Intersection length	c	(m)	50			
Number of lanes	d	(-)	1			
Cycle length	C	(s)	60			
Green time	G	(s)	20			
Free flow speed	v^f	(m/s)		7.6	6.2	4
Capacity	Q	(veh/s)		0.149	0.06	1.6
Jam density	κ	(veh/m)		0.15	0.05	0.6
Backward wave speed	w	(m/s)		1.6	1.6	2
Initial smoothing	λ^0	(-)		0.038	0.02	0.02
Density scaling factor	h	(-)		1	1	1
Bus network configuration						
Bus stop spacing	p	(m)	333			
Bus network design	α	(-)	0.8			
Average number of transfers	e_T	(-)	0			
Fixed dwell time	ζ	(s)	30			
Dwell time per passenger	ζ'	(s/pax)	1			
Interaction situations						
Avg. bicycle travel time	T_a	(s)	17.5			
SD bicycle travel time	σ	(s)	15.9			
Platoon clearing factor	ψ	(-)	0.98			
Bicycle clearing limit	$\overline{b}$	(bicycles)	50.16			
Bicycle spacing threshold 1	k'_v	(veh/m)	0.01			
Bicycle spacing threshold 2	k''_v	(veh/m)	0.15			
Vehicle spacing threshold	k'	(veh/m)		0.01	0.003	
Reduced speed	v_{red}	(m/s)	4			
Speed reduction factor	δ	(-)	0.88			
Inv. speed-density FD parameter	μ_1	(m)	2×10^{-3}			
Inv. speed-density FD parameter	μ_2	(m²/s)	8×10^{-2}			
Inv. speed-density FD parameter	μ_3	(1/s)	2×10^{-4}			
Speed coupling parameter	ν_b	(-)	1.5×10^{-5}			
Passenger car unit intercept	φ_0	(-)	2			
Passenger car unit slope	φ'	(s/m)	3.32×10^{-2}			
Mixing parameter	ψ_0	(s)	9×10^{-2}			

TABLE B.2: List of parameters for the multi-modal interaction MFD.

B.3 TABLES

Table B.3 lists all cities with the number of days and number of analyzed regions (i.e. number of critical points) included in Section 4.3. Table B.4 lists all parameters used for the estimation of the geometric approach for the estimation of the 3D-MFD in Section 4.4.2.

No	City	Country	Population [1000]	Detectors	Days	Regions
1	Augsburg	Germany	277	777	20	1
2	Basel	Switzerland	167	83	7	1
3	Bern	Switzerland	129	769	7	1
4	Birmingham	United Kingdom	1097	114	6	1
5	Bolton	United Kingdom	128	202	22	1
6	Bordeaux	France	754	591	7	4
7	Bremen	Germany	549	583	14	2
8	Cagliari	Italy	154	133	50	1
9	Constance	Germany	81	129	7	1
10	Darmstadt	Germany	150	393	5	1
11	Duisburg	Germany	487	590	14	1
12	Essen	Germany	570	38	36	1
13	Frankfurt	Germany	701	112	1	1
14	Graz	Austria	270	300	10	1
15	Groningen	Netherlands	198	55	6	1
16	Hamburg	Germany	1746	419	105	1
17	Innsbruck	Austria	125	49	30	1
18	Kassel	Germany	194	601	4	3
19	London	United Kingdom	8478	5804	22	16
20	Los Angeles	USA	3970	4072	14	3
21	Luzern	Switzerland	81	159	361	1
22	Madrid	Spain	3142	2123	20	10
23	Manchester	United Kingdom	517	221	22	1
24	Marseille	France	1054	178	32	2
25	Melbourne	Australia	4820	1649	15	3
26	Munich	Germany	1408	548	1	2
27	Paris	France	3236	513	366	4
28	Rotterdam	Netherlands	618	227	6	1
29	Santander	Spain	176	378	3	2
30	Speyer	Germany	50	199	14	1
31	Strasbourg	France	228	220	25	1
32	Stuttgart	Germany	604	298	8	1
33	Taipei	Taiwan	2674	445	14	6
34	Tokyo	Japan	9273	2111	30	9
35	Torino	Italy	902	787	21	3
36	Toronto	Canada	2809	298	61	2
37	Toulouse	France	747	910	7	4
38	Utrecht	Netherlands	328	1072	4	1
39	Vilnius	Lithuania	540	581	1	2
40	Wolfsburg	Germany	122	405	14	1
41	Zürich	Switzerland	385	1225	7	7

TABLE B.3: List of cities in the critical point model

Parameter	Simulation		London	Zürich
	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 B.4: Network parameters for empirical validation of the 3D-MFD functional form.

C

ECONOMIC DATA

For demand side analysis as well as the calibration of the traffic assignment and equilibrium model we require also economic data that we will discuss in this Appendix. In Section C.1 we introduce the household travel survey that we use for the demand side analysis as well as calculation of benchmark variables for the equilibrium model. In Section C.2 we describe the available origin and destination matrix and in Section C.3 we summarize the sources and data processing of the economic activity data. Last, we list additional tables for the economic equilibrium model of Chapter 7 in Section C.4.

C.1 HOUSEHOLD DATA

The household data originates from the Swiss travel survey carried out every five years, called the *Mikrozensus*. This thesis uses data from 2010 and 2015 (Swiss Federal Statistical Office and Swiss Federal Office for Spatial Development, 2012, 2017). In each of the two surveys, around 60.000 inhabitants or around 1,% of the Swiss population is surveyed on their economic status, mobility tool ownership as well as travel behavior.

For the demand side analysis in Chapter 3, we extract from the 2010 survey the following information of individuals: gender, age (grouped by age categories), employment status, university degree and monthly gross household income[1]. Aside the introduced accessibility variable in Section 3.2, we use two further spatial control variables: First, a spatial typology definition by Swiss Federal Office of Spatial Development *et al.* (2011) that differentiates between urban, agglomeration and non-urban environment and, second, the quality of public transport, the latter two at the household location. For each location, the quality of public transport based on distance to the next station, frequency at this station and available lines on a five-level scale as proposed by Swiss Federal Office of Spatial De-

1 We recode the stated gross monthly household income classes into a continuous scale by assigning the midpoint value of each class to the household. As 24 % of all households did not report on their income, we impute the income with an ordered logit model. For each household that did not report income, we assign the sum of the product of probability of belonging to a class with the midpoint income class value. Results available on request.

velopment (2011). The scale ranks from E (worst) to A (best). Table C.1 summarizes all variables used for the demand side model.

For the calibration of the economic equilibrium model, we require benchmark shares for the modal shares across origin and destination pairs as well as benchmark shares of the mobility tool ownership levels. Table C.2 summarizes the benchmark values for each zone in the analysis regarding mobility tool ownership and mode choice. People living within the Zürich city limits (Kreise 1-12) largely have a season ticket and rarely rely on just a car for travel. This contrasts with individuals commuting from outside areas.

C.2 ORIGIN-DESTINATION DATA

We require information on where commuters live and work (n_{ij}), the observed mode shares as well as the mobility tool portfolio choices of commuters. A commuter matrix for Switzerland was previously made to generate a synthetic population for MATSim (a Multi-Agent Transportation Simulation) (Bösch *et al.*, 2016). The essential information used in that commuter matrix had been collected by the Swiss government for the years 2010-2012 via a nation-wide census called a *Strukturerhebung*. The spatial resolution of the MATSim commuter matrix is higher than that of our applied zoning in Figure 6.1. We mapped the origins and destinations of that matrix to the zones used for this work and aggregated the data to line up with other data used in this analysis. Because the totals of the commuter matrix do not equal the workplace totals used in the Swiss national transport model for the city of Zürich, we scaled the commuting matrix flows in n_{ij} to match aggregate official employment statistics.

C.3 ECONOMIC ACTIVITY DATA

The economic data used in the model was largely obtained from the city's statistical offices. Table C.3 lists descriptive statistics and their sources. We match all data from a comparable time period to provide consistency, limiting any outside exogenous variation. We proxy for regional output levels ($\overline{Y}_j$) by using a combination of regional gross product data, which was disaggregated based on employment levels in each Kreis, and Kreis-level employment and commercial property rental data. The number of employed workers was collected from a national employment and workplace survey in 2008. Wages were obtained from a nation-wide transportation survey

Categorical variables				
	Share			
Person is male	45.53 %			
Age categories				
> 70	16.96 %			
61-70	17.39 %			
51-60	17.59 %			
41-50	19.09 %			
31-40	15.11 %			
< 31	13.86 %			
Employed	62.26 %			
University degree	16.32 %			
Quality of public transport at household location				
Level A: very good	12.60 %			
Level B: good	16.11 %			
Level C: moderate	20.85 %			
Level D: low	26.70 %			
Level E: very low	23.76 %			
Spatial typology at household location				
City	32.50 %			
Agglomeration	48.45 %			
Countryside	19.05 %			
Continuous variables				
	Mean	SD	Min	Max
General accessibility				
	1.53	1.56	-10.09	5.14
Better accessibility by public transport				
	-0.01	0.61	-1.76	2.26
Better job accessibility				
	0.03	0.12	-0.40	0.42
Log of gross monthly household income in CHF				
	8.75	0.56	7.31	9.90

TABLE C.1: Sample summary statistics. The upper part lists categorical variables and the lower part continuous variables.

conducted in 2015 by the Swiss Federal Statistical Office and Swiss Federal Office for Spatial Development (2017). To calculate relevant wage levels, all full-time employed respondents working in Zürich were selected. As only household income was reported in the survey, the reported household income is divided by the square root of household members older than 18 years. Hourly wages are calculated by dividing monthly income by 160 working hours.[2] The value of the regional capital stock is composed of the value of commercial floor space and a specific factor. The specific factor is the unobservable component of the regional gross product, or rather, the difference between gross product levels and observed labour costs and commercial rent. In other words, we distinguish between the observable capital stock component labelled as capital generally, and the unobservable component labelled as a specific factor. The capital rental rate, i.e., the rent for commercial floor space, is available from a private company survey. Regional housing stock levels and prices are characterized by residential floor space and average rent. The average rent is a weighted average over all rented flats featuring 1 to 5 rooms.

The economic data present a relatively intuitive reference point. Many people work in Kreis 1, which features high wages and a limited housing stock, and live further from the city centre, primarily north or along the coast of the lake. Notably, there is limited variation in the housing rental prices which is consistent with the high cost of living in Zürich.

C.4 TABLES

2 Note that given our characterization of the labour market in Zürich, the average Kreis-level labour value share of production is 0.8.

	Mobility tool portfolio[1]			Mode[2]	
Zone	Car	Season ticket	Both	Car	Bus
1	0.041	0.679	0.279	0.058	0.942
2	0.034	0.556	0.410	0.125	0.875
3	0.032	0.570	0.398	0.110	0.890
4	0.038	0.605	0.358	0.077	0.923
5	0.013	0.516	0.471	0.073	0.927
6	0.034	0.526	0.440	0.138	0.862
7	0.047	0.518	0.435	0.145	0.855
8	0.012	0.395	0.593	0.157	0.843
9	0.034	0.574	0.392	0.090	0.910
10	0.036	0.545	0.418	0.090	0.910
11	0.043	0.531	0.425	0.146	0.854
12	0.054	0.525	0.421	0.183	0.817
101	0.235	0.508	0.257	0.322	0.678
102	0.380	0.332	0.288	0.569	0.431
103	0.258	0.450	0.291	0.352	0.648
104	0.281	0.502	0.218	0.394	0.606
105	0.238	0.325	0.437	0.410	0.590
106	0.262	0.475	0.263	0.398	0.602
107	0.655	0.100	0.244	0.827	0.173
108	0.317	0.369	0.315	0.468	0.532
109	0.219	0.510	0.271	0.296	0.704
110	0.231	0.425	0.344	0.351	0.649
111	0.130	0.521	0.350	0.222	0.778

[1] Calculated as the share of all people employed in each zone based on the 2015 Swiss transportation survey (Swiss Federal Statistical Office and Swiss Federal Office for Spatial Development, 2017), irrespective of their actual mode choices. Those without any mobility tool were assigned to the group of season ticket holders.

