Macroscopic Tools for Monitoring, Modeling, Design, and Optimization of Multi-Modal Urban Transportation Systems

Doctoral Thesis

Author(s):
Dakic, Igor

Publication date:
2021

Permanent link:
https://doi.org/10.3929/ethz-b-000546096

Rights / license:
In Copyright - Non-Commercial Use Permitted

Funding acknowledgement:
165644 - DIPLOMAT – Design, Optimization and control of urban Multimodal systems (SNF)
MACROSCOPIC TOOLS FOR MONITORING, MODELING, DESIGN, AND OPTIMIZATION OF MULTI-MODAL URBAN TRANSPORTATION SYSTEMS

A dissertation submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by
IGOR DAKIC
MSc., Florida Atlantic University
MSc., University of Belgrade
BSc., University of Belgrade

citizen of Serbia

accepted on the recommendation of
Prof. Dr. Monica Menendez, examiner
Prof. Dr. Kay W. Axhausen, examiner
Prof. Dr. Hani Mahmassani, co-examiner
Prof. Dr. Nigel Wilson, co-examiner
Prof. Dr. Nikolas Geroliminis, co-examiner

2021
As traffic congestion grows in cities across the globe, more emphasis is being placed on developing multi-modal transportation systems with the hope of adopting more sustainable practices. The work of Mr. Igor Dakic explores such multi-modal systems, in particular, those concerning car traffic and public transportation, from multiple perspectives. This is valuable, as public and private transportation typically share and compete for the same limited road space. Thus, improvements to one mode normally come at the expense of the other one. Therefore, to really improve the system, it makes sense to do it from a holistic perspective, taking into account not only the performance of the individual modes, but also the interactions between them.

To that end, Mr. Dakic has borrowed concepts from different disciplines, including traffic flow theory, queueing theory, simulation, and optimization. He has also developed novel methodologies that build on the three-dimensional macroscopic fundamental diagram (3D-MFD), linking the accumulation of each vehicle type (e.g. cars and buses) to the overall throughput of the network. Moreover, he has evaluated such throughput in terms of vehicles and in terms of passengers. This is especially important as buses have a much higher capacity than cars, and can carry many more passengers in a smaller space. Overall, this dissertation offers a holistic approach, proposing new tools for the monitoring, modeling, design, and optimization of multi-modal transportation systems.

On behalf of the Traffic Engineering research group at the Swiss Federal Institute of Technology, Zurich, I thank Mr. Dakic for his enthusiasm, perseverance, and attention to detail; as well as for his contributions to advance our knowledge on multi-modal transportation.

Prof. Dr. Monica Menendez

Former Director of the Traffic Engineering research group at ETH Zurich
Currently Associate Professor of Civil Engineering at New York University Abu Dhabi (NYUAD) / Global Network Associate Professor of Civil and Urban Engineering at New York University (NYU) / Director of Research Center for Interacting Urban Networks (CITIES)
ABSTRACT

Multi-modal urban cities are complex systems where different modes of transport compete for and share the limited road infrastructure to serve transport demand. These modes interact with each other, creating conflicts at different road infrastructure levels. The goal of this dissertation is to develop novel monitoring, modeling, design, and optimization tools to better understand and manage the overall performance of such complex multi-modal transportation systems, with a special emphasis on the interactions between the private and public modes of transport. To this end, we apply the three-dimensional Macroscopic Fundamental Diagram (3D-MFD), recognized as a powerful framework to investigate traffic dynamics of multi-modal urban cities and develop network-wide control strategies.

In particular, we make several methodological contributions. First, we propose novel estimation methods for the space-mean speed of cars based on: (i) the automatic vehicle location (AVL) data of public transport where no floating car data (FCD) is available; and (ii) the fused FCD and AVL data sources where both are available, but FCD is not complete. By combining the estimated car space-mean speed with the flow measurements for car traffic, acquired either from loop detector data (LDD) or other kinds of sensors such as video cameras, the proposed estimation methods allow to macroscopically monitor multi-modal urban systems with a limited amount of available data.

Second, we propose a methodology to analytically approximate the 3D-MFD for bi-modal urban corridors while accounting for the stochastic nature of bus operations. The proposed framework extends the existing variational theory (VT) approaches by: (i) introducing a probabilistic VT graph, where the costs are computed using an efficient stochastic shortest path algorithm; (ii) capturing the effects of stochastic moving bus bottlenecks and the correlation of bus arrival times; (iii) incorporating a macroscopic passenger model that reflects the passenger dynamics for the different modes; and (iv) accounting for the effects that the traffic conditions might have on bus operations. As such, the proposed method allows to macroscopically model multi-modal urban systems.

Third, we provide a general framework for the bus network design problem that considers multiple trip length patterns, two types of user behavior, and the effects that the bus network structure might have on the traffic...
performance and passenger mode choice. This framework can be used to **macroscopically design multi-modal urban systems**. To capture complex modal interactions and quantify the operating speeds, we apply the 3D-MFD. We use the operating speed for each mode to determine the mode choice at the trip length level. This way, we are able to solve the optimal bus network design problem under the free-flow/saturated traffic conditions in an analytical way, while considering more realistic settings including a dynamic description of the peak hour demand, mixed traffic, and different mode choice decisions depending on trip lengths and walking preferences.

Finally, we propose a novel concept, called flexible bus dispatching system, which offers new perspectives and enormous flexibility to better manage the dispatching frequencies and the allocation of the vehicle resources, reducing thereby the operating cost. In such a flexible bus dispatching system, the bus fleet consists not only of conventional buses, but also of modular and fully automated bus units that can either operate individually or combined together (forming thereby a single modular unit of a higher passenger capacity). As such, the proposed approach allows to **macroscopically optimize multi-modal urban systems**. To determine the optimal number of combined modular bus units and the optimal frequency at which the units (both conventional and modular) should be dispatched across different bus lines, while accounting for the traffic dynamics at the network level, we propose an optimization framework based on the 3D-MFD. To the best of our knowledge, this is the first application of the 3D-MFD and modular bus units for the frequency setting problem in the domain of bus operations.
Multimodale Städte sind komplexe Systeme, in denen verschiedene Verkehrsträger um die begrenzte Straßeninfrastruktur konkurrieren und sich diese teilen, um die Verkehrsnachfrage zu bedienen. Diese Verkehrsträger interagieren miteinander, was zu Konflikten auf verschiedenen Ebenen der Straßeninfrastruktur führt. Ziel dieser Dissertation ist es, neuartige Überwachungs-, Modellierungs-, Entwurfs- und Optimierungswerkzeuge zu entwickeln, um die Gesamtleistung solcher komplexen multimodalaren Verkehrssysteme besser zu verstehen und zu steuern, wobei der Schwerpunkt auf den Wechselwirkungen zwischen privaten und öffentlichen Verkehrsträgern liegt. Zu diesem Zweck wenden wir das dreidimensionale makroskopische Fundamentaldiagramm (3D-MFD) an, das als leistungsfähiger Rahmen für die Untersuchung der Verkehrsdynamik multimodaler urbaner Städte und die Entwicklung netzwerkweiter Steuerungsstrategien anerkannt ist.

Wir leisten insbesondere mehrere methodische Beiträge. Erstens schlagen wir neuartige Schätzverfahren für die mittlere Geschwindigkeit von Fahrzeugen im Raum vor, die auf (i) den automatischen Fahrzeugortungsdaten (AVL) des öffentlichen Verkehrs und (ii) den fusionierten FCD- und AVL-Daten basieren: (i) den automatischen Fahrzeugortungsdaten (AVL) des öffentlichen Verkehrs, wenn keine Floating Car Data (FCD) verfügbar sind; und (ii) den fusionierten FCD- und AVL-Datenquellen, wenn beide verfügbar sind, die FCD aber nicht vollständig sind. Durch die Kombination der geschätzten mittleren Geschwindigkeit des Pkw-Verkehrs mit den Verkehrsflussmessungen für den Pkw-Verkehr, die entweder aus Schleifen-Detektordaten (LDD) oder anderen Arten von Sensoren wie Videokameras gewonnen werden, ermöglichen die vorgeschlagenen Schätzverfahren die makroskopische Überwachung multimodaler städtischer Systeme mit einer begrenzten Menge an verfügbaren Daten.