[2] Calculated as the share of all outbound commuting trips from each zone according to the 2015 Swiss transportation survey.

TABLE C.2: Benchmark shares of mobility tool ownership and mode choice.

Kreis	Output ($\overline{Y}$)[1] (Mil CHF/day)	Employment (Mil)	Wages ($\overline{W}$)[2] (CHF/hr)	Capital stk. ($\overline{K}$)[3] (Mil m^2)	Office rent ($\overline{R}$)[4] (CHF/m^2/day)	Housing stk. ($\overline{H}$)[3] (Mil m^2)	Housing price ($\overline{PH}$)[3] (CHF/m^2/day)
1	32.6	0.071	45.5	1.2	1.4	0.3	1.1
2	20.1	0.034	48.9	0.9	1.3	1.6	0.7
3	16.3	0.034	43.7	0.9	0.7	2.1	0.7
4	12.3	0.030	39.4	0.7	0.8	1.2	0.7
5	13.2	0.035	43.1	1.3	0.9	0.7	0.6
6	5.9	0.016	38.3	0.4	0.8	1.6	0.7
7	7.8	0.019	39.4	0.5	1.0	2.2	0.7
8	9.1	0.023	42.6	0.5	1.2	0.9	0.8
9	18.4	0.040	39.6	1.3	0.7	2.2	0.6
10	4.1	0.014	38.2	0.3	0.7	1.8	0.7
11	15.0	0.044	42.5	1.4	0.7	3.1	0.6
12	1.4	0.005	33.8	0.1	0.6	1.1	0.6

[1] Statistisches Jahrbuch der Stadt Zürich 2015; the city's total GDP is divided across Kreise according to employment.
[2] Swiss Federal Statistical Office and Swiss Federal Office for Spatial Development (2017)
[3] Statistisches Jahrbuch der Stadt Zürich 2014
[4] Market survey by JLL (2015).

Table C.3: Benchmark values for the economic equilibrium.

Variable type	Symbol	Description
Economic	Y_i	Production at node i
	U_{ij}	Utility realized living at i and working at j
	P	Macro output price
	W_i	Wage rate in zone i
	R_i	Rental rate for capital in zone i
	PSF_j	Price of specific factor in zone i
	PH_i	Housing price in zone i
	PLS_{ij}	Shadow value of leisure for living in i and working in j
	PU_{ij}	Unit price of utility for living in i and working in j
	LL	Absentee landlord
	RA_{ij}	Representative agent living at i and working in j
	LD_j	Labour demand at node j
	X_j	Productivity index at node j
	NX_{ij}	Effective level of labour net of productivity gains
	NLS_{ij}	Total leisure travelling from i to j
	ACC_j	Accessibility index at node j
Transport	T_{ijmr}	Total travel times from commuting from i to j
	C_{ijmr}	Total path cost from i to j
	$\check{C}_{ijmr}$	Total perceived path cost from i to j
	V_{km}	Speed of mode m in zone k
	A_{km}	Accumulation of vehicles of mode m in zone k
	M_{ij}	Minimum perceived path costs between i and j
	ρ	Shadow price (mobility tools, parking, bus capacity)
	Q_{ijt}	Shares of mobility tool ownership between i and j
Sorting	UL_i	Utility from living at node i
	NL_i	Number of people living in i
	NLW_{ij}	Number of people living in i and working in j
	N_{ijmr}	NLW_{ij} using mode m and route r

TABLE C.4: Economic model variables

Parameter type	Symbol	Description
Economic	θ_i^l	Value share of labour in zone i
	θ_i^k	Value share of capital in zone i
	θ_i^s	Value share of the specific factor in zone i
	θ^{ls}	Value share of leisure
	θ^c	Value share of consumption
	θ_{ij}^u	Proportion of total residents at i working in j
	θ_i^{li}	Proportion of total residents living at i
	σ^c	Elasticity of substitution in demand for the composite good
	$\overline{K}_i$	Reference level of capital stock in zone i
	$\overline{SF}_i$	Reference level of specific factor stock in zone i
	$\overline{H}_i$	Reference level of housing stock in zone i
	$\overline{u}_{ij}$	Reference utility level for living in i and working in j
	β	Productivity parameter measuring changes in productivity
	μ^A	Parameter measuring impact of worker mass on productivity
	θ_{ij}^u	Proportion of total residents at i working in j
	θ_i^{li}	Proportion of total residents living at i
	μ^L	Elasticity of housing location choice
	μ^W	Elasticity of work location choice
	μ^A	Impact of economic mass on productivity.
	β	Scaling of impact of economic mass on output.
Transport network	L_k	Infrastructure length
	B_k	Length of bus network
	φ_k	Share of dedicated bus lanes
	H_k	Headway of buses
	z_k	Bus line overlapping factor
	α_k	Bus network design
	$\overline{\varphi}_{ij}$	Amenity factor of mode m using route r
	d_{ijmr}	Macro route distance
	θ_{ijkmr}^d	Share of route distance in zone k
	μ^R	Scale parameter of route and mode choice
	μ^M	Price elasticity of mobility tool ownership
	$\overline{Z}_k$	Public transport capacity in k
	π_{ijt}^{total}	Total cost of mobility tool ownership for portfolio t between i and j
	π_t^{fix}	Fixed cost of mobility tool for mode m
	π_{tm}^{var}	Variable cost of mobility tool for mode m
	n_{ij}	Benchmark origin-destination matrix

TABLE C.5: Economic model parameters

Kreis	$V_{0,c}$[1] (km h^{-1})	$\overline{V}_c$[2] (km h^{-1})	$V_{0,b}$[4] (km h^{-1})	Δ[6] (s)	h[3] (min)	$\overline{A}_b$[4] (-)	B_k[5] (lane-km)
1	31.2	9.7	16.5	27	6	141	26.9
2	37.9	13.1	25.3	26	6	89	34.7
3	34.4	13.0	22.7	22	6	116	39.4
4	31.8	10.16	21.3	29	6	64	14.5
5	36.3	11.6	23.1	18	6	41	15.9
6	36.5	13.3	22.9	20	6	72	30.1
7	36.0	15.0	22.3	32	6	65	28.1
8	36.9	18.1	22.3	17	6	38	18.8
9	36.0	14.7	25.1	26	6	118	47.4
10	38.1	15.8	26.2	18	6	116	34.1
11	36.8	13.7	24.5	26	6	145	43.9
12	36.9	13.3	27.5	23	6	37	23.2

[1] Estimated average speed for car journeys between 05:30 to 06:00 based on TomTom's speed profiles data-base as used by Ambühl *et al.* (2017).
[2] Estimated average journey speeds for the weekday morning peak from 08:00 to 08:30 based on TomTom's speed profiles data-base as used by Ambühl *et al.* (2017).
[3] Provided by public time table information.
[4] Estimated based on a trajectory data set of all public transport lines. Estimated as the mean during the weekday morning peak from 08:00 to 08:30.
[5] Following the definition given by Daganzo (2010).

Table C.6: Observations for the calibration of the 3D-MFD.

Kreis	L_k	Share of dedicated bus lanes	α_k	z_k
	(lane-km)	(-)	(-)	(-)
1	93.3	0.14	0.605	5.12
2	177.2	0.10	0.459	4.21
3	143.8	0.14	0.569	4.00
4	91.9	0.08	0.780	5.49
5	69.4	0.12	0.734	3.49
6	141.9	0.11	0.954	3.39
7	77.8	0.18	0.733	3.03
8	89.36	0.11	0.752	2.67
9	145.9	0.16	0.675	3.64
10	123.4	0.14	0.635	5.23
11	165.1	0.13	0.584	4.82
12	116.7	0.09	0.743	2.54

TABLE C.7: 3D-MFD calibration parameters.

D

MATHEMATICAL FORMULATIONS

D.1 MAXIMUM LIKELIHOOD ESTIMATION

The discrete choice model presented in Chapter 3 requires the estimation of the model parameters, usually denoted as β's. Thus, we need to find an estimator that is, first, *efficient*[1], and second that it has the asymptotic property of *consistency*[2]. A widely used and very general estimator is maximum likelihood, which "*stated simply, a maximum likelihood estimator is the value of the parameters for which the observed sample is most likely to have occurred*" (Ben-Akiva and Lerman, 1985).

Consider now the follow notation. Assuming that all observations in the sample of size N are independently and randomly drawn from the population. Further, let $\mathbf{x_i}$ be a vector of observed and non-stochastic variables of the i-th observation, $\#$ is the parameter vector, y_i is the observed choice of individual i, but assumed to be a random variable and f is a function mapping y_i, $\mathbf{x_i}$ and $\#$ to a probability value. Then, the likelihood of the sample is defined as given by Eqn. D.1.

$$\mathcal{L} = \prod_{i=1}^{N} f\left(y_i | \mathbf{x_i}, \#\right) \tag{D.1}$$

However, for convenience, the logarithm of $\mathcal{L}$ is used, which does not change the values of the parameter estimates as the logarithmic function is strictly monotonically increasing. Then, the problem to estimate the optimal set of parameters $\hat{\#}$ is given by Eqn. D.2.

$$\max_{\hat{\#}} \log \mathcal{L} = \max_{\hat{\#}} \sum_{i=1}^{N} \log f\left(y_i | \mathbf{x_i}, \hat{\#}\right) \tag{D.2}$$

In case a solution to Eqn. D.2 exists, it must satisfy the usual first order conditions $\dfrac{\partial \log \mathcal{L}}{\partial \hat{\#}} = 0$ and the second order condition that at the maximum the Hessian matrix $\nabla^2 \log \mathcal{L}$ must be negative semi-definite at $\hat{\#}$.

1 The estimator is unbiased and no other unbiased estimator has smaller variance

2 The estimator is consistent as when the sample size increases the distribution converges or even collapses to the true parameters.

In the case of f can be formulated in a closed way, the problem in Eqn. D.2 can be solved with a conventional optimizer that is using, for example, the Newton-Raphson algorithm. When f cannot be formulated in a closed form, which is the case when f is the multivariate cumulative normal distribution function, one has to rely on simulation methods (Train, 2009) or approximation methods, namely the maximum approximate composite marginal likelihood (MACML) (Bhat and Sidharthan, 2011; Bhat, 2011) or advanced matrix decomposition methods (Bhat, 2018).

D.2 MATHEMATICAL FORMULATION OF OPTIMIZATION PROBLEMS

In this thesis, I refer several times to optimization problems, as well as equilibrium problems and mixed complementarity problems. In this section, I refine these problems mathematically. I have taken this section in large parts in revised form from Bliemer (2001). Section D.2.1 generally formulates an optimization problem, then we discuss this formulation as a variational inequality problems in Section D.2.2 and then as (mixed) complementarity problem in Section D.2.3.