Zweitens schlagen wir eine Methode zur analytischen Annäherung der 3D-MFD für bimodale städtische Korridore vor, wobei die stochastische Natur des Busbetriebs berücksichtigt wird. Der vorgeschlagene Rahmen erweitert die bestehenden Ansätze der Variations-Theorie (VT) durch: (i) Einführung eines probabilistischen VT-Graphen, bei dem die Kosten mit Hilfe eines effizienten stochastischen Algorithmus für den kürzesten Weg berechnet werden; (ii) Erfassung der Auswirkungen von stochastischen,
sich bewegenden Busengpässen und der Korrelation von Busankunftszei-
ten; (iii) Einbeziehung eines makroskopischen Fahrgastmodells, das die
Fahrgastdynamik für die verschiedenen Verkehrsträger widerspiegelt; und
(iv) Berücksichtigung der Auswirkungen, die die Verkehrsbedingungen
auf den Busbetrieb haben könnten. Somit ermöglicht die vorgeschlage-
e Prof. Dr. Michael M. Müller
ne Methode die **makroskopische Modellierung multimodaler städtischer
Systeme**.

Drittens stellen wir einen allgemeinen Rahmen für das Problem der Bus-
netzgestaltung zur Verfügung, der mehrere Reiselängenmuster, zwei Ar-
ten von Nutzerverhalten und die Auswirkungen berücksichtigt, die die
Busnetzstruktur auf die Verkehrsleistung und die Verkehrsmittelwahl der
Fahrgäste haben könnte. Dieser Rahmen kann für die **makroskopische Ge-
staltung multimodaler städtischer Systeme** verwendet werden. Um kom-
plexe modale Interaktionen zu erfassen und die Betriebsgeschwindigkei-
ten zu quantifizieren, wenden wir die 3D-MFD an. Wir verwenden die
Betriebsgeschwindigkeit für jeden Verkehrsträger, um die Verkehrsmittel-
wahl auf der Ebene der Reiselänge zu bestimmen. Auf diese Weise sind
wir in der Lage, das Problem der optimalen Gestaltung des Busnetzes bei
freiem/gesättigtem Verkehr auf analytische Weise zu lösen und gleichzei-
tig realistischere Einstellungen zu berücksichtigen, einschließlich einer
dynamischen Beschreibung der Nachfrage in der Spitzenstunde, gemischten
Verkehrs und unterschiedlicher Verkehrsmittelwahlentscheidungen in Ab-
hängigkeit von der Reiselänge und den Gehpräferenzen.

Schließlich schlagen wir ein neuartiges Konzept vor, das als flexibles Bus-
Dispatching-System bezeichnet wird und neue Perspektiven und enorme
Flexibilität bietet, um die Dispatching-Frequenzen und die Zuweisung der
Fahrzeugressourcen besser zu verwalten und dadurch die Betriebskosten
tszen zu senken. In einem solchen flexiblen Bus-Dispatching-System besteht die
Busflotte nicht nur aus konventionellen Bussen, sondern auch aus modu-
laren und vollautomatischen Buseinheiten, die entweder einzeln oder ge-
mensam betrieben werden können (und so eine einzige modulare Einheit
mit einer höheren Fahrgastkapazität bilden). Der vorgeschlagene Ansatz
ermöglicht somit eine **makroskopische Optimierung multimodaler städ-
tischer Systeme**. Um die optimale Anzahl kombinierter modularer Busein-
heiten und die optimale Häufigkeit zu bestimmen, mit der die Einheiten
(sowohl konventionelle als auch modulare) auf verschiedenen Buslinien
gesetzt werden sollten, während gleichzeitig die Verkehrsdynamik auf
Netzebene berücksichtigt wird, schlagen wir einen Optimierungsrahmen
auf der Grundlage der 3D-MFD vor. Nach unserem Kenntnisstand ist dies

viii
die erste Anwendung der 3D-MFD und modularer Buseinheiten für das Problem der Frequenzeinstellung im Bereich des Busbetriebs.
ACKNOWLEDGEMENTS

Writing acknowledgments is perhaps the most fulfilling part of completing my dissertation. It symbolizes the end of my PhD, and allows me to express my gratitude to all people who have given me support on this exciting and challenging journey.

First and foremost, I would like to thank my supervisor, Prof. Monica Menendez, who recognized my potential and gave me the opportunity to be part of her amazing team. I am extremely grateful for her wise guidance in shaping my research direction and developing new ideas, as well as for exposing me to constant challenges that contributed to my professional and personal development. Thank you Monica for being inspiring scientist, supportive mentor, and a great friend! It was a real privilege working with you!

I am also grateful to my official supervisor, Prof. Kay W. Axhausen, and external examiners, Prof. Hani Mahmassani, Prof. Nigel Wilson, and Prof. Nikolas Geroliminis, for serving on my doctoral examination committee, reviewing my dissertation in detail, and attending my defense. Their insightful comments and constructive suggestions led to fruitful discussions and have significantly improved this dissertation.

Furthermore, I would like to express my sincere gratitude to my dear collaborators and friends, Prof. Ludovic Leclercq and Prof. Jorge Laval, who hosted me at the University of Lyon and Georgia Institute of Technology, respectively. Their rigorous mathematical approach, enthusiasm for traffic flow theory, vision, and creativity were truly inspiring and have undoubtedly shaped my knowledge. Thank you Ludo and Jorge for sharing your extensive knowledge and creating a very inspiring environment during my research stays!

Being a member of the Institute for Transport Planning and Systems (IVT) team was a great pleasure. I will always cherish the moments of good times we spent together, inspiring conversations during coffee breaks and conferences, and fruitful collaborations. Special thanks go to Lukas Ambühl, Kaidi Yang, Mireia Roca-Riu, Javier Ortigosa, Qiao Ge, Haitao He, Jin Cao, Adrian Meister, Allister Loder, Sergio Guidon, Felix Becker, Anastasios Kouvelas, Kimia Chavoshi, Alexander Genser, Zahra Ghandeharioun, Francesco Corman, Beda Büchel, and Peter Lorch. Equally, I thank Prof. Aleksandar Stevanovic and Nikola Mitrovic for a very exciting collab-
oration during my PhD, as well as to all CITIES and LICIT lab members for creating a very friendly atmosphere and making my research stays extremely pleasant.

This journey would not have been possible without the great help and support of my friends. Special thanks to Nikola Vasic, Dejan Mijailovic, Filip Mijailovic, Dusko Bekcic, Alessio Guffanti, and Nicolo De Rita. Thank you guys for cheering me up in difficult moments! Sharing my passion for sports, outdoors, politics, languages, and life with you is priceless.

Moreover, I want to thank my family, my father, brother, and sister, for their continuous support not only in pursuing my PhD, but in everything I do. I am extremely grateful to have you, and I would like to thank you for always putting a smile on my face and being there for me!

Last, but certainly not least, I would like to dedicate this dissertation to my mother, whose efforts, time, and patience made me the person I am today. Thank you for everything you have done for me! I will always keep you in my heart. May the God give your soul eternal peace.
CONTENTS

List of Figures xvi

List of Tables xviii

1 INTRODUCTION 1
   1.1 Research motivation ........................................ 1
   1.2 Research objectives and scope ............................. 4
       1.2.1 Objectives .......................................... 4
       1.2.2 Scope ............................................... 5
   1.3 Research contributions .................................... 5
   1.4 Dissertation outline ..................................... 8

2 LITERATURE REVIEW 11
   2.1 Macroscopic fundamental relationships ..................... 11
   2.2 Macroscopic monitoring of urban systems ................. 12
   2.3 Macroscopic modeling of urban systems .................... 14
   2.4 Macroscopic design of urban systems ..................... 15
   2.5 Macroscopic optimization of urban systems ............... 17

3 TOOL FOR MONITORING MULTI-MODAL SYSTEMS 19
   3.1 General methodology ...................................... 20
       3.1.1 Eulerian method .................................... 20
       3.1.2 Lagrangian method based on the FCD (Probes method) 22
       3.1.3 Lagrangian method based on the AVL data ........... 22
       3.1.4 Lagrangian method based on the fused FCD and AVL data ................................................. 27
       3.1.5 Determining the accuracy of the estimated space-mean speed of cars ................................... 30
       3.1.6 Evaluating the performance of the proposed fusion method ................................................. 31
   3.2 Case studies ............................................... 32
       3.2.1 Abstract network .................................... 32
       3.2.2 City of Zurich network ............................... 34
   3.3 Results ...................................................... 35
       3.3.1 Simulation outputs ................................... 35
### 3.3.2 Empirical analysis ............................................ 41