D.2.1 *Optimization problems*

First, we define that $\mathbf{x} \in \mathbb{R}^n$ is a vector of decision variables with n elements. Further, $\mathbf{X} \subseteq \mathbb{R}^n$ is a nonempty, closed and convex set; and $g : \mathbf{X} \to \mathbb{R}^n$ is a continuous function. With this definitions, a problem for finding a minimum is formulated as given in Eqn. D.3. Here, the problem is to find that $\overline{\mathbf{x}} \in \mathbf{X}$ that minimizes g.

$$\overline{\mathbf{x}} = \arg\min_{x \in X} g(\mathbf{x}) \tag{D.3}$$

Contrary, a problem for finding a maximum is similarly defined as given by Eqn. D.4. Here, then the the problem is to find that $\overline{\mathbf{x}} \in \mathbf{X}$ that maximizes g.

$$\overline{\mathbf{x}} = \arg\max_{x \in X} g(\mathbf{x}) \tag{D.4}$$

D.2.2 *Variational inequality problems*

A Variational inequality problem is defined as follows. Again $\mathbf{x} \in \mathbb{R}^n$ is a vector of decision variables and, again, $\mathbf{X} \subseteq \mathbb{R}^n$ is a nonempty, closed

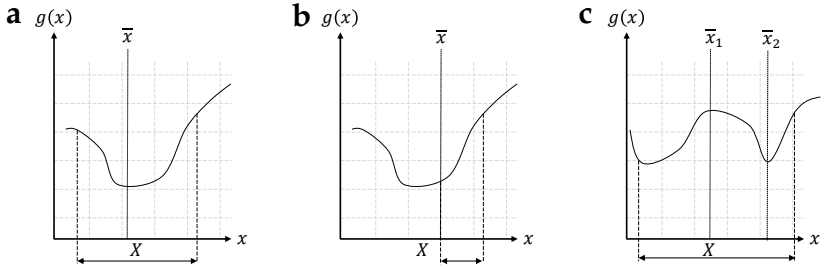


FIGURE D.1: Comparing solutions to the variational inequality problem and the optimization problem. The illustration is a revised version of Bliemer (2001).

and convex set. Now, we define that $\mathbf{f} : X \rightarrow \mathbb{R}^n$ is a vector of continuous functions. Then, the variational inequality problem is defined as by Eqn. D.5. Here, the problem is to find that $\overline{\mathbf{x}}$ such that the inequality holds.

$$\mathbf{f}\left(\overline{\mathbf{x}}\right)^T \left(\mathbf{x} - \overline{\mathbf{x}}\right) \geq 0, \qquad \forall \mathbf{x} \in \mathbf{X}. \tag{D.5}$$

Following Nagurney's 1993 proof, the solution $\overline{\mathbf{x}}$ to Eqn. D.5 is also a solution to the optimization problem from Eqn. D.3 only if the following two conditions hold: First, $\nabla_{\mathbf{x}}\mathbf{f}\left(\mathbf{x}\right)$ is symmetric for all $\mathbf{x} \in \mathbf{X}$ and, second, a pseudoconvex function $h\left(\mathbf{x}\right)$ exists that satisfies $\nabla_{\mathbf{x}}\mathbf{h}\left(\mathbf{x}\right) = \mathbf{f}\left(\mathbf{x}\right)$.

We can graphically illustrate this condition for the solution of both problems in Figure D.1. In this example, we have a one-dimensional problem, i.e. $n = 1$. Further, we set $f\left(x\right) = g'\left(x\right)$. In Figure D.1**a** we see plot $g\left(x\right)$ over $\mathbf{X}$. We see that the solution $\overline{x}$ is clearly a solution to the optimization problem from Eqn. D.3 and to the variational inequality problem from Eqn. D.5 because $g'\left(\overline{x}\right) = 0$. Figure D.1**b** then emphasizes that $\overline{x}$ is also a solution of both problems when it is at the boundary of $\mathbf{X}$. However, Figure D.1**c** shows two points, $\overline{x}_1$ and $\overline{x}_2$, which are both solutions to the variational inequality problem, but $\overline{x}_1$ is not a solution to the optimization problem because $g\left(x\right)$ is not pseudoconvex on $\mathbf{X}$.

D.2.3 *Complementarity problems*

The mixed complementarity problem (MCP) used in Chapters 5, 6 and 7 are closely related to variational inequality problems formulated in the previous section. The key difference between a variational inequality and

a complementarity problem is that the latter is only defined on the non-negative orthant $\mathbb{R}^n_+$. For such a complementarity problem, we again set $\mathbf{x} \in \mathbb{R}^n_+$ as the vector of decision variables. Let then $\mathbf{h}(\mathbf{x}) : \mathbb{R}^n_+ \to \mathbb{R}^n$ be a vector of continuous functions. The complementarity problem is then to find a vector $\bar{\mathbf{x}} \in \mathbb{R}^n_+$ such that

$$\mathbf{h}(\bar{\mathbf{x}})^T \bar{\mathbf{x}} = 0 \tag{D.6}$$

$$\mathbf{h}(\bar{\mathbf{x}})^T \geq \mathbf{0} \tag{D.7}$$

In Eqn. D.6 we clearly see the key feature of a complementarity problem: either the constraint or the associated variable has to be zero. This makes the complementarity a feature rather than a requirement of the model. Here, we use Rutherford's 1995 notion of the $\perp$ symbol to indicate the complementarity between the constraint and associated variable as given by Eqn. D.8.

$$\mathbf{h}(\bar{\mathbf{x}})^T \geq \mathbf{0} \perp \bar{\mathbf{x}} \geq 0 \tag{D.8}$$

Then, mixed complementarity problem (MCP) can also have equations with free associated variables, i.e. no associated constraints. The associated variables are then not constraint to the non-negative orthant. In other words, we a MCP is defined as given by Eqn. D.9.

$$\begin{aligned} & h_i(\bar{\mathbf{x}}) = 0 \quad \text{and} \quad l_i < \bar{x}_i < u_i \\ \text{or} \quad & h_i(\bar{\mathbf{x}}) \geq 0 \quad \text{and} \quad \bar{x}_i = l_i \\ \text{or} \quad & h_i(\bar{\mathbf{x}}) \leq 0 \quad \text{and} \quad \bar{x}_i = u_i \end{aligned} \tag{D.9}$$

Here, l and u are lower and upper bounds of the associated variables, which can be infinite given the mixed nature of the problem. We then define a box $\mathbf{B} \equiv [l, u] = \{x \in \mathbb{R}^n : l_i \leq x_i \leq u_i\}$ with $l \in < \{\mathbb{R} \cup -\infty\}$ and $u \in < \{\mathbb{R} \cup \infty\}$. Intuitively, for both bounds, $-\infty \leq l_i < u_i \leq \infty$ must always hold. Lastly, this formulation allows to formulate a MCP rather compact with MCP $(\mathbf{h}, \mathbf{B})$.

D.3 MATHEMATICAL PROBLEM WITH EQUILIBRIUM CONSTRAINTS

The mathematical problem with equilibrium constraints is a mathematical program with an optimization problem in its constraints (Bracken and

McGill, 1973). In particular, a mathematical problem with equilibrium constraints (MPEC) is a constrained nonlinear programming problem where the primary (first level) objective to be minimized or maximized and the some or all constraints of the problem are formulated of the solution of a second level problem that describe an equilibrium (Luo *et al.*, 1996). Thus, it is crucial element of the MPEC to define equilibrium constraints. Here, the constraints can result from an optimization problem, variational inequalities, the min operator, or a complementarity problem (Robinson, 1980; Pang and Harker, 1990).

In Chapter 6, we are able to derive the second level constraints as a mixed complementarity problem (MCP) as introduced in Section D.2.3. Here, we follow Dirkse and Ferris (1998). The MPEC as given by Eqn. D.10 is defined with a set of design variables $x \in \mathbf{X} \subseteq \mathbb{R}^n$ and a set of state variables $y \in \mathbf{Y} \subseteq \mathbb{R}^m$. Variables from both sets of variables may enter the objective function $\theta : \mathbb{R}^{m+n} \to \mathbb{R}$ that is to be optimized. Then, the problem has two different kind of constraints. The first constraints require the feasibility of variables x and y, determined by functions h, g and box $\mathbf{X}$. These constraints can either be an inequality or an equation. Then the second set gives the equilibrium constraints with box $\mathbf{X}$ and function $F : \mathbb{R}^{m+n} \to \mathbb{R}^m$ or, in other words, that it solves the MCP defined by F and $\mathbf{Y}$.

$$\begin{aligned} \text{minimize} \quad & \theta(x, y) \\ \text{subject to} \quad & h(x) \in H \\ & x \in \mathbf{X} \\ & g(x, y) \in G \\ \text{and} \quad & y \text{ solves MCP}\left(F(x, \cdot), \mathbf{Y}\right) \end{aligned} \tag{D.10}$$

BIBLIOGRAPHY

Aboudolas, K. and N. Geroliminis (2013) Perimeter and boundary flow control in multi-reservoir heterogeneous networks, *Transportation Research Part B: Methodological*, **55**, 265–281.

Adbulaal, M. and L. J. LeBlanc (1979) Continuous equilibrium network design models, *Transportation Research Part B: Methodological*, **13** (1) 19–32.

Adler, M. W. and J. N. van Ommeren (2016) Does public transit reduce car travel externalities? Quasi-natural experiments' evidence from transit strikes, *Journal of Urban Economics*, **92**, 106–119.

Aghamohammadi, R. and J. A. Laval (2018) Dynamic traffic assignment using the macroscopic fundamental diagram: A Review of vehicular and pedestrian flow models, *Transportation Research Part B: Methodological*, **in press**.

Ahlfeldt, G. M., S. J. Redding, D. M. Sturm and N. Wolf (2015) The Economics of Density: Evidence From the Berlin Wall, *Econometrica*, **83** (6) 2127–2189.

Alessandretti, L., P. Sapiezynski, V. Sekara, S. Lehmann and A. Baronchelli (2018) Evidence for a conserved quantity in human mobility, *Nature Human Behaviour*, **2** (7) 485–491.

Allen, T. and C. Arkolakis (2019) The Welfare Effects of Transportation Infrastructure Improvements, *Technical Report*, **25487**, National Bureau of Economic Research.

Ambühl, L., A. Loder, M. C. Bliemer, M. Menendez and K. W. Axhausen (2018a) A functional form for the macroscopic fundamental diagram with a physical meaning, *Transportation Research Part B: Methodological*, **accepted**.

Ambühl, L., A. Loder, M. C. Bliemer, M. Menendez and K. W. Axhausen (2018b) Introducing a Re-Sampling Methodology for the Estimation of Empirical Macroscopic Fundamental Diagrams, *Transportation Research Record*, **2672**, 239–248.

Ambühl, L., A. Loder, M. Menendez and K. W. Axhausen (2017) Empirical macroscopic fundamental diagrams: New insights from loop detector and floating car data, paper presented at the *96th Annual Meeting of the Transportation Research Board*, Washington D.C.

Ambühl, L., A. Loder, M. Menendez and K. W. Axhausen (2018c) A case study of Zurich's two-layered perimeter control, paper presented at the *7th Transport Research Arena*, Vienna.

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*, **accepted**.

Ambühl, L. and M. Menendez (2016) Data fusion algorithm for macroscopic fundamental diagram estimation, *Transportation Research Part C: Emerging Technologies*, **71**, 184–197.

Amirgholy, M. and H. O. Gao (2017) Modeling the dynamics of congestion in large urban networks using the macroscopic fundamental diagram: User equilibrium, system optimum, and pricing strategies, *Transportation Research Part B: Methodological*, **104**, 215–237.

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.

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.

Anas, A. and I. Kim (1996) General Equilibrium Models of Polycentric Urban Land Use with Endogenous Congestion and Job Agglomeration, *Journal of Urban Economics*, **40** (2) 232–256.

Anas, A. and R. Lindsey (2011) Reducing urban road transportation externalities: Road pricing in theory and in practice, *Review of Environmental Economics and Policy*, **5**, 66–88.

Anas, A. and R. Xu (1999) Congestion, Land Use, and Job Dispersion: A General Equilibrium Model, *Journal of Urban Economics*, **45** (3) 451–473.

Anderson, M. L. (2014) Subways, Strikes, and Slowdowns: The Impacts of Public Transit on Traffic Congestion, *American Economic Review*, **104** (9) 2763–2796.