### 3.4 Summary .......................................................... 46

### 4 TOOL FOR MODELING MULTI-MODAL SYSTEMS ............. 49

#### 4.1 General methodology ........................................ 50

- 4.1.1 Uni-modal MFD in a deterministic environment .......... 50
- 4.1.2 Multi-modal p-MFD in a stochastic environment .......... 53
- 4.1.3 Traffic conditions and their effects on bus stop shortcuts 68

#### 4.2 Implementation and discussion .............................. 70

- 4.2.1 Empirical implementation and validation ................ 70
- 4.2.2 Value of the proposed analytical framework ............. 75
- 4.2.3 Application example: Evaluation of different bus lane layouts .................................................. 78

#### 4.3 Summary .......................................................... 79

### 5 TOOL FOR THE DESIGN OF MULTI-MODAL SYSTEMS ....... 81

#### 5.1 General methodology ........................................ 82

- 5.1.1 General network settings ................................. 82
- 5.1.2 Mathematical formulation of the objective function .... 88
- 5.1.3 Derivation of the maximum bus occupancy ............. 94
- 5.1.4 Derivation of the operating car speed for each lane allocation type ............................................. 96

#### 5.2 Numerical results ............................................... 101

- 5.2.1 Solution algorithm ........................................ 101
- 5.2.2 Case study and analyzed scenarios ....................... 102
- 5.2.3 General results and the effects of demand intensity ... 102
- 5.2.4 Value of the proposed methodological extensions ..... 107

#### 5.3 Summary .......................................................... 113

### 6 TOOL FOR OPTIMIZING MULTI-MODAL SYSTEMS .......... 115

#### 6.1 General methodology ........................................ 116

- 6.1.1 System description and network representation .......... 116
- 6.1.2 Mathematical formulation of the optimization framework .......................................................... 121
- 6.1.3 Modeling vehicular dynamics ............................. 126
- 6.1.4 Modeling passenger dynamics ............................. 130

#### 6.2 Numerical experiments and results ......................... 133

- 6.2.1 Case study and simulation scenarios ..................... 133
- 6.2.2 Value of considering modular vehicle technology ...... 135
6.2.3 Value of using the 3D-MFD to capture complex system dynamics ............................................. 140
6.2.4 Effects of the operating unit cost of modular bus units .............................................................. 143
6.2.5 Effects of the size of the bus network and the bus fleet ............................................................. 144
6.3 Summary ................................................................. 145

7 CONCLUSIONS AND OUTLOOK .................................................. 149
7.1 Summary of contributions and findings ......................................................................................... 149
7.2 Outlook ........................................................................................................................................ 153
7.3 Concluding remarks ..................................................................................................................... 155

A APPENDIX ............................................................................. 157

BIBLIOGRAPHY ................................................................. 165

Curriculum Vitae ................................................................. 179

List of publications ............................................................... 181
LIST OF FIGURES

Figure 2.1 Illustration of (a) the vehicular 3D-MFD and (b) the passenger 3D-MFD for the city center of Zurich, Switzerland. ........................................... 12

Figure 3.1 Studied networks: a) Snapshot of the simulation network layout; b) Zones of interest in the city of Zurich - note that the map also shows residential streets, which are not included in this study. ............ 33

Figure 3.2 Comparison of the AVL-based estimation methods for the space-mean speed of cars, for a one-lane network with: a) 40% mixed-lanes; b) 15% mixed-lanes; and a two-lane network with: c) 37.5% mixed-lanes; d) 12.5% mixed-lanes. .................................................. 36

Figure 3.3 Average error on the space-mean speed of cars when $\varphi_{AVL} = 37.5\%$, $f_b = 7.5$ min, and the spatial distribution of probe vehicles is: a) Homogeneous; b) Inhomogeneous. .................................................. 38

Figure 3.4 Average error on the space-mean speed of cars when $\varphi_{AVL} = 12.5\%$, $f_b = 7.5$ min, and the spatial distribution of probe vehicles is: a) Homogeneous; b) Inhomogeneous. .................................................. 38

Figure 3.5 Contour plot of the range of improvements obtained by the proposed AVL-based estimation method in comparison to the FCD-based method, when the spatial distribution of probe vehicles is: a) homogeneous; b) inhomogeneous. ................................. 39

Figure 3.6 Macroscopic fundamental relationship for the car mode between: a) flow and density; b) speed and density. .............................................................. 42

Figure 3.7 Empirical validation of the proposed AVL-based Lagrangian method for an average: a) working day; b) weekend day. .................................................. 42

Figure 3.8 Network coverages of FCD and AVL data in the city of Zurich, for the analyzed weekdays and time periods. .................................................. 44
Figure 3.9 Average error on the space-mean speed of cars in case of the limited FCD network coverage (5%, 10%, and 15%). ................................................................. 45
Figure 4.1 Modeling the bottleneck probability in case of: (a) bus stop shortcuts; (b) moving bus shortcuts. .............. 57
Figure 4.2 Schematic representation of the study site for empirical analysis. ................................................................. 71
Figure 4.3 Comparison of the estimated and empirical multi-modal MFD for a real corridor in the City of Zurich, Switzerland. ................................................................. 73
Figure 4.4 Differences between the stochastic and the deterministic VT approaches. ................................................................. 76
Figure 4.5 Comparison of the p-MFDs from a real corridor with a mixed or dedicated bus lane. ......................... 78
Figure 5.1 City structure and the corresponding bus network. 84
Figure 5.2 A city and its vertical and horizontal cordons. 95
Figure 5.3 Computing the average car delay experienced along a corridor with allocation \(a\) in direction \(p\). 100
Figure 5.4 Schematic representation of the solution algorithm. 102
Figure 5.5 Concentration of user origins and destinations, and the corresponding trip length pattern in case of: (a-b) uniform city; (c-d) a mono-centric city; (e-f) a commuter city; and (g-h) a twin city. 104
Figure 5.6 Mode choice distribution for the optimal bus network design in scenario SC-I in case of: (a) uniform city; (b) a mono-centric city; (c) a commuter city; and (d) a twin city. 107
Figure 6.1 Schematic representation of the types of bus units that form the bus fleet in the proposed flexible bus dispatching system. ................................................................. 118
Figure 6.2 Schematic representation of the multi-modal reservoir model (a) used to model traffic dynamics based on the 3D-MFD (b). ................................................................. 120
Figure 6.3 Schematic illustration of the studied network, where solid lines represent dedicated bus segments, whereas dashed lines represent mixed bus segments. ......................... 133
Figure 6.4 Simulation settings: (a) car demand profile; and (b) public transport passenger demand profile. 134
Figure 6.5  Comparison of the improvement in the total system cost for different penetration rates of modular units. 136
Figure 6.6  Evolution of the average number of units contained in a modular bus for different penetration rates of modular units: (a) before the peak; (b) during the peak; and (c) after the peak. 137
Figure 6.7  Comparison of the improvement in system delay for different penetration rates of modular units. 138
Figure 6.8  Comparison of the improvement in the operator cost for different penetration rates of modular units. 138
Figure 6.9  Spatial-temporal distribution of the accumulation of modular buses, conventional buses, and modular bus units for different bus lines. 139
Figure 6.10 Comparison of the improvement in the system cost made by accounting for the complex multimodal interactions and congestion propagation for different penetration rates of modular units. 141
Figure 6.11 Evolution of the vehicle accumulation of both modes in the 3D-MFD plane for the tested traffic scenarios and two penetration rates of modular bus units ($p_m = 0$ and $p_m = 0.1$). 142
Figure A.1  Regions defining potential locations of an origin/destination within a rectangle defined by sides $l_x$ and $l_y$ in case: (a) $l_x \geq l_y$; and (b) $l_x < l_y$. 158

**LIST OF TABLES**

<p>| Table 3.1 | Nomenclature. .................................................. 21 |
| Table 3.2 | Comparison between the fusion and the reference method for: $\phi^{avl} = {10%, 5%}; f_b = 15$ min. .................. 40 |
| Table 4.1 | Nomenclature. .................................................. 51 |
| Table 5.1 | Nomenclature. .................................................. 83 |
| Table 5.2 | Experimental settings. ........................................ 103 |
| Table 5.3 | Optimal design parameters, operator cost, user cost, and the corresponding mode choice. .................. 106 |
| Table 5.4 | Effects of user behavior on the cost functions. ........... 108 |</p>
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.5</td>
<td>Effects of trip length patterns on the cost functions.</td>
<td>109</td>
</tr>
<tr>
<td>5.6</td>
<td>Effects of modal interactions on the cost functions.</td>
<td>111</td>
</tr>
<tr>
<td>5.7</td>
<td>Effects of heterogeneous design parameters across cardinal directions on the cost functions.</td>
<td>113</td>
</tr>
<tr>
<td>6.1</td>
<td>Nomenclature.</td>
<td>117</td>
</tr>
<tr>
<td>6.2</td>
<td>Effects of the operating unit cost of modular bus units on the cost functions ($p_m = 0.1$).</td>
<td>144</td>
</tr>
<tr>
<td>6.3</td>
<td>Effects of the size of the bus network and the bus fleet on the cost functions.</td>
<td>146</td>
</tr>
</tbody>
</table>
INTRODUCTION

Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less.

— Marie Curie

1.1 RESEARCH MOTIVATION

In the past few decades, urbanization has become a prevalent trend for many places around the globe. As cities continue to grow, they require substantial infrastructure resources, placing various burdens to society and economic development. To address the social, economic, and environmental challenges associated with rapid urbanization, it is crucial to provide a sustainable transportation system. This is because transportation connects people and goods for social and economic interactions (1).

Multi-modality is considered to be a crucial aspect of a sustainable urban development and mobility management. If not managed well, traffic congestion will be increasingly pervasive in urban areas. Constructing new infrastructure is an expensive solution, as the cost needed to keep pace with an increase in demand, including the induced demand, is very high (2). Therefore, special attention should be given to the design, management, and operation of the existing road facilities for different modes. This is, however, challenging given that these modes interact with each other, creating conflicts at different road infrastructure levels. While there is a vast and well-established literature towards understanding and mitigating vehicular congestion in single-mode networks, the knowledge on multi-modal networks with passenger mobility consideration is in its infancy.

To provide an efficient multi-modal transportation system, it is necessary to develop the right tools for the following four pillars of urban networks: monitoring, modeling, design, and optimization. Monitoring tools allow us to infer the current traffic congestion level and determine whether we should perform certain actions. Modeling tools help us to understand the complex interactions between different modes, and how a given action would affect both the traffic performance and the level of interactions. Design tools provide us with the optimal configuration of the public transport
system that yields a level of service competitive to that of the automobile at a reasonable cost. This can be very beneficial to both, the policy makers and the practitioners, when constructing new or redesigning existing public transport networks. Finally, optimization tools give us the optimal set of actions that maximizes the utilization of the existing road infrastructure when it is not possible to (re)design urban networks in order to keep pace with the everlasting increase in travel demand. The development of these tools is, however, challenging due to the complexity of the system dynamics and the lack of methods for understanding multi-modal traffic performance under different network structures and traffic flow operations.

With the emerging concepts related to the macroscopic modeling of large-scale networks, we can now study urban traffic dynamics in a parsimonious way. These concepts are mostly based on the Macroscopic Fundamental Diagram (MFD) \((3, 4)\), also known as the Network Fundamental Diagram \((5)\). The MFD has been shown to be a useful and elegant tool to determine the current state of traffic and quantify the capacity of an urban transportation network. It links the vehicle accumulation and the travel production with a well-defined and reproducible curve. Although initially the MFDs were studied within the context of uni-modal (i.e. car) traffic, recent research efforts have resulted in the development of a three-dimensional MFD \((3D\text{-MFD})\), applicable to multi-modal traffic \((6–9)\). These works demonstrated that, in contrast to the vehicular \(3D\text{-MFD}\), the passenger \(3D\text{-MFD}\) exhibits the maximum throughput for a non-zero accumulation of buses, revealing the impact of higher bus occupancy. As such, the \(3D\text{-MFD}\) offers new ways to model the complex system dynamics of urban networks while capturing the interactions between different transportation modes, in particular those between buses and cars \((10)\).

Despite the potential benefits of applying the \(3D\text{-MFD}\) for the aforementioned four pillars, i.e. monitoring, modeling, design, and optimization of multi-modal urban systems at the network level, there are still challenges in the utilization of such macroscopic approach, as summarized below.

- To empirically estimate the \(3D\text{-MFD}\) and use it for monitoring purposes, data needs to be acquired for different modes. In most urban cities, public transport vehicles are equipped with GPS devices \((11, 12)\) that can be used to determine the ground truth traffic variables (flow, speed, and density) for the public mode of transport. For the private mode, however, data is very limited or non-existent. Therefore, it is worth to investigate the potential for using the trajectory data from public transport to infer the state of car traffic (e.g. speed)
when no other data source exists. Such potential has not yet been explored.

- While there exist analytical recipes for uni-modal MFDs (13–17), there is no such formulation for the 3D-MFD that can be used for modeling purposes. Uni-modal MFDs can be modeled with the variational formulation of kinematic wave theory (18, 19). However, due to the complex and stochastic nature of multi-modal traffic, variational theory-based derived MFDs have not yet been extended to account for multiple modes at the network level.

- Current studies on the design of multi-modal urban systems do not consider the effects of public transport network structure on private car users, the level of interactions between the modes, and the passenger mode choice that depends on traffic conditions. The 3D-MFD offers a great potential to capture all these dependencies for the public transport network design problem. Unfortunately, we have not yet seen such application of the 3D-MFD.

- Since public transport allows more passengers to efficiently travel across an urban area (20), maximizing the efficiency of multi-modal urban systems is usually achieved by optimizing the frequency and allocation of public transport vehicles, in particular buses. To ensure that the determined bus frequency and vehicle allocation actually lead to an efficient bus service, it is crucial to account for the impact of traffic conditions and the interactions between the modes, especially in case of mixed traffic. This can be addressed with the aid of the 3D-MFD. However, the potential to apply the 3D-MFD for the purpose of capturing the dynamics of traffic congestion and complex modal interactions for the frequency setting problem has not yet been investigated.

The solution to the aforementioned challenges still remains as open research questions. In this dissertation, we address these questions by proposing novel macroscopic tools to monitor, model, design, and optimize multi-modal urban transportation systems based on the 3D-MFD.
1.2 RESEARCH OBJECTIVES AND SCOPE

1.2.1 Objectives

The current dissertation aims to accomplish the following:

1. Develop novel estimation methods for the average speed of cars at the network level based on: (i) the automatic vehicle location (AVL) data from public transport where no floating car data (FCD) is available; and (ii) the fused FCD and AVL data sources where both are available, but FCD is not complete. By combining the estimated car space-mean speed with the flow measurements for car traffic, acquired either from loop detector data (LDD) or other kinds of sensors such as video cameras, the proposed estimation methods allow to macroscopically monitor multi-modal urban systems with a limited amount of available data.

2. Design a methodology to analytically approximate the 3D-MFD while accounting for the stochastic nature of bus operations, passenger dynamics, correlation of bus arrival times, effects that the traffic conditions might have on bus operations, and complex interactions between the modes. Such analytical framework can be used to macroscopically model multi-modal urban systems.

3. Provide a general framework for the bus network design problem that considers multiple trip length patterns, two types of user behavior, and the effects that the bus network structure might have on the traffic performance and passenger mode choice. This framework can be used to macroscopically design multi-modal urban systems.

4. Propose a novel flexible bus dispatching system in which a fleet of fully automated modular bus units together with conventional buses serves the passenger demand. Since it explicitly accounts for the dynamics of traffic congestion and complex interactions between the modes at the network level, the proposed approach can be used to macroscopically optimize multi-modal urban systems.
1.2.2 Scope

The scope of the current thesis is limited to the following points:

- We consider multi-modal urban systems consisting of buses and cars only due to their high level of interactions. We disregard the effects and impacts of other public modes of transport and bicycles, as they usually operate along dedicated infrastructure.

- We focus on the supply side of multi-modal systems, and thus assume that the traffic demand and the routing information of bus passengers are exogenously given. For the frequency setting problem, we do not consider a dynamic change in the route choice or departure time as a function of the dispatching policy. This is because we aim to develop short-term dispatching strategies to improve the overall operator and passenger performance, rather than making long-term planning decisions that might affect the demand as in the bus network design problem.

- We apply the 3D-MFD as an aggregated model for multi-modal systems and use it to investigate the effects of network topology and configuration, road space allocation, and public transport operations in the context of multi-modal traffic performance. Consequently, we study monitoring, modeling, design, and optimization of multi-modal urban networks only from a macroscopic perspective.

1.3 Research Contributions

The main contributions of this dissertation are in the following four applications of the 3D-MFD:

1. **Monitoring multi-modal systems from the macroscopic perspective with a limited amount of available data.**
   
   (1.1) We derive a new Lagrangian method for estimating the space-mean speed of cars using the AVL data from public transport where no FCD is available.
   