Andrés, J. and G. Gélvez (2014) Joint Disaggregate Modeling of Car and Motorcycle Ownership: A Case Study of Bogotá, Colombia, *Transportation Research Record*, **2451**, 149–156.

Anowar, S., N. Eluru and L. F. Miranda-Moreno (2014) Alternative Modeling Approaches Used for Examining Automobile Ownership: A Comprehensive Review, *Transport Reviews*, **34** (4) 441–473.

Antonius, M. A. (1949) *Des Kaisers Marcus Aurelius Antonius Selbstbetrachtungen*, Verlag von Philipp Reclam jun., Lepzig.

Ardekani, S. and R. Herman (1982) Quality of Traffic Service, *Technical Report*, **2**, Center for Transportation Research, The University of Texas at Austin.

Ardekani, S. A. and R. Herman (1985) A comparison of the quality of traffic service in downtown networks of various cities around the world, *Traffic Engineering and Control*, **26**, 574–581.

Ardekani, S. A., J. C. Williams and S. Bhat (1992) Influence of urban network features on quality of traffic service, *Transportation Research Record*, **1358**, 6–12.

Arnott, R. (2007) Congestion tolling with agglomeration externalities, *Journal of Urban Economics*, **62** (2) 187–203.

Arnott, R. (2013) A bathtub model of downtown traffic congestion, *Journal of Urban Economics*, **76**, 110–121.

Arnott, R. and K. Small (1994) The Economics of Traffic Congestion, *American Scientist*, **82** (5) 446–455.

Arnott, R. and A. Yan (2000) The two-mode problem: Second-best pricing and capacity, *Review of urban & regional development studies*, **12** (3) 170–199.

Aschauer, D. A. (1989) Is public expenditure productive?, *Journal of Monetary Economics*, **23** (2) 177–200.

Axhausen, K. W., T. Bischof, R. Fuhrer, R. Neuenschwander, G. Sarlas and P. Walker (2015) Gesamtwirtschaftliche Effekte des öffentlichen Verkehrs mit besonderer Berücksichtigung der Verdichtungs- und Agglomerationseffekte, *Technical Report*, IVT, ETH Zürich und Ecoplan, St. Gallen. Bern und Zürich.

Axhausen, K. W. and H.-G. Körling (1987) Some Measurements of Robertson's Platoon Dispersion Factor, *Transportation Research Record*, **1112**, 71–77.

Axhausen, K. W., A. Zimmermann, S. Schönfelder, G. Rindsfüser and T. Haupt (2002) Observing the rhythms of daily life: A six-week travel diary, *Transportation*, **29** (2) 95–124.

Baker, R. J., J. Collura, J. J. Dale, L. Head, B. Hemily, M. Ivanovic, J. T. Jarzab, D. McCormick, J. Obenberger and L. Smith (2002) An overview of transit signal priority, *Technical Report*, ITS America, Washington DC.

Banister, D. and Y. Berechman (2001) Transport investment and the promotion of economic growth, *Journal of Transport Geography*, **9** (3) 209–218.

Batista, S. and L. Leclercq (2018a) Introduction of multi-regional MFD-based models with route choices : the definition of regional paths, paper presented at the *8th Luso-Brazilian Congress for Urban, Regional, Integrated and Sustainable Planning*, Coimbra.

Batista, S. and L. Leclercq (2018b) Regional dynamic traffic assignment framework for multi-reservoir MFD models, paper presented at the *7th Symposium of the European Association for Research in Transportation (hEART 2018)*, Athens.

Batty, M. (2008) The size, scale, and shape of cities., *Science*, **319**, 769–771.

Baum-Snow, N., L. Brandt, J. V. Henderson, M. A. Turner and Q. Zhang (2017) Roads, Railroads, and Decentralization of Chinese Cities, *The Review of Economics and Statistics*, **99** (3) 435–448.

Beaudoin, J., Y. H. Farzin and C. Y. Lin Lawell (2015) Public transit investment and sustainable transportation: A review of studies of transit's impact on traffic congestion and air quality, *Research in Transportation Economics*, **52**, 15–22.

Becker, H., F. Ciari and K. W. Axhausen (2017a) Comparing car-sharing schemes in Switzerland: User groups and usage patterns, *Transportation Research Part A: Policy and Practice*, **97**, 17–29.

Becker, H., F. Ciari and K. W. Axhausen (2017b) Modeling free-floating car-sharing use in Switzerland: A spatial regression and conditional logit approach, *Transportation Research Part C: Emerging Technologies*, **81**, 286–299.

Becker, H., A. Loder, B. Schmid and K. W. Axhausen (2017c) Modeling car-sharing membership as a mobility tool: A multivariate Probit approach with latent variables, *Travel Behaviour and Society*, **8**, 26–36.

Beckmann, M., C. B. McGuire and C. B. Winsten (1956) *Studies in the Economics of Transportation*, Yale University Press, New Haven, CT.

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

Bento, A. M., M. L. Cropper, A. M. Mobarak and K. Vinha (2005) The effects of urban spatial structure on travel demand in the United States, *Review of Economics and Statistics*, **87** (3) 466–478.

Berger, T., C. Chen and C. B. Frey (2018) Drivers of disruption? Estimating the Uber effect, *European Economic Review*, **110**, 197–210.

Bettencourt, L. M., J. Lobo, D. Strumsky and G. B. West (2010) Urban scaling and its deviations: Revealing the structure of wealth, innovation and crime across cities, *PLoS ONE*, **5** (11) 20–22.

Bettencourt, L. M. A. (2013) The origins of scaling in cities., *Science*, **340**, 1438–1441.

Bettencourt, L. M. A., J. Lobo, D. Helbing, C. Kühnert and G. B. West (2007) Growth, innovation, scaling and the pace of life in cities, *Proceedings of the National Academy of Sciences*, **104** (17) 7301–7306.

Bhat, C. R. (2005) A multiple discrete–continuous extreme value model: formulation and application to discretionary time-use decisions, *Transportation Research Part B: Methodological*, **39**, 679–707.

Bhat, C. R. (2011) The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models, *Transportation Research Part B: Methodological*, **45** (7) 923–939.

Bhat, C. R. (2015) A new generalized heterogeneous data model (GHDM) to jointly model mixed types of dependent variables, *Transportation Research Part B: Methodological*, **79**, 50–77.

Bhat, C. R. (2018) New matrix-based methods for the analytic evaluation of the multivariate cumulative normal distribution function, *Transportation Research Part B: Methodological*, **109**, 238–256.

Bhat, C. R., S. Astroza, R. Sidharthan, M. J. B. Alam and W. H. Khushefati (2014) A joint count-continuous model of travel behavior with selection based on a multinomial probit residential density choice model, *Transportation Research Part B: Methodological*, **68**, 31–51.

Bhat, C. R., A. R. Pinjari, S. K. Dubey and A. S. Hamdi (2016) On accommodating spatial interactions in a Generalized Heterogeneous Data Model (GHDM) of mixed types of dependent variables, *Transportation Research Part B: Methodological*, **94**, 240–263.

Bhat, C. R. and S. Sen (2006) Household vehicle type holdings and usage: an application of the multiple discrete-continuous extreme value (MDCEV) model, *Transportation Research Part B: Methodological*, **40** (1) 35–53.

Bhat, C. R. and R. Sidharthan (2011) A simulation evaluation of the maximum approximate composite marginal likelihood (MACML) estimator for mixed multinomial probit models, *Transportation Research Part B: Methodological*, **45**, 940–953.

Bickel, P., C. Chen, J. Kwon, J. Rice, E. v. Zwet and P. Varaiya (2007) Measuring Traffic, *Statistical Science*, **22** (4) 581–597.

Bliemer, M. C. (2001) Analytical dynamic traffic assignment with interacting user-classes, Ph.D. Thesis, TU Delft.

Bliemer, M. C. and M. P. Raadsen (2018) Continuous-time general link transmission model with simplified fanning, part I: Theory and link model formulation, *Transportation Research Part B: Methodological*, **in press**.

Bliemer, M. C. J., M. P. H. Raadsen, E.-S. Smits, B. Zhou and M. G. H. Bell (2014) Quasi-dynamic traffic assignment with residual point queues incorporating a first order node model, *Transportation Research Part B: Methodological*, **68**, 363–384.

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.

Bösch, P. M. (2018) Autonomous Vehicles - The next Revolution in Mobility, Ph.D. Thesis, ETH Zurich.

Bösch, P. M., F. Becker, H. Becker and K. W. Axhausen (2017) Cost-based analysis of autonomous mobility services, *Transport Policy*, **64**, 76–91.

Bösch, P. M., K. Müller and F. Ciari (2016) The IVT 2015 baseline scenario, paper presented at the *16th Swiss Transport Research Conference*, Ascona.

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

Boyce, D., B. Janson and R. Eash (1981) The effect on equilibrium trip assignment of different link congestion functions, *Transportation Research Part A: General*, **15** (3) 223–232.

Boyce, D. E. (1984) Urban Transportation Network-Equilibrium and Design Models: Recent Achievements and Future Prospects, *Environment and Planning A: Economy and Space*, **16** (11) 1445–1474.

Bracken, J. and J. T. McGill (1973) Mathematical Programs with Optimization Problems in the Constraints, *Operations Research*, **21** (1) 37–44, 2 1973.

Branston, D. (1976) Link capacity functions: A review, *Transportation Research*, **10** (4) 223–236.

Brinkman, J. C. (2016) Congestion, agglomeration, and the structure of cities, *Journal of Urban Economics*, **94**, 13–31.

Brueckner, J. K. and D. Fansler (1983) The economics of urban sprawl: Theory and evidence on the spatial size of cities, *Review of Economics and Statistics*, **55**, 479–482.

Buehler, R., J. Pucher, R. Gerike and T. Götschi (2016) Reducing car dependence in the heart of Europe: lessons from Germany, Austria, and Switzerland, *Transport Reviews*, **37** (1) 4–28.

Buisson, C. and C. Ladier (2009) Exploring the Impact of Homogeneity of Traffic Measurements on the Existence of Macroscopic Fundamental Diagrams, *Transportation Research Record*, **2124**, 127–136.

Bureau of Public Roads (1964) Traffic Assignment Manual, *Technical Report*, US Department of Commerce, Urban Planning Division, Washington D.C.

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.

Cats, O., T. Reimal and Y. Susilo (2014) Public Transport Pricing Policy - Empirical Evidence from a Fare-Free Scheme in Tallinn, Estonia, *Transportation Research Record*, **2415**, 89–96.

Ceder, A. and N. H. M. Wilson (1986) Bus network design, *Transportation Research Part B: Methodological*, **20** (4) 331–344.

Cervero, R. (2002) Induced Travel Demand: Research Design, Empirical Evidence, and Normative Policies, *Journal of Planning Literature*, **17** (1) 3–20, 8 2002.

Cervero, R. (2003) Road expansion, urban growth, and induced travel: A path analysis, *Journal of the American Planning Association*, **69** (2) 145–163.

Cervero, R. and M. Hansen (2002) Induced travel demand and induced road investment: A simultaneous equation analysis, *Journal of Transport Economics and Policy*, **36** (3) 469–490.

Cervero, R. and K. Kockelman (1997) Travel demand and the 3Ds: Density, diversity, and design, *Transportation Research Part D-Transport and Environment*, **2** (3) 199–219.

Cervero, R. and J. Murakami (2010) Effects of built environments on vehicle miles traveled: Evidence from 370 US urbanized areas, *Environment and Planning A*, **42** (2) 400–418.

Chatman, D. G. and R. B. Noland (2011) Do public transport improvements increase agglomeration economies? A review of literature and an agenda for research, *Transport Reviews*, **31**, 725–742.