   (1.2) We derive a new Lagrangian method for improving the accuracy of the estimated space-mean speed of cars by fusing the FCD and AVL data sources where both are available.
   
   (1.3) We use empirical data for demonstrating the applicability and validating the proposed methods in real-life traffic scenarios.
(1.4) By combining the estimated car space-mean speed with the LDD, we cross-compare different 3D-MFD estimation methods using both simulation and empirical data. Such empirical comparison is, to the best of our knowledge, the first of its kind.

2. **Modeling multi-modal systems from the macroscopic perspective while incorporating stochasticity in bus operations.**

(2.1) We propose a new methodology to account for, and include the stochastic behavior of buses into the variational theory (VT) framework. Such methodology introduces a probabilistic VT graph, where the costs are computed using an efficient stochastic shortest path algorithm.

(2.2) We capture the effects of stochastic moving bus bottlenecks and the correlation of bus arrival times.

(2.3) We extend VT to incorporate passenger dynamics, modeled as a function of the passenger demand and bus operations, and further derive a multi-modal passenger MFD.

(2.4) We further extend VT to account for the impact of the traffic conditions on both stochastic bus operations and passenger dynamics.

(2.5) We introduce an innovative application example for the evaluation of different bus lane layouts, aiming to maximize the passenger throughput along a multi-modal urban corridor.

3. **Designing multi-modal systems from the macroscopic perspective while taking into account the complex modal interactions.**

(3.1) We use the distribution of the user trip lengths as an intermediate level of abstraction to determine the optimal design parameters. Such a level of abstraction allows not only to account for different trip length patterns (per and across cardinal directions) for the same level of passenger demand, but also to capture more accurately the network loading for all modes.

(3.2) We consider the complex modal interactions and passenger mode choice, thus include the travel costs for both, the bus and the car mode, into the objective function. This way, we optimize the performance of the whole network while taking into account multiple transport modes.
(3.3) We account for spatially non-uniform network topology and distribution of the passenger demand across cardinal directions.

(3.4) We consider heterogeneous design parameters across cardinal directions, incorporating different lane allocations into the bus network design problem and accounting for their different influences on car traffic.

(3.5) We consider the influence of the passenger demand and network topology on the traffic performance, which, in turn, affects the passenger mode choice (i.e. demand). In other words, we use a full feedback loop to model all aforementioned dependencies.

(3.6) We investigate the effects of demand intensity, user behavior, and trip length patterns on the optimal bus network design and passenger mode choice.

4. Optimizing the performance of urban systems from the macroscopic perspective while capturing multi-modal traffic dynamics.

(4.1) We propose a flexible bus dispatching system that allows to dynamically combine and split modular and fully automated bus units for the purpose of improving the operation of the bus system.

(4.2) We develop an optimization framework that jointly determines the optimal number of combined modular bus units and the optimal frequency at which the units, both conventional and modular, should be dispatched across different bus lines. The proposed optimization model accounts for the effects of the dispatching policy on the operation and travel cost of both modes.

(4.3) We apply the 3D-MFD concept for the frequency setting problem to capture the complex multi-modal interactions and congestion propagation, accounting for both vehicle and passenger dynamics. As such, the proposed modeling framework can also be applied to the existing bus dispatching system consisting only of conventional buses, addressing the limitations of the current state of the art regarding the modeling of traffic dynamics.

(4.4) We study the effect of the operating unit cost of modular bus units, the size of the bus network, and the size of the bus fleet, testing the robustness of the proposed optimization framework.
Following the objectives defined above, this dissertation consists of four main parts, each related to a specific application of the 3D-MFD. A short overview of each chapter is given below.

**Chapter 2** provides a comprehensive review of the most important research studies on the network-level monitoring, modeling, design, and optimization of multi-modal urban systems using macroscopic tools.

**Chapter 3** presents novel Lagrangian methods for estimating the space-mean speed of cars in multi-modal urban networks: (i) the AVL-based method, which uses only the data provided by public transport vehicles where no FCD is available, and which accounts for the network configuration layout and the configuration of the public transport system; and (ii) the fusion algorithm, which combines the FCD and AVL data sources where both are available, and which does not require a priori knowledge about FCD penetration rates. The estimated car space-mean speed is then combined with the LDD and further used to estimate the 3D-MFD for monitoring the overall network performance.

**Chapter 4** extends the estimation analysis of Chapter 3 and proposes a methodology to analytically approximate the 3D-MFD for multi-modal urban corridors while accounting for the stochastic nature of bus operations. The proposed framework is used as an efficient modeling tool to identify a proper lane allocation strategy along a corridor, showing its pragmatic significance for practitioners.

**Chapter 5** uses the modeling framework of Chapter 4 to capture complex modal interactions for the bus network design problem and quantify the operating speed for each mode. These operating speeds are used to determine the mode choice at the trip length level. This allows to investigate the effects of demand intensity, user behavior, and trip length patterns on the optimal bus network configuration and passenger mode choice in an analytical way.

**Chapter 6** builds upon the design analysis of Chapter 5 and discusses the subsequent phase of the public transport network planning process, the frequency setting problem. In particular, it develops an optimization model used to determine the optimal composition of modular bus units and the optimal service frequency at which the buses (both conventional and modular) should be dispatched across each bus line.

**Chapter 7** concludes this dissertation, highlighting the main findings and providing an outlook for future research.
This dissertation is based on the following refereed archival journal articles and conference contributions by the author. All publications below are original work and first authored by the doctoral candidate.

**Journal articles**


**Conference proceedings**


I don’t care that they stole my idea... I care that they don’t have any of their own.

— Nikola Tesla

Over the last decade, the macroscopic fundamental relationships have gained enormous research interest in the context of monitoring, modeling, design, and optimization of multi-modal systems at the network level. In this chapter, we provide a comprehensive overview of the current state of the art related to the aforementioned applications of the 3D-MFD.

2.1 MACROSCOPIC FUNDAMENTAL RELATIONSHIPS

In traffic flow theory, various concepts have been proposed to macroscopically model the dynamics of urban networks under steady state traffic conditions (21–23). Following these ideas, (3) suggested a macroscopic relationship between the total outflow from the system and the aggregated accumulation, commonly known today as the MFD. Later, using empirical data from the city of Yokohama, (4) confirmed the existence of the MFD.

Scientific literature that followed this direction showed that the MFD only applies to a homogeneous urban region (24) and that it is not very sensitive to slight changes in demand (25). The first property implies that, in order to apply the MFD concept to a heterogeneous network, such network should be partitioned into multiple regions, each exhibiting homogeneous traffic conditions (26, 27). The second property can be very useful for control purposes, as it allows one to establish an efficient network-level control scheme without a detailed knowledge of origin-destination patterns (28–32).

Taking into account that, in most urban networks, multiple modes often compete for, and share the limited road infrastructure, the explanatory power of a single-mode traffic modeling might not be sufficient. Not only because the interactions between different modes are neglected, but also because by just looking at the vehicle throughput, the actual goal of improving the passenger mobility cannot be achieved, as not all modes carry the same number of passengers (33).
While uni-modal MFDs have been extensively studied during the past decade, multi-modal MFDs have only received some attention very recently. (34) were among the first to model traffic for a multi-modal network in Nairobi, Kenya, showing that a similar pattern as for the classical MFD can also be observed for bi-modal traffic. Nevertheless, the influence of competing modes on each other, as well as on the global traffic performance was not investigated. This was addressed by (35) and (6), who extended the concept of a single-mode MFD to the 3D-MFD through the means of simulation. The existence of the 3D-MFD was later proven empirically in (7). It was demonstrated that the passenger 3D-MFD (Fig. 2.1b) shows a completely different pattern from the vehicular 3D-MFD (Fig. 2.1a), highlighting the importance of modeling passenger traffic dynamics for multi-modal systems.

![Figure 2.1: Illustration of (a) the vehicular 3D-MFD and (b) the passenger 3D-MFD for the city center of Zurich, Switzerland.](image)

2.2 MACROSCOPIC MONITORING OF URBAN SYSTEMS

The (3D-)MFD has been recognized as a powerful framework to monitor urban transportation systems due to its ability to quickly determine the current traffic congestion level. In most cases, the macroscopic traffic variables are estimated using either LDD or FCD. If none of these data sources is available, traffic information might also come from some other sources, such as adaptive traffic control systems that report density-like measures, which can be potentially used for constructing the macroscopic fundamental relationships (36).
Recently, it has been shown that the incomplete LDD can be successfully used to inform an MFD perimeter control \((37, 38)\). However, both studies assumed that the detector measurements are not impacted by its position. This, unfortunately, cannot be generally expected in real transport networks as pointed out by \((39)\), and empirically shown by \((27)\). As a matter of fact, \((40)\) have concluded that, for an accurate density estimation using LDD, loop detectors should be uniformly positioned across the streets and cover as many different traffic situations as possible (e.g. some detectors should be placed close to the stop line, some in the middle of the street, and some far from traffic signals). Optimal location of fixed measurement points for estimating the MFD was also studied in \((41)\).