Chen, H.-K. (1999) *Dynamic travel choice models: a variational inequality approach*, Springer-Verlag, Berlin-Heidelberg.

Chiabaut, N. and A. Barcet (2018) Demonstration and evaluation of intermittent bus lane strategy, paper presented at the *97th Annual Meeting of the Transportation Research Board*, Washington DC.

Chiabaut, N., X. Xie and L. Leclercq (2012) Road Capacity and Travel Times with Bus Lanes and Intermittent Priority Activation, *Transportation Research Record*, **2315**, 182–190.

Chowdhury, D., L. Santen and A. Schadschneider (2000) Statistical physics of vehicular traffic and some related systems, *Physcis Reports*, **329**, 199–329.

Ciccone, A. (2002) Agglomeration effects in Europe, *European Economic Review*, **46**, 213–227.

Çolak, S., A. Lima and M. C. González (2016) Understanding congested travel in urban areas, *nature Communications*, **7**, 10793.

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

Cramer, J. and A. B. Krueger (2016) Disruptive Change in the Taxi Business: The Case of Uber, *National Bureau of Economic Research*, **106** (5) 177–182.

Dafermos, S. (1980) Traffic Equilibrium and Variational Inequalities, *Transportation Science*, **14** (1) 42–54.

Dafermos, S. C. and F. T. Sparrow (1969) The traffic assignment problem for a general network, *Journal of Research of the National Bureau of Standards B*, **73** (2) 91–118.

Daganzo, C. F. (1977) On the traffic assignment problem with flow dependent costs—II, *Transportation Research*, **11** (6) 439–441.

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** (4) 269–287.

Daganzo, C. F. (1997) *Fundamentals of Transportation and Traffic Operations*, Pergamon, Oxford.

Daganzo, C. F. (2005) A variational formulation of kinematic waves: Basic theory and complex boundary conditions, *Transportation Research Part B: Methodological*, **39** (2) 187–196.

Daganzo, C. F. (2007) Urban gridlock: Macroscopic modeling and mitigation approaches, *Transportation Research Part B: Methodological*, **41** (1) 49–62.

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

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.

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.

Daganzo, C. F. and L. J. Lehe (2015) Distance-dependent congestion pricing for downtown zones, *Transportation Research Part B: Methodological*, **75**, 89–99.

Daganzo, C. F., L. J. Lehe and J. Argote-Cabanero (2018) Adaptive offsets for signalized streets, *Transportation Research Part B: Methodological*, **117**, 926–934.

Daganzo, C. F. and Y. Sheffi (1977) On Stochastic Models of Traffic Assignment, *Transportation Science*, **11** (3) 253–274.

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.

Dantsuji, T., D. Fukuda and N. Zheng (2017) A macroscopic approach for optimizing road space allocation of bus lanes in multimodal urban networks through simulation analysis: An application to the Tokyo CBD network, paper presented at the *IEEE Conference on Intelligent Transportation Systems (ITSC)*.

de Jong, G., J. Fox, A. Daly, M. Pieters and R. Smit (2004) A comparison of car ownership models, *Transport Reviews*, **24**, 379–408.

de Jong, G. C. and R. Kitamura (2009) A review of household dynamic vehicle ownership models: Holdings models versus transactions models, *Transportation*, **36** (6) 733–743.

de Palma, A. and R. Lindsey (2011) Traffic congestion pricing methodologies and technologies, *Transportation Research Part C: Emerging Technologies*, **19** (6) 1377–1399.

Del Castillo, J. M. and F. G. Benitez (1995a) On the functional form of the speed-density relationship - I: General theory, *Transportation Research Part B: Methodological*, **29** (5) 373–389.

Del Castillo, J. M. and F. G. Benitez (1995b) On the functional form of the speed-density relationship- II: Empirical investigation, *Transportation Research Part B: Methodological*, **29** (5) 391–406.

Department of Economic and Social Affairs of the United Nations Secretariat (2014) World Urbanization Prospects: The 2014 Revision, *Technical Report*, United Nations, New York.

Derrible, S. and C. Kennedy (2009) Characterizing metro networks: state, form, and structure, *Transportation*, **37** (2) 275–297.

Derrible, S. and C. Kennedy (2010) The complexity and robustness of metro networks, *Physica A: Statistical Mechanics and its Applications*, **389** (17) 3678–3691.

Derrible, S. and C. Kennedy (2011) Applications of Graph Theory and Network Science to Transit Network Design, *Transport Reviews*, **31** (4) 495–519.

Dirkse, S. P. and M. C. Ferris (1998) Modeling and solution environments for MPEC: GAMS & MATLAB, in M. Fukushima and L. Qi (eds.) *Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, 127–147, Springer, Boston.

Dixit, V. V. (2013) Behavioural foundations of two-fluid model for urban traffic, *Transportation Research Part C: Emerging Technologies*, **35**, 115–126.

Dönier, G., J. Bernhard, B. Stefan and L. Raymann (2013) Die adaptive Verkehrssteuerung, *Technical Report*, Dienstabteilung Verkehr, Stadt Zürich.

Downs, A. (1962) The law of peak-hour expressway congestion, *Traffic Quarterly*, **16** (3) 393–409.

Duranton, G. and D. Puga (2004) Micro-Foundations of Urban Agglomeration Economies, in J. Henderson and J.-F. Thisse (eds.) *Handbook of Regional and Urban Economics*, vol. 4, 2063–2117, Elsevier, Amsterdam.

Duranton, G. and M. A. Turner (2011) The Fundamental Law of Road Congestion: Evidence from US Cities, *American Economic Review*, **101** (6) 2616–2652.

Duranton, G. and M. A. Turner (2012) Urban Growth and Transportation, *The Review of Economic Studies*, **79** (4) 1407–1440.

Edie, L. (1963) Discussion of traffic stream measurements and definitions, in J. Almond (ed.) *Proceedings of the 2nd International Symposium on the Theory of Traffic Flow*, 139–154, OECD, Paris.

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.

Eliasson, J. (2009) A cost–benefit analysis of the Stockholm congestion charging system, *Transportation Research Part A: Policy and Practice*, **43** (4) 468–480.

Enquist, B. J., J. H. Brown and G. B. West (1998) Allometric scaling of plant energetics and population density, *Nature*, **395**, 163–165.

Erath, A. and K. W. Axhausen (2010) *Long Term Fuel Price Elasticity: Effects on Mobility Tool Ownership and Residential Location Choice*, Swiss Federal Office of Energy (SFOE), Federal Office for the Environment (FOEN), Bern.

Ewing, R. and R. Cervero (2001) Travel and the built environment: a synthesis, *Transportation Research Record*, **1780**, 87–114.

Ewing, R. and R. Cervero (2010) Travel and the Built Environment, *Journal of the American Planning Association*, **76** (3) 265–294.

Ewing, R. and R. Cervero (2017) "Does Compact Development Make People Drive Less?" The Answer Is Yes, *Journal of the American Planning Association*, **83** (1) 19–25.

Fagnant, D. J. and K. Kockelman (2015) Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations, *Transportation Research Part A: Policy and Practice*, **77**, 167–181.

Farag, S., M. Dijst and M. Lanzendorf (2003) Exploring the Use of E-Shopping and Its Impact on Personal Travel Behavior in the Netherlands, *Transportation Research Record*, **1858**, 47–54.

Farahani, R. Z., E. Miandoabchi, W. Y. Szeto and H. Rashidi (2013) A review of urban transportation network design problems, *European Journal of Operational Research*, **229** (2) 281–302.

Ferdous, N., A. R. Pinjari, C. R. Bhat and R. M. Pendyala (2010) A comprehensive analysis of household transportation expenditures relative to other goods and services : an application to United States consumer expenditure data, *Transportation*, **37** (3) 363–390.

Fernald, J. G. (1999) Roads to Prosperity? Assessing the Link Between Public Capital and Productivity, *American Economic Review*, **89** (3) 619–638.

Ferris, M. C., A. Meeraus and T. F. Rutherford (1999) Computing Wardropian equilibria in a complementarity framework, *Optimization Methods and Software*, **10** (5) 669–685.

Fisk, C. (1980) Some developments in equilibrium traffic assignment, *Transportation Research Part B: Methodological*, **14** (3) 243–255.

Fisk, C. (2000) Commentary: Road pricing and optimal transport networks, *Road and Transport Research*, **9** (1) 92–101.

Forschungsgesellschaft für Straßen- und Verkehrswesen e.V. (2009) *Handbuch für die Bemessung von Strassenverkehrsanlagen: HBS*, 3 edn., Forschungsgesellschaft für Strassen- und Verkehrswesen, Köln.

Fosgerau, M. (2015) Congestion in the bathtub, *Economics of Transportation*, **4** (4) 241–255.

Friedrich, M., K. Immisch, P. Jehlicka, T. Otterstätter and J. Schlaich (2010) Generating Origin-Destination Matrices from Mobile Phone Trajectories, *Transportation Research Record*, **2196**, 93–101.

Friesz, T. L. (1985) Transportation network equilibrium, design and aggregation: Key developments and research opportunities, *Transportation Research Part A: Policy and Practice*, **19** (5-6) 413–427.

Fujita, M., P. R. Krugman and A. J. Venables (2001) *The spatial economy: Cities, regions, and international trade*, MIT press, Cambridge, MA.

Fujita, M. and J.-F. Thisse (2002) *Economics of agglomeration: cities, industrial location, and regional growth*, Cambridge University Press, Cambridge.

Fujita, M. and J.-F. Thisse (2013) *Economics of agglomeration: cities, industrial location, and globalization*, 2 edn., Cambridge University Press, Cambridge.

Gao, X. and V. V. Gayah (2017) An analytical framework to model uncertainty in urban network dynamics using Macroscopic Fundamental Diagrams, *Transportation Research Procedia*, **23**, 497–516.

Garrow, M. and R. Machemehl (1999) Development and Evaluation of Transit Signal Priority Strategies, *Journal of Public Transportation*, **2** (2) 65–90.

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.

Geroliminis, N. and B. Boyaci (2012) The effect of variability of urban systems characteristics in the network capacity, *Transportation Research Part B: Methodological*, **46**, 1607–1623.

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., J. Haddad and M. Ramezani (2013) Optimal Perimeter Control for Two Urban Regions With Macroscopic Fundamental Diagrams: A Model Predictive Approach, *IEEE Transactions on Intelligent Transportation Systems*, **14** (1) 348–359.

Geroliminis, N. and D. M. Levinson (2009) Cordon Pricing Consistent with the Physics of Overcrowding, in W. H. K. Lam, S. C. Wong and H. K. Lo (eds.) *Transportation and Traffic Theory 2009: Golden Jubilee*, 219–240, Springer, Boston.

Geroliminis, N. and J. Sun (2011) Properties of a well-defined macroscopic fundamental diagram for urban traffic, *Transportation Research Part B: Methodological*, **45** (3) 605–617.

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.

Geurs, K. T., K. J. Krizek and A. Reggiani (2012) *Accessibility analysis and transport planning: challenges for Europe and North America*, Edward Elgar, Cheltenham.

Geurs, K. T. and B. van Wee (2004) Accessibility evaluation of land-use and transport strategies: review and research directions, *Journal of Transport Geography*, **12** (2) 127–140.

Glaeser, E. L. (1999) Learning in Cities, *Journal of Urban Economics*, **46** (2) 254–277.

Glaeser, E. L. (2008) *Cities, agglomeration, and spatial equilibrium*, Oxford University Press, Oxford.

Godfrey, J. W. (1969) The Mechanism of a Road Network, *Traffic Engineering and Control*, **11**, 323–327.

Goethe, J. W. v. (1808) *Faust: der Tragödie erster Teil*, Cotta'sche Verlagsbuchhandlung, Tübingen.

Gonzales, E. J. and C. F. Daganzo (2012) Morning commute with competing modes and distributed demand: User equilibrium, system optimum, and pricing, *Transportation Research Part B: Methodological*, **46** (10) 1519–1534.

González, M. C., C. A. Hidalgo and A.-L. Barabási (2008) Understanding individual human mobility patterns, *Nature*, **453**, 779–782.

Gonzalez-Navarro, M. and M. A. Turner (2018) Subways and urban growth: Evidence from earth, *Journal of Urban Economics*, **108**, 85–106.

Goodwin, P. B. (1992) A review of new demand elasticities with special reference to short and long run effects of price changes, *Journal of Transport Economics and Policy*, **26** (2) 155–169.

Goodwin, P. B. (1996) Empirical evidence on induced traffic a review and synthesis, *Transportation*, **23**, 35–54.

Graham, D. J. (2007a) Agglomeration, Productivity and Transport Investment, *Journal of Transport Economics and Policy*, **41** (3) 317–343.

Graham, D. J. (2007b) Variable returns to agglomeration and the effect of road traffic congestion, *Journal of Urban Economics*, **62** (1) 103–120.

Graham, D. J., S. Gibbons and R. Martin (2010) The spatial decay of agglomeration economies: estimates for use in transport appraisal, *Technical Report*, Imperial College London, London School of Economics, London.

Graham, D. J. and S. Glaister (2004) Road traffic demand elasticity estimates: A review, *Transport Reviews*, **24** (3) 261–274.

Greene, W. H. (2012) *Econometric analysis*, Pearson, Boston.

Greenshields, B. D. (1935) A study of highway capacity, *Highway Research Board Proceedings*, **14**, 448–477.

Guberinić, S., G. Šenborn and B. Lazić (2008) *Optimal traffic control: urban intersections*, Taylor & Francis, Boca Raton.

Guihaire, V. and J.-K. Hao (2008) Transit network design and scheduling: A global review, *Transportation Research Part A: Policy and Practice*, **42** (10) 1251–1273.

Guler, S. I., V. V. Gayah and M. Menendez (2016) Bus priority at signalized intersections with single-lane approaches: A novel pre-signal strategy, *Transportation Research Part C: Emerging Technologies*, **63**, 51–70.

Guler, S. I. and M. Menendez (2014) Analytical formulation and empirical evaluation of pre-signals for bus priority, *Transportation Research Part B: Methodological*, **64**, 41–53.

Hadas, Y. and A. Ceder (2018) Optimal Connected Urban Bus Network of Priority Lanes, *Transportation Research Record*, **2418**, 49–57.

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.

Haddad, J. and B. Mirkin (2017) Coordinated distributed adaptive perimeter control for large-scale urban road networks, *Transportation Research Part C: Emerging Technologies*, **77**, 495–515.

Hall, J. D., C. Palsson and J. Price (2018) Is Uber a substitute or complement for public transit?, *Journal of Urban Economics*, **108**, 36–50.

Hamilton, A., B. Waterson, T. Cherrett, A. Robinson and I. Snell (2013) The evolution of urban traffic control: changing policy and technology, *Transportation Planning and Technology*, **36**, 24–43.

Hamilton Ailsa, B. W. R. (1982) Wasteful Commuting, *Journal of Political Economy*, **90** (5) 1035–1053.

Handy, S. L. and D. A. Niemeier (1997) Measuring Accessibility: An Exploration of Issues and Alternatives, *Environment and Planning A*, **29** (7) 1175–1194.

Hanlon, W. W. and A. Miscio (2017) Agglomeration: A long-run panel data approach, *Journal of Urban Economics*, **99**, 1–14.

Hansen, M. and Y. Huang (1997) Road supply and traffic in California urban areas, *Transportation Research Part A: Policy and Practice1*, **31** (3) 205–218.

Hansen, W. G. (1959) How Accessibility Shapes Land Use, *Journal of the American Institute of Planners*, **25**, 73–76.

Harford, J. D. (2006) Congestion, pollution, and benefit-to-cost ratios of US public transit systems, *Transportation Research Part D: Transport and Environment*, **11** (1) 45–58.

He, H., S. I. Guler and M. Menendez (2016) Adaptive control algorithm to provide bus priority with a pre-signal, *Transportation Research Part C: Emerging Technologies*, **64**, 28–44.

Heckman, J. J. (1979) Sample Selection Bias as a Specification Error, *Econometrica*, **47**, 153–161.

Helbing, D. (2001) Traffic and related self-driven many-particle systems, *Reviews of Modern Physics*, **73** (4) 1067–1141.

Helbing, D. (2009) Derivation of a fundamental diagram for urban traffic flow, *The European Physical Journal B*, **70** (2) 229–241.

Henderson, J. (2003) Marshall's scale economies, *Journal of Urban Economics*, **53**, 1–28.

Herman, R. and S. Ardekani (1984) Characterizing Traffic Conditions in Urban Areas, *Transportation Science*, **18** (2) 101–140.

Herman, R. and I. Prigogine (1979) A two-fluid approach to town traffic, *Science*, **204**, 148–151.

Holl, A. (2016) Highways and productivity in manufacturing firms, *Journal of Urban Economics*, **93**, 131–151.

Horni, A., K. Nagel and K. W. Axhausen (2016) *The Multi-Agent Transport Simulation MATSim*, Ubiquity Press, London.

Hsu, W.-T. and H. Zhang (2014) The fundamental law of highway congestion revisited: Evidence from national expressways in Japan, *Journal of Urban Economics*, **81**, 65–76.

Hunt, P. B., D. I. Robertson, R. D. Bretherton and M. C. Royle (1982) The SCOOT on-line traffic signal optimization technique, *Traffic Engineering & Control*, **23** (4) 190–192.

Hymel, K. (2009) Does traffic congestion reduce employment growth?, *Journal of Urban Economics*, **65** (2) 127–135.

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

Jäggi, B., A. Erath, C. Dobler and K. W. Axhausen (2012) Modeling household fleet choice as a function of fuel price using a multiple discrete-continuous choice model, *Transportation Research Record*, **2302**, 174–183.

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

Ji, Y., J. Luo and N. Geroliminis (2014) Empirical observations of congestion propagation and dynamic partitioning with probe data for large-scale systems, *Transportation Research Record*, **2422**, 1–11.

JLL (2015) Switzerland Office Market – 2015.

Jolliffe, I. (2002) *Principal component analysis*, Springer, New York.

Kepaptsoglou, K. and M. Karlaftis (2009) Transit Route Network Design Problem: Review, *Journal of Transportation Engineering*, **135** (8) 491–505.

Kohla, B. and M. Fellendorf (2015) Urban traffic control and management, *Elektrotechnik und Informationstechnik*, **132** (7) 389–394.

Kouvelas, A., M. Saeedmanesh and N. Geroliminis (2017) Enhancing model-based feedback perimeter control with data-driven online adaptive optimization, *Transportation Research Part B: Methodological*, **96**, 26–45.

Kraus, M. (2012) Road pricing with optimal mass transit, *Journal of Urban Economics*, **72** (2-3) 81–86.

Krugman, P. (1991) *Geography and Trade*, MIT Press, Cambrdige, MA.

Kühnert, C., D. Helbing and G. B. West (2006) Scaling laws in urban supply networks, *Physica A: Statistical Mechanics and its Applications*, **363** (1) 96–103.

Kuzmyak, J., C. Baber and D. Savory (2006) Use of Walk Opportunities Index to Quantify Local Accessibility, *Transportation Research Record*, **1977**, 145–153.

Lämmer, S. and D. Helbing (2008) Self-control of traffic lights and vehicle flows in urban road networks, *Journal of Statistical Mechanics: Theory and Experiment*, **P04019**.

Lamotte, R. and N. Geroliminis (2018) The morning commute in urban areas with heterogeneous trip lengths, *Transportation Research Part B: Methodological*, **117**, 794–810.

Laval, J. A. and F. Castrillón (2015) Stochastic approximations for the macroscopic fundamental diagram of urban networks, *Transportation Research Part B: Methodological*, **81** (3) 904–916.

Laval, J. A., L. Leclercq and N. Chiabaut (2018) Minimal parameter formulations of the dynamic user equilibrium using macroscopic urban models: Freeway vs city streets revisited, *Transportation Research Part B: Methodological*, **117**, 676–686.

Le Vine, S., M. Lee-Gosselin, A. Sivakumar and J. Polak (2013) A new concept of accessibility to personal activities: development of theory and application to an empirical study of mobility resource holdings, *Journal of Transport Geography*, **31**, 1–10.

Leape, J. (2006) The London Congestion Charge, *Journal of Economic Perspectives*, **20**, 157–176.

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.

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

Lessing, G. E. (1779) Nathan der Weise, `http://www.digbib.org/Gotthold_Ephraim_Lessing_1729/Nathan_der_Weise_.pdf`.

Levinsohn, J. and E. E. Leamer (1995) International trade theory: The evidence, *Handbook of International Economics*, **3**, 1341–1375.

Levinson, D. (2012) Network structure and city size, *PLoS One*, **7** (1) e29721.

Levinson, D. and R. Karamalaputi (2003) Induced supply - A model of highway network expansion at the microscopic level, *Journal of Transport Economics and Policy*, **37** (3) 297–318.

Li, D., B. Fu, Y. Wang, G. Lu, Y. Berezin, H. E. Stanley and S. Havlin (2015) Percolation transition in dynamical traffic network with evolving critical bottlenecks, *Proceedings of the National Academy of Sciences*, **112** (3) 669–672.

Lighthill, M. J. and G. B. Whitham (1955) On Kinematic Waves. II. A Theory of Traffic Flow on Long Crowded Roads, *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **229** (1178) 317–345.

Litman, T. (2012) Transit Price Elasticities and Cross-Elasticities, *Journal of Public Transportation*, **7** (2) 37–58.

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

Loder, A. and K. W. Axhausen (2018) Mobility tools and use: Accessibility's role in Switzerland, *Journal of Transport and Land Use*, **11** (1) 367–385.

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

Loo, B. P. Y. and A. H. T. Cheng (2010) Are there useful yardsticks of population size and income level for building metro systems? Some worldwide evidence, *Cities*, **27** (5) 299–306.

Lopez, C., L. Leclercq, P. Krishnakumari, N. Chiabaut and H. van Lint (2017) Revealing the day-to-day regularity of urban congestion patterns with 3D speed maps, *Scientific Reports*, **7**, 14029.

Louf, R. and M. Barthelemy (2014) How congestion shapes cities: from mobility patterns to scaling, *Scientific Reports*, **4**, 5561.

Louf, R., C. Roth and M. Barthelemy (2014) Scaling in transportation networks, *PLoS ONE*, **9** (7) 1–8.

Lu, S., V. L. Knoop and M. Keyvan-Ekbatani (2018) Using taxi GPS data for macroscopic traffic monitoring in large scale urban networks: calibration and MFD derivation, *Transportation Research Procedia*, **34**, 243–250.

Luo, Z.-Q., J.-S. Pang and D. Ralph (1996) *Mathematical programs with equilibrium constraints*, Cambridge University Press, Cambridge.

Madre, J.-L., K. W. Axhausen and W. Brög (2007) Immobility in travel diary surveys, *Transportation*, **34** (1) 107–128.

Magnanti, T. L. and R. T. Wong (1984) Network Design and Transportation Planning: Models and Algorithms, *Transportation Science*, **18** (1) 1–55.

Mahmassani, H., J. C. Williams and R. Herman (1987) Performance of urban traffic networks, paper presented at the *oth International Symposium on Transportation and Traffic Theory*.

Maniccam, S. (2003) Traffic jamming on hexagonal lattice, *Physica A: Statistical Mechanics and its Applications*, **321** (3) 653–664.