Similarly, based on the assumption that GPS equipped probe vehicles are homogeneously distributed across the network, \((42)\) proposed an MFD estimation method using the FCD. Such assumption, however, is not realistic, as, in most cases, the market penetration level of probes is not uniform across the entire set of origin-destination (OD) pairs. To account for unequally distributed probes across the network, \((43)\) derived an MFD from FCD, using a new probe penetration estimation method based on the \(k\)-means clustering analysis. Although the findings are encouraging, the implementation of the proposed method is difficult due to the following reasons. First, the detector location can have a great impact on the estimated number of probes for a given OD pair, which has not been addressed in the study. Second, the optimal number of clusters, shown to have a significant impact on the accuracy of the results, can be network-specific, making the generalization of the procedure very challenging.

It has also been shown that by fusing LDD and FCD for both flow and density we can acquire substantial improvements in the accuracy of an estimated MFD \((44)\). However, this fusion algorithm needs to have a priori knowledge about the penetration rate of probe vehicles, and assumes the macroscopic traffic variables (flow and density) are properly measured by loop detectors. Considering that the rate of FCD is not very often (if ever) available, and that, in most cases, loop detectors cannot report the exact flow and density values, practical implementation of such a fusion procedure remains challenging.

In an effort to address the issue that punctual observations are not able to properly capture the spatial average of speed, \(v\), or density, \(k\), within links (i.e. due to the length and position of loop detectors along the link), \((45)\) advocated the use of FCD, herein called \textit{probes method}, for estimating the network speed. Combining such estimated space-mean speed
of cars with the spatially average flow of cars, \( q \) (i.e. based on LDD), it is possible to obtain the density of cars at the network level using the basic fundamental relationship, \( q = kv \) \((4)\). It has been demonstrated that the probes method yields significant improvements in estimating an MFD compared to the single LDD data source (i.e. the method is less sensitive to the loop detector placement bias). Even though the information regarding the FCD penetration rate is not explicitly required, the accuracy of the proposed methodology highly depends on: (i) the number of trajectories used to determine the space-mean speed of cars for the entire network; and (ii) the level of homogeneity in terms of the spatial distribution of probe vehicles.

2.3 MACROSCOPIC MODELING OF URBAN SYSTEMS

Shortly after introducing the MFD, \((13)\) derived an analytical approximation method based on variational theory (VT). This pioneering analytical work inspired further analytical approaches, used to demonstrate how network topology and signal settings (e.g. link length, number of lanes, signal offsets, street organization, etc.) affect the shape of the MFD \((14, 17, 39, 45–48)\). A functional form with a physical meaning has also been proposed \((49)\).

Recognizing the importance of accounting for passenger dynamics, \((50)\) and \((51)\) introduced the passenger MFD (p-MFD), and used it to assess the user and system optimum, and to understand how the traffic conditions are changed by the passenger mode choice. The authors neglected, however, the impact of bus stops and, instead, assigned a lower average speed to buses compared to private vehicles, transforming the problem into a moving bottleneck problem. As such, the problem could be solved with VT, similarly to \((52)\) and \((53)\).

Although there are several analytical approximations for modeling car MFDs \((13–17)\), there is no such formulation for 3D-MFDs or p-MFDs. Car MFDs can be modeled in several ways, e.g. using a mesoscopic model or the method of cuts. The latter one is based on the variational formulation of kinematic wave theory \((18, 19)\) and allows to derive a uni-modal MFD in a single run, for all possible boundary conditions. Consequently, it is one of the most frequently used approaches to develop analytical approximations for an MFD on a corridor level. However, due to the complex and stochastic nature of multi-modal traffic, VT-derived MFDs have not been extended to account for multiple modes at the network level. To the best of
the author’s knowledge, the only attempt in this direction has been made by (54) and (52). While the first reference investigates the impact of bus operations (i.e. dwell time) on the passenger throughput, their approach has the following limitations: (i) it is simulation-based, requiring a large number of simulation experiments to analyze the effects of stochastic bus operations on a p-MFD; (ii) it does not capture possible dependencies of bus arrival times; (iii) it assumes a linear relationship between the bus occupancy and the dwell time, without analyzing passenger dynamics as a function of both the passenger demand and the bus operations; and (iv) it does not account for the impact of the traffic conditions affecting the actual bus speed, and the resulting bus arrival time and dwell time at any given stop. Similarly, the second reference is only focused on the vehicle throughput and deterministic bus operations.

2.4 MACROSCOPIC DESIGN OF URBAN SYSTEMS

Motivated by the need to address the everlasting increase in travel demand, support sustainability, and preserve existing transport land use, transportation solutions are often sought in the domain of public transport systems (20). These systems are regarded as a public service that should provide mobility access to all citizens in an urban area.

One of the commonly explored problems in the public transport sector, more specifically in the domain of bus operations, is the bus network design. In this problem, the arrangement of bus lines atop a street network needs to be determined in a way that it provides a good level of accessibility and service between every pair of points in the city throughout the day.

Several studies have looked at the optimal bus network configuration, investigating different city network structures: grid systems (55), radial systems (56), systems of corridors (57), and hub-and-spoke systems (58). All these studies have a common objective - to determine the topological (e.g. line and stop spacings) and operational (e.g. service frequency) characteristics of the bus system that minimize the user and the operator cost. To achieve a service and accessibility level competitive with that of the automobile at a reasonable cost, (59) proposed a hybrid concept, obtained by combining hub-and-spoke and grid systems. This concept has further been generalized and extended to account for more realistic network configurations (60–62). The model developed by (60) also served as an inspiration to
construct a real transfer-based bus network, the Nova Xarxa in Barcelona, Spain.

A recent study by (63) empirically proved that such a well-designed transfer-based network could attract new users, who would not be opposed to transferring. However, most scientific literature on the bus network design problem (see e.g. 59–62, 64, 65) assume, up to now, that bus users choose the closest origin and destination stops such that they minimize the number of transfers, adjusting their walking distance to meet this criteria. This, in turn, may result in longer access (including the egress) time, thus longer total user time traveled, depending on the selected design parameters (e.g. stop and line spacings). Potential effects of other types of user behavior (e.g. assuming that users are willing to adjust the number of transfers in order to minimize the walking distance) on the user and/or operator cost function have not been investigated.

Furthermore, these studies have also assumed trip origins and destinations to be uniformly and independently distributed across the network, reducing thereby the complexity in the mathematical modeling. Consequently, the optimal bus network design is determined for one particular trip length pattern, imposed by the homogeneously placed origins and destinations. To the best of the author’s knowledge, the only attempt made to address this limitation can be found in (66), who studied the optimal bus network configuration under spatially heterogeneous demand patterns.

Although the last reference investigated an irregular design of the bus network across a city, it still used a single trip length for all users in the network, similarly to all the previous studies. If, however, we look at the urban mobility patterns, we can see that such assumption is not realistic, since the trip lengths are not uniform. In addition, the authors considered only one mode (i.e. buses), without accounting for the collective effect of the analyzed topology in the global traffic performance. This has been relaxed in (67), who studied the impact of the design of transit system on the available street capacities, modal split, and route choice, in (68), who studied the effects of bus network design on traffic, and in (69), who empirically evaluated the effects of bus operations on network’s capacity and critical density. Nevertheless, none of these studies took into account the mode choice resulting from the complex bi-modal interactions and their significant effects on the traffic performance. A recent empirical study by (70) found that the travel time is highly correlated with the topology of a road network. Hence, to maximize mobility in multi-modal urban systems,
the impact of the bus network structure on the traffic performance should be carefully investigated.

2.5 MACROSCOPIC OPTIMIZATION OF URBAN SYSTEMS

To continue to be attractive and a suitable alternative for the car owners, public modes of transport need to provide a good level of service. Among the different aspects of public transport that one can consider to be of influence for the passenger perception of the level of service, its operating regime stands out, i.e. how the dispatched frequencies of public transport vehicles, notably buses, are adjusted to meet urban mobility patterns.

This is known as the frequency setting problem and has been the subject of several studies. Early approaches for determining the optimal bus frequency were based on analytical models (71–74) or heuristic methods (75), and were usually formulated with the aid of graph theory, with nodes and arcs representing bus stops and street segments, respectively. (76) introduced the maximum load section-based method to determine the service frequency that can meet the passenger demand. Following a similar approach, (77) established four alternative methods based on the passenger count data. The objectives were two-fold: to minimize the fleet size and to meet the passenger demand. The minimum fleet size was also used as the objective function by (72), who derived a continuous model for the optimal bus frequency in case of both, single-bus routes (72) and multiple-bus routes (78). (79) and (80) studied the frequency setting problem with stochasticity in the demand, arrival times, boarding/alighting times, and travel times. Using a non-linear program to formulate the problem, (81) and (82) provided an optimal allocation of resources over space and time, while minimizing the weighted sum of ridership and wait time savings. Recent work by (83) determined the optimal bus frequency based on a travel time variability parameter.

To ensure that the determined bus frequency actually leads to an efficient bus service, it is crucial to account for the dynamics of traffic congestion and the interactions between the modes, especially in case of mixed traffic (69). However, existing studies on this topic, up to now, have rarely addressed the frequency setting problem under varying traffic conditions. To the best of the authors’ knowledge, the only attempt made in this direction can be found in (84), who proposed a responsive bus dispatching strategy based on a doubly dynamical approach (as in 85). Although that study investigated a time-dependent bus dispatching problem in a
bi-modal network, it only considered a simple network with two main directions, without analyzing possible effects of the dispatched frequency on the passenger dynamics across different bus stops, and how the available vehicle fleet should be optimally distributed across different bus lines during the planning horizon.

In fact, when determining the optimal bus frequency, one of the most constraining parameters is the available vehicle fleet. The vehicle fleet can consist not only of conventional buses, but also of modular bus units that can either operate individually or combined together (forming thereby a single modular bus of a higher passenger capacity). In railway systems, for example, the vehicle fleet corresponds to a stock of cars and locomotives (commonly referred to as train units) that, when combined, form a single train. The number of train units assigned to a train is, in most cases, determined according to the predefined dispatched frequency of trains and the level of passenger demand (86–89). Only recently did researchers investigate the potential for combining the allocation of train units and the optimization of the dispatched frequency (90–93).

Recent advances in vehicle technology have opened new opportunities to apply similar concepts (i.e. combining and splitting of vehicle units) also for the bus systems (93–95). Such concept, herein called flexible bus dispatching system, can be especially useful for public transport operators, as it offers new perspectives and enormous flexibility to better manage the allocation of the vehicle resources and reduce the operating cost. For example, rather than dispatching a bus with a high passenger capacity in case of low passenger demand, an operator can send a group of few combined modular and fully automated bus units that, overall, has lower passenger capacity (increasing thereby the passenger occupancy) and lower operating cost. Furthermore, as the bus units are fully automated, there is no cost for assigning bus drivers to them.

Despite the potential benefits of combining and splitting modular vehicle units along fixed bus lines, no such research has been conducted in the current state of the art. While some studies can be found in the scientific literature for railway systems (86–92, 96) and more recently for the metro system (93, 95), strategies developed for railway systems are not readily applicable to bus systems. This is because the railway system represents a closed environment, where trains do not interact with other modes. For the bus system, however, we need to consider the complex interactions between buses and cars.
In reality, having a perfect macroscopic picture about the traffic performance is almost never achievable, due to the limited amount of available data. Consequently, different estimation methods have been proposed, which rely on either LDD or FCD, as one of the most commonly used data sources to determine macroscopic traffic variables for the car mode. These existing estimation methods, however, fail to provide a representative outcome regarding the performance of multi-modal traffic when LDD and/or FCD is very limited or non-existent, especially for some time intervals. This poses the need for the development of new estimation tools that are based on other available data sources, such as the AVL data from public transport, and which allow to monitor multi-modal systems even in the extreme cases where no LDD nor FCD exists.

In this chapter, we close this gap and propose novel estimation methods for the space-mean speed of cars based on: (i) the AVL data of public transport where no FCD is available; and (ii) the fused FCD and AVL data sources where both are available, but FCD is not complete. By combining the estimated car space-mean speed with the flow measurements for car traffic, the proposed tools allow to further derive the 3D-MFD and use it for monitoring purposes. Considering that the data needed to apply the proposed methods is readily available and relatively easy to collect, the proposed methodology is pragmatic, adding a unique contribution to the existing body of knowledge on this subject. The remainder of this chapter is mainly based on the following publication:

3.1 General Methodology

In this section, we first describe how the macroscopic traffic variables can be estimated in multi-modal systems based on LDD and FCD. Then, we present a novel Lagrangian method that uses the AVL data from public transport to estimate (or improve the estimation of) the space-mean speed of cars at the network level. For the readers’ convenience, Table 3.1 summarizes the notation used in this chapter. Note that, due to a large number of variables used in this dissertation, the notation might slightly differ from chapter to chapter. Therefore, for consistency, we provide the list of the most important notation used in that chapter.

3.1.1 Eulerian method

For a given transport network, index links by \( l \in L \), where \(|L|\) is the total number of links\(^1\). Let \( N_l \) be the set of lanes on link \( l \), indexed by \( n \in N_l \). Denote the subset of links with LDD by \( L_{LDD} \subseteq L \). Using punctual observations (i.e. data from loop detectors) for each lane \( l \) of length \( \ell_{l,n} \), including flow \( q_{c,l,n}(\tau) \) and density \( k_{c,l,n}(\tau) \), we can derive the fundamental relationships for car traffic as:

\[
\hat{q}_{LDD}^c(\tau) = \frac{\sum_{l \in L_{LDD}} \sum_{n \in N_l} q_{c,l,n}(\tau) \cdot \ell_{l,n}}{\sum_{l \in L_{LDD}} \sum_{n \in N_l} \ell_{l,n}}, \tag{3.1a}
\]

\[
\hat{k}_{LDD}^c(\tau) = \frac{\sum_{l \in L_{LDD}} \sum_{n \in N_l} k_{c,l,n}(\tau) \cdot \ell_{l,n}}{\sum_{l \in L_{LDD}} \sum_{n \in N_l} \ell_{l,n}} \approx \frac{\sum_{l \in L_{LDD}} \sum_{n \in N_l} o_{c,l,n}(\tau) \cdot \ell_{l,n}}{\bar{\ell}_c \cdot \sum_{l \in L_{LDD}} \sum_{n \in N_l} \ell_{l,n}}, \tag{3.1b}
\]

where \( \hat{q}_{LDD}^c(\tau) \) and \( \hat{k}_{LDD}^c(\tau) \) stand for the average weighted flow and density during time interval \( \tau \), respectively. The space-mean speed of cars can then be computed as \( \hat{v}_{LDD}^c(\tau) = \hat{q}_{LDD}^c(\tau) / \hat{k}_{LDD}^c(\tau) \). Considering that, in most cases, the detector measures occupancy \( o_{c,l,n} \) rather than density \( k_{c,l,n} \), we can approximate the density by dividing the occupancy by the space-mean effective vehicle length, \( \bar{\ell}_c \) (4). Although this estimation method is commonly used in practice, it might not be very accurate given that the accuracy of density/occupancy is subject to the spatial distribution of loop detectors within the length of the links, the network coverage of LDD, and the length of the loop detectors. Research by (27) aims to overcome this problem by

---

\(^1\) Two-way links are deemed as two separate one-way links, each having the number of lanes for the corresponding direction of travel.
proposing a new correction method for both flow and occupancy measured by loops, allowing one to use Eq. 3.1 in the future.