Marcotte, P. (1986) Network design problem with congestion effects: A case of bilevel programming, *Mathematical Programming*, **34** (2) 142–162.

Mariotte, G., L. Leclercq and J. A. Laval (2017) Macroscopic urban dynamics: Analytical and numerical comparisons of existing models, *Transportation Research Part B: Methodological*, **101**, 245–267.

Marshall, A. (1920) *Principles of Economics*, Macmillan, London.

Mazloumian, A., N. Geroliminis and D. Helbing (2010) The spatial variability of vehicle densities as determinant of urban network capacity, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **368**, 4627–4647.

Melo, P. C., D. J. Graham, D. Levinson and S. Aarabi (2017) Agglomeration, accessibility and productivity: Evidence for large metropolitan areas in the US, *Urban Studies*, **54** (1) 179–195.

Melo, P. C., D. J. Graham and R. B. Noland (2009) A meta-analysis of estimates of urban agglomeration economies, *Regional Science and Urban Economics*, **39** (3) 332–342.

Mendes, G. A., K. W. Axhausen, J. S. Andrade and H. J. Herrmann (2014) A scenario planning approach for disasters on Swiss road network, *International Journal of Modern Physics C*, **25** (11) 1450067.

Metz, D. (2008) The Myth of Travel Time Saving, *Transport Reviews*, **28**, 321–336.

Migdalas, A. (1995) Bilevel programming in traffic planning: Models, methods and challenge, *Journal of Global Optimization*, **7** (4) 381–405.

Miller, A. J. (1970) Letter to the Editor—The Amount of Traffic Which Can Enter a City Center During Peak Periods, *Transportation Science*, **4** (4) 409–411.

Mogridge, M. J. H. (1990) *Travel in towns: jam yesterday, jam today and jam tomorrow?*, Macmillan, London.

Mogridge, M. J. H. (1997) The self-defeating nature of urban road capacity policy, *Transport Policy*, **4**, 5–23.

Mokhtarian, P. L., I. Salomon and S. L. Handy (2006) The impacts of ict on leisure activities and travel: A conceptual exploration, *Transportation*, **33** (3) 263–289.

Morris, J. M., P. L. Dumble and M. R. Wigan (1979) Accessibility Indicators for Transport Planning, *Transportation Research Part A: Policy and Practice*, **13** (2) 91–109.

Muhlich, N., V. V. Gayah and M. Menendez (2015) An examination of MFD hysteresis patterns for hierarchical urban street networks Using microsimulation, *Transportation Research Record*, **2491**, 117–126.

Muñoz, J. C. and C. F. Daganzo (2002) Moving Bottlenecks: A Theory Grounded on Experimental Observation, in M. A. Taylor (ed.) *Transportation and Traffic Theory in the 21st Century*, 441–461, Emerald Group Publishing Limited.

Nagai, R., T. Nagatani and N. Taniguchi (2005) Traffic states and jamming transitions induced by a bus in two-lane traffic flow, *Physica A: Statistical Mechanics and its Applications*, **350** (2-4) 548–562.

Nagatani, T. (1993) Jamming transition in the traffic-flow model with two-level crossings, *Physical Review E*, **48** (5) 3290–3294.

Nagatani, T. (2002) The physics of traffic jams, *Reports on Progress in Physics*, **65** (9) 1331.

Nagel, K. and M. Schreckenberg (1992) A Cellular Automaton Model for Freeway Traffic, *Journal De Physique I*, **2** (12) 2221–2229.

Nagurney, A. (1993) *Network economics: a variational inequality approach*, Kluwer Academic Publishers, Boston.

Newman, P. W. G. and J. R. Kenworthy (1991) Transport and urban form in thirty-two of the world's principal cities, *Transport Reviews*, **11** (3) 249–272.

Nguyen-Phuoc, D. Q., G. Currie, C. De Gruyter and W. Young (2018) Exploring the impact of public transport strikes on travel behavior and traffic congestion, *International Journal of Sustainable Transportation*, **12** (8) 613–623.

Olmos, L. E., S. Çolak, S. Shafiei, M. Saberi and M. C. González (2018) Macroscopic dynamics and the collapse of urban traffic, *Proceedings of the National Academy of Sciences*, **115** (50) 12654–12661.

Olszewski, P., H. S. Fan and Y. W. Tan (1995) Area-wide traffic speed-flow model for the Singapore CBD, *Transportation Research Part A: Policy and Practice*, **29** (4) 273–281.

Ortigosa, J. and M. Menendez (2014) Traffic performance on quasi-grid urban structures, *Cities*, **36**, 18–27.

Ortúzar, J. d. D. and L. G. Willumsen (2011) *Modelling transport*, Wiley-Blackwell, Chichester.

Owen, A. and D. M. Levinson (2015) Modeling the commute mode share of transit using continuous accessibility to jobs, *Transportation Research Part A: Policy and Practice*, **74**, 110–122.

Paleti, R., C. Bhat and R. Pendyala (2013) Integrated Model of Residential Location, Work Location, Vehicle Ownership, and Commute Tour Characteristics, *Transportation Research Record*, **2382**, 162–172.

Pang, J.-S. and P. T. Harker (1990) Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications, *Mathematical Programming*, **48** (1) 161–220.

Papageorgiou, M., C. Diakaki, V. Dinopoulou, A. Kotsialos and Y. Wang (2008) Review of road traffic control strategies, *Proceedings of the IEEE*, **91** (12) 2043–2067.

Parry, I. W. H. (2009) Pricing Urban Congestion, *Annual Review of Resource Economics*, **1** (1) 461–484.

Parry, I. W. H. and K. A. Small (2009) Should Urban Transit Subsidies Be Reduced?, *American Economic Review*, **99** (3) 700–724.

Patriksson, M. (1994) *The Traffic Assignment Problems: Models and Methods*, Topics in Transportation, VSP, Utrecht.

Patz, A. (1925) Die richtige Auswahl der Verkehrslinien bei grossen Strassenbahnnetzen, *Verkehrstechnik*, **50/51**, 977–983.

Paul, B., S. Mitra and B. Maitra (2016) Calibration of Robertson's Platoon Dispersion Model in Non-lane Based Mixed Traffic Operation, *Transportation in Developing Economies*, **2** (11).

Pigou, A. C. (1920) *The economics of welfare*, Macmillan, London.

Prager, W. (1954) Problems of traffic and transportation, *Proceedings of the Symposium on Operations Research in Business and Industry*, 105–113.

Prud'homme, R. and J. P. Bocarejo (2005) The London congestion charge: a tentative economic appraisal, *Transport Policy*, **12**, 279–287.

Ramezani, M., J. Haddad and N. Geroliminis (2015) Dynamics of heterogeneity in urban networks: aggregated traffic modeling and hierarchical control, *Transportation Research Part B: Methodological*, **74**, 1–19.

Redding, S. J. and D. M. Sturm (2008) The costs of remoteness: Evidence from German division and reunification, *American Economic Review*, **98** (5) 1766–1797.

Redmond, L. S. and P. L. Mokhtarian (2001) The positive utility of the commute: Modeling ideal commute time and relative desired commute amount, *Transportation*, **28** (2) 179–205.

Ricardo, D. (1817) *On the Principles of Political Economy and Taxation*, John Murray.

Robertson, D. I. (1969) A Traffic Network Study Tool, *TRRL Laboratory Report*, **253**.

Robinson, S. M. (1980) Strongly Regular Generalized Equations, *Mathematics of Operations Research*, **5** (1) 43–62.

Rokicki, B. and M. Stępniak (2018) Major transport infrastructure investment and regional economic development – An accessibility-based approach, *Journal of Transport Geography*, **72**, 36–49.

Rosenthal, S. S. and W. C. Strange (2001) The Determinants of Agglomeration, *Journal of Urban Economics*, **89**, 311–316.

Rosenthal, S. S. and W. C. Strange (2004) Evidence on the Nature and Sources of Agglomeration Economies, in J. Henderson and J.-F. Thisse (eds.) *Handbook of Regional and Urban Economics*, vol. 4, 2119–2171, Elsevier, Amsterdam.

Rotaris, L., R. Danielis, E. Marcucci and J. Massiani (2010) The urban road pricing scheme to curb pollution in Milan, Italy: Description, impacts and preliminary cost–benefit analysis assessment, *Transportation Research Part A: Policy and Practice*, **44** (5) 359–375.

Roth, C., S. M. Kang, M. Batty and M. Barthelemy (2012) A long-time limit for world subway networks, *Journal of the Royal Society Interface*, **9** (75) 2540–2550.

Rutherford, T. F. (1995) Extension of GAMS for complementarity problems arising in applied economic analysis, *Journal of Economic Dynamics and Control*, **19** (8) 1299–1324.

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.

Saeedmanesh, M. and N. Geroliminis (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** (2) 137–148.

Samaniego, H. and M. E. Moses (2008) Cities as Organisms: Allometric Scaling of Urban Road Networks, *Journal of Transport and Land Use*, **1** (1) 21–39.

Santos, G. and B. Shaffer (2004) Preliminary Results of the London Congestion Charging Scheme, *Public Works Management & Policy*, **9** (2) 164–181.

Sarlas, G. and K. W. Axhausen (2019) Commuting distance and individual accessibility, *Presented at 98th Annual Meeting of the Transportation Research Board, Washington, D.C.*

Sarlas, G., R. Fuhrer and K. W. Axhausen (2015) Quantifying the agglomeration effects of Swiss public transport between 2000 and 2010, paper presented at the *15th Swiss Transport Research Conference*, Ascona.

Sasaki, M. and T. Nagatani (2003) Transition and saturation of traffic flow controlled by traffic lights, *Physica A: Statistical Mechanics and its Applications*, **325** (3-4) 531–546.

Schaar (1925) Die Leistungsfähigkeit einer Strasse für den Kraftverkehr, *Verkehrstechnik*, **23**.

Schafer, A. and D. G. Victor (2000) The future mobility of the world population, *Transportation Research Part A: Policy and Practice*, **34**, 171–205.

Schéele, S. (1980) A supply model for public transit services, *Transportation Research Part B: Methodological*, **14** (1-2) 133–146.

Schläpfer, M., L. M. Bettencourt, S. Grauwin, M. Raschke, R. Claxton, Z. Smoreda, G. B. West and C. Ratti (2014) The scaling of human interactions with city size, *Journal of the Royal Society Interface*, **11** (98) 20130789.

Schlich, R. and K. W. Axhausen (2003) Habitual travel behaviour: Evidence from a six-week travel diary, *Transportation*, **30** (1) 13–36.

Schmid, B. (2019) Connecting Time-Use, Travel and Shopping Behavior: Results of a Multi-Stage Household Survey, Ph.D. Thesis, ETH Zurich.

Schmid, B. and K. W. Axhausen (2018) In-store or online shopping of search and experience goods: A hybrid choice approach, *Journal of Choice Modelling*, **in press**.

Schönfelder, S. and K. W. Axhausen (2003) Activity spaces: measures of social exclusion?, *Transport Policy*, **10** (4) 273–286, 10 2003.

Schreiber, A., A. Loder and K. W. Axhausen (2016) Urban mode and subscription choice- An application of the three-dimensional MFD, paper presented at the *16th Swiss Transport Research Conference*.

Scott, D. M. and K. W. Axhausen (2006) Household mobility tool ownership: modeling interactions between cars and season tickets, *Transportation*, **33**, 311–328.

Seddon, P. A. (1972) The prediction of platoon dispersion in the combination methods of linking traffic signals, *Transportation Research*, **6** (2) 125–130.

Shen, Q. (2000) Spatial and social dimensions of Commuting, *Journal of the American Planning Association*, **66**, 68–83.

Sherman, R. (1971) Congestion interdependence and urban transit fares, *Econometrica*, **39** (3) 565–576.

Simma, A. and K. W. Axhausen (2001) Structures of commitment in mode use: a comparison of Switzerland, Germany and Great Britain, *Transport Policy*, **8**, 279–288.

Small, K. A. and E. T. Verhoef (2007) *The economics of urban transportation*, Routledge, London.

Smeed, R. J. (1961) *The Traffic Problem in Towns*, Statistical Society, Manchester.

Smeed, R. J. (1966) Road capacity of city centers, *Traffic Engineering and Control*, **8** (7) 455–458.

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

Smith, A. (1776) *An Inquiry into the Nature and Causes of the Wealth of Nations*, W. Strahan and T. Cadell, London.

Smith, H. R., B. Hemily and M. Ivanovic (2005) Transit signal priority (TSP): A planning and implementation handbook, *Technical Report*, ITS America, Washington DC.

Solow, R. M. (1972) Congestion, Density and the Use of Land in Transportation, *The Swedish Journal of Economics*, **74** (1) 161–173.

Song, C., Z. Qu, N. Blumm and A. L. Barabasi (2010) Limits of predictability in human mobility, *Science*, **327**, 1018–1021.

Sonntag, H. (1977) Linienplanung im öffentlichen Personennahverkehr, Ph.D. Thesis, Technische Universität Berlin.

Spiess, H. (1990) Technical Note — Conical Volume-Delay Functions, *Transportation Science*, **24** (2) 153–158.

Stadt Zürich (2015) Loop detector data in Zurich, *Technical Report*, Dienstabteilung Verkehr (DAV), Zürich.

Stadt Zürich (2016) Fahrzeiten der VBZ im Soll-Ist-Vergleich, *Technical Report*, Verkehrsbetriebe Zürich, Zürich.

Stadt Zürich (2017) Private Communication.

Stevanovic, A. (2010) *Adaptive traffic control systems : domestic and foreign state of practice*, vol. 403 of *NCHRP Synthesis*, Transportation Research Board, Washington, D.C.

Stopher, P. R. (2004) Reducing road congestion: A reality check, *Transport Policy*, **11** (2) 117–131.

Swiss Federal Office of Spatial Development (2011) *ÖV-Güteklassen - Berechnungsmethodik ARE*, Bundesamt für Raumenticklung ARE, Neuchâtel.

Swiss Federal Office of Spatial Development, Swiss Federal Office of Environment and Swiss Federal Statistical Office (2011) *Landschaftstypologie Schweiz*, Bundesamt für Raumenticklung ARE, Bundesamt für Umwelt BAFU, Bundesamt für Statistik BFS.

Swiss Federal Statistical Office and Swiss Federal Office for Spatial Development (2012) *Mobilität in der Schweiz - Ergebnisse des Mikrozensus Mobilität und Verkehr 2010*, Bundesamt für Statistik BFS, Bundesamt für Raumentwicklung ARE, Neuchâtel.

Swiss Federal Statistical Office and Swiss Federal Office for Spatial Development (2017) *Verkehrsverhalten der Bevölkerung- Ergebnisse des Mikrozensus Mobilität und Verkehr 2015*, Bundesamt für Statistik BFS, Bundesamt für Raumentwicklung ARE, Neuchâtel.

Tabuchi, T. (1998) Urban Agglomeration and Dispersion: A Synthesis of Alonso and Krugman, *Journal of Urban Economics*, **44** (3) 333–351.

Tanner, R. and D. Bolduc (2014) The Multiple Discrete-continuous Extreme Value Model (MDCEV) with Fixed Costs, *Procedia - Social and Behavioral Sciences*, **111**, 390–399.

Thøgersen, J. (2009) Promoting public transport as a subscription service: Effects of a free month travel card, *Transport Policy*, **16** (6) 335–343.

Thomson, J. M. (1967) Speed and flows of traffic in Central London: 2. Speed-flow relations, *Traffic Engineering and Control*, **8** (12) 721–725.

Thünen, J. v. (1826) *Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*, Perthes, Hamburg.

Tirachini, A. and D. A. Hensher (2012) Multimodal Transport Pricing: First Best, Second Best and Extensions to Non-motorized Transport, *Transport Reviews*, **32**, 181–202.

Tirachini, A., D. A. Hensher and J. M. Rose (2014) Multimodal pricing and optimal design of urban public transport: The interplay between traffic congestion and bus crowding, *Transportation Research Part B: Methodological*, **61**, 33–54.

Tomasello, M. (2009) *Why we cooperate*, MIT Press, Cambridge, MA.

Train, K. E. (2009) *Discrete choice methods with simulation*, Cambridge University Press, Cambridge.

Transportation Research Board (2011) 75 Years of the Fundamental Diagram for Traffic Flow Theory: Greenshields Symposium.

Transportation Research Board (2016) *Highway Capacity Manual*, 6 edn., Transportation Research Board, Washington, D.C.

Treiber, M. and A. Kesting (2013) *Traffic Flow Dynamics*, Springer-Verlag, Berlin Heidelberg.

Tsekeris, T. and N. Geroliminis (2013) City size, network structure and traffic congestion, *Journal of Urban Economics*, **76**, 1–14.

Underwood, R. T. (1961) Speed, Volume and density relationships, in *Quality and Theory of Traffic Flow*, 141–188, Pennsylvania State University, University Park.

van Nieuwkoop, R. H. (2014) Transportation Networks and Economic Equilibrium Modeling Issues and Applications, Ph.D. Thesis, ETH Zürich.

Venables, A. J. (2007) Evaluating urban transport improvements, *Journal of Transport Economics and Policy*, **41**, 173–188.

Venables, A. J. (2017a) Breaking into tradables: Urban form and urban function in a developing city, *Journal of Urban Economics*, **98**, 88–97.

Venables, A. J. (2017b) Expanding cities and connecting cities: appraising the effects of transport improvements, *Journal of Transport Economics and Policy*, **51** (1) 1–19.

Vickerman, R. (2008) Transit investment and economic development, *Research in Transportation Economics*, **23** (1) 107–115.

Vickerman, R., K. Spiekermann and M. Wegener (1999) Accessibility and economic development in Europe, *Regional Studies*, **33** (1) 1–15.

Vickrey, W. S. (1969) Congestion theory and transportation investment, *The American Economic Review*, **59**, 251–260.

Wang, J., Y. Mao, J. Li, Z. Xiong and W. X. Wang (2015) Predictability of road traffic and congestion in urban areas, *PLoS One*, **10** (4) e0121825.

Wang, P., T. Hunter, A. M. Bayen, K. Schechtner and M. C. González (2012) Understanding road usage patterns in urban areas, *Scientific Reports*, **2**, 1001.

Wardrop, J. G. (1952) Some Theoretical Aspects of Road Traffic Research, *Proceedings of the Institution of Civil Engineers*, **Part II**, 325–378.

Wardrop, J. G. (1968) Journey speed and flow in central urban areas, *Traffic Engineering & Control*, **9**, 528–532.

Webster, F. V. (1958) Traffic signal settings, *Road Research Laboratory*, **39**.

Weis, C. and K. W. Axhausen (2009) Induced travel demand: Evidence from a pseudo panel data based structural equations model, *Research in Transportation Economics*, **25** (1) 8–18.

Weiss, D. J., A. Nelson, H. S. Gibson, W. Temperley, S. Peedell, A. Lieber, M. Hancher, E. Poyart, S. Belchior, N. Fullman, B. Mappin, U. Dalrymple, J. Rozier, T. C. D. Lucas, R. E. Howes, L. S. Tusting, S. Y. Kang, E. Cameron, D. Bisanzio, K. E. Battle, S. Bhatt and P. W. Gething (2018)

A global map of travel time to cities to assess inequalities in accessibility in 2015, *Nature*, **553** (7688) 333–336.

West, G. B., J. H. Brown and B. J. Enquist (1997) A General Model for the Origin of Allometric Scaling Laws in Biology, *Science*, **276**, 122–126.

Wheaton, W. C. (1998) Land Use and Density in Cities with Congestion, *Journal of Urban Economics*, **43** (2) 258–272.

Wheaton, W. C. (2004) Commuting, congestion, and employment dispersal in cities with mixed land use, *Journal of Urban Economics*, **55** (3) 417–438.

Wierbos, M., V. L. Knoop, F. Hänseler and S. P. Hoogendoorn (2018) A Macroscopic Flow Model for Mixed Bicycle-Car Traffic, paper presented at the *7th Symposium of the European Association for Research in Transportation (hEART 2018)*, Athens.

Williams, J. C. (2001) Macroscopic Flow Models, in N. Gartner, C. Messer and A. Rathi (eds.) *Traffic flow Theory: A state-of-the-Art Report*, 1–31, Transportation Research Board, Washington DC.

Williams, J. C., H. S. Mahmassani, S. Iani and R. Herman (1987) Urban Traffic Network Flow Models, *Transportation Research Record*, **1112**, 78–88.

Willumsen, L. G. (1978) Estimation of an O-D Matrix from Traffic Counts ? A Review., *Technical Report*, Institute of Transport Studies, Leeds, UK.

Wong, W. and S. C. Wong (2016) Network topological effects on the macroscopic Bureau of Public Roads function, *Transportmetrica A: Transport Science*, **12** (3) 272–296.

Wu, J. J., Z. Y. Gao, H. J. Sun and H. J. Huang (2006) Congestion in different topologies of traffic networks, *Europhysics Letters*, **74** (3) 560–566.

Yamamoto, T. (2009) Comparative analysis of household car, motorcycle and bicycle ownership between Osaka metropolitan area, Japan and Kuala Lumpur, Malaysia, *Transportation*, **36**, 351–366.

Yang, H. and M. G. H. Bell (1998) A capacity paradox in network design and how to avoid it, *Trransportation Research Part A: Policy and Practice*, **32** (7) 539–545.

Yang, J., S. Chen, P. Qin, F. Lu and A. A. Liu (2018a) The effect of subway expansions on vehicle congestion: Evidence from Beijing, *Journal of Environmental Economics and Management*, **88**, 114–133.

Yang, K., N. Zheng and M. Menendez (2018b) Multi-scale perimeter control approach in a connected-vehicle environment, *Transportation Research Part C: Emerging Technologies*, **94**, 32–49.

Yildirimoglu, M. and N. Geroliminis (2014) Approximating dynamic equilibrium conditions with macroscopic fundamental diagrams, *Transportation Research Part B: Methodological*, **70**, 186–200.

Yu, B., L. Kong, Y. Sun, B. Yao and Z. Gao (2015) A bi-level programming for bus lane network design, *Transportation Research Part C: Emerging Technologies*, **55**, 310–327.

Zahavi, Y. (1972) Traffic performance evaluation of road networks by the α-relationship, *Traffic Engineering and Control*, **14** (5-6) 228–231.

Zhang, F., N. Zheng, H. Yang and N. Geroliminis (2018) A systematic analysis of multimodal transport systems with road space distribution and responsive bus service, *Transportation Research Part C: Emerging Technologies*, **96**, 208–230.

Zhang, W. and K. M. Kockelman (2016) Optimal policies in cities with congestion and agglomeration externalities: Congestion tolls, labor subsidies, and place-based strategies, *Journal of Urban Economics*, **95**, 64–86.

Zhao, L., Y. C. Lai, K. Park and N. Ye (2005) Onset of traffic congestion in complex networks, *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, **71** (2) 26125.

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*, **2651**, 42–51.

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

Zheng, N., R. A. Waraich, N. Geroliminis and K. W. Axhausen (2012) A dynamic cordon pricing scheme combining a macroscopic and an agent-based traffic model, *Transportation Research Part A: Policy and Practice*, **46**, 1291–1303.