<table>
<thead>
<tr>
<th>General variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{M}$ set of data sources indexed by $m$, such that $\mathcal{M} = {\text{LDD, FCD, AVL, FUS}}$</td>
</tr>
<tr>
<td>$\mathcal{T}$ set of time intervals indexed by $\tau$</td>
</tr>
<tr>
<td>$\mathcal{L}$ set of links indexed by $l$</td>
</tr>
<tr>
<td>$\mathcal{N}_l$ set of lanes, indexed by $n$, on link $l$</td>
</tr>
<tr>
<td>$\ell_{i,n}$ length of lane $n$ on link $l$</td>
</tr>
<tr>
<td>$\bar{\ell}_c$ space-mean effective vehicle length</td>
</tr>
<tr>
<td>$\theta$ ratio of the free flow speed of cars and buses, i.e. $\theta = \frac{u_c}{u_b}$</td>
</tr>
<tr>
<td>$T$ time interval duration</td>
</tr>
<tr>
<td>$\hat{\bar{v}}_c(\tau)$ estimated space-mean speed of cars in interval $\tau$ using data source $m$</td>
</tr>
<tr>
<td>$\hat{k}^m_c(\tau)$ estimated density of cars in interval $\tau$ using data source $m$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Loop detector data (LDD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{L}_{\text{LDD}}(\tau)$ subset of links with loop detectors in interval $\tau$</td>
</tr>
<tr>
<td>$\hat{q}^L_{\text{LDD}}(\tau)$ estimated flow of cars from LDD in interval $\tau$</td>
</tr>
<tr>
<td>$q_{l,n}(\tau)$ flow of cars given by LDD on lane $n$ along link $l$ in interval $\tau$</td>
</tr>
<tr>
<td>$o_{l,n}(\tau)$ occupancy of cars given by LDD on lane $n$ along link $l$ in interval $\tau$</td>
</tr>
<tr>
<td>$k_{l,n}(\tau)$ density of cars given by LDD on lane $n$ along link $l$ in interval $\tau$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Floating car data (FCD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{L}_{\text{FCD}}(\tau)$ subset of links containing FCD in interval $\tau$</td>
</tr>
<tr>
<td>$\mathcal{P}_l(\tau)$ set of probe vehicles, indexed by $p$, on link $l$ in interval $\tau$</td>
</tr>
<tr>
<td>$S_l$ set of FCD segments, indexed by $s$, included in link $l$</td>
</tr>
<tr>
<td>$\ell'<em>{l,s}$ length of FCD segment $s$, such that $\ell_l = \sum</em>{s \in S_l} \ell'_{l,s}$</td>
</tr>
<tr>
<td>$d_{c,l,p}(\tau)$ distance traveled by probe vehicle $p$ on link $l$ in interval $\tau$</td>
</tr>
<tr>
<td>$t_{c,l,p}(\tau)$ time traveled by probe vehicle $p$ on link $l$ in interval $\tau$</td>
</tr>
<tr>
<td>$N_{c,l,s}$ number of probe vehicles on FCD segment $s$ along link $l$ in interval $\tau$</td>
</tr>
<tr>
<td>$\bar{v}_{c,l,s}$ space-mean speed of probe vehicles on FCD segment $s$ along link $l$ in interval $\tau$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Public transport data (AVL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{L}_{\text{AVL}}(\tau)$ set of mixed-lane links containing AVL data in interval $\tau$</td>
</tr>
<tr>
<td>$\mathcal{L}'_{\text{AVL}}(\tau)$ subset of mixed-lane links with bus stop containing AVL data in interval $\tau$</td>
</tr>
<tr>
<td>$B_l(\tau)$ set of buses, indexed by $b'$, operating along link $l$ in interval $\tau$</td>
</tr>
<tr>
<td>$d_{b,l,b'}(\tau)$ distance traveled by bus $b'$ on link $l$ in interval $\tau$</td>
</tr>
<tr>
<td>$t_{b,l,b'}(\tau)$ time traveled (including dwell time) by bus $b'$ on link $l$ in interval $\tau$</td>
</tr>
<tr>
<td>$\Omega_{b,l,b'}(\tau)$ dwell (including acceleration/deceleration) time of bus $b'$ on link $l$ in interval $\tau$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fused floating car and public transport data (FUS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{L}'_{\text{AVL}}(\tau)$ subset of mixed-lane links containing only AVL data in interval $\tau$</td>
</tr>
<tr>
<td>$\mathcal{L}''_{\text{AVL}}(\tau)$ subset of mixed-lane links with bus stop containing only AVL data in interval $\tau$</td>
</tr>
<tr>
<td>$\phi^m(\tau)$ network coverage of data source $m$ in interval $\tau$</td>
</tr>
</tbody>
</table>

**Table 3.1: Nomenclature.**
3.1.2 Lagrangian method based on the FCD (Probes method)

Using the trajectory data from probe vehicles it is possible to estimate the average speed of cars at the network level. Let $P_l(\tau)$ and $L_{FCD}(\tau) = \{l \in L : P_l(\tau) \neq \emptyset\}$ stand for a dynamic set of probe vehicles on each link $l$ and a dynamic set of links with FCD in time interval $\tau$, respectively. Then, the estimated space-mean speed of cars from FCD can be derived using the generalized definitions of Edie (97):

$$\hat{\nu}^{\text{FCD}}_c(\tau) = \frac{TDT_c(\tau)}{TTT_c(\tau)} = \frac{\sum_{l \in L_{FCD}(\tau)} \sum_{p \in P_l(\tau)} d_{c,l,p}(\tau)}{\sum_{l \in L_{FCD}(\tau)} \sum_{p \in P_l(\tau)} t_{c,l,p}(\tau)},$$

(3.2)

where $TDT_c(\tau)$ and $TTT_c(\tau)$ stand for the total distance and total time traveled by probe vehicles in time interval $\tau$, respectively; $d_{c,l,p}(\tau)$ is the distance and $t_{c,l,p}(\tau)$ is the time traveled by a single probe vehicle $p$ on link $i$ during time interval $\tau$. The estimated speed of cars can then be further combined with the flow of cars, $\hat{q}_c(\tau)$, to estimate the density of cars, i.e. $\hat{k}^{\text{FCD}}_c(\tau) = \hat{q}_c(\tau) / \hat{\nu}^{\text{FCD}}_c(\tau)$. In cases where the network is surveyed by loop detectors, $\hat{q}_c(\tau)$ can be simple substituted with the detectors counts, i.e. $\hat{q}_c \equiv \hat{q}^{\text{LDD}}_c(\tau)$. Otherwise, other kinds of sensors, such as videos cameras (98) or connected vehicle-based technology (85), can be utilized for acquiring the flow measurements. This makes the FCD-based estimation method very useful, especially in the absence of LDD, if and only if the FCD is available at any given time. Nonetheless, the accuracy highly depends on the fraction of vehicles whose trajectory data are used to estimate the average speed for the entire network. We will elaborate more on this issue later in this chapter.

3.1.3 Lagrangian method based on the AVL data

In multi-modal systems, data can be acquired from different modes. This can be especially instrumental when there is a lack of information from a certain mode (e.g. cars). In that case, the data from other modes (e.g. buses) can be used to draw some conclusions regarding the operating conditions of the mode for which the data is not available (35), for example, estimated the speed of buses based on the known speed of cars, following the concept of the two-fluid theory, according to which the average running speed in a street network is proportional to the fractional running time (22, 99). However, in most urban cities, public transport vehicles are equipped with GPS devices (11, 12) that can be used to determine the ground truth speed
of the public mode of transport. In such cases, it is not necessary to use the speed of cars to estimate the already available speed of buses. Instead, we investigate if and how the data provided by public transport vehicles can be utilized to make an inference on the state of traffic for the private mode of transport, especially when the FCD is limited or not available. In the following, we present a novel Lagrangian method for estimating the macroscopic fundamental relationships, in particular the space-mean speed of cars, using the AVL data from buses. The following assumptions are made for the proposed model:

**Assumption 3.1** (Car accumulation). Traffic conditions are assumed to be homogeneous across the network, as one of the prerequisites for implementing the (3D-)MFD concept. If the network is heterogeneous, some of the existing partitioning algorithms \cite{26,27} can be applied to obtain multiple regions with homogeneous conditions, each revealing a well defined and low scatter (3D-)MFD. In such cases, the proposed method could still be applied to each of the defined regions.

**Assumption 3.2** (Vehicle dynamics). Similarly to \cite{35}, the free-flow speed of buses $u_b$ is assumed to be the same as the free-flow speed of cars $u_c$, i.e. $\theta = u_c / u_b = 1$. If this is not the case, we can relax this assumption by estimating parameter $\theta$ using empirical data.

**Assumption 3.3** (Lane changing). Since we are interested in aggregated traffic dynamics, we do not consider potential lane changing maneuvers that can occur in the proximity of a bus stop.

**Assumption 3.4** (Bus-car interactions). Taking into account that the number of vehicles traveling in front of, or behind a bus is not available, it is assumed that cars are uniformly distributed along a given link.

**Assumption 3.5** (Bus operations). When estimating the cruising speed of buses, we assume that the distance traveled during the deceleration and acceleration periods is negligible in comparison to the total distance traveled by buses for a given time interval (similarly to \cite{35}). Consequently, the estimated cruising speed of buses might be overestimated to a small degree when a bus needs to stop. Nevertheless, if the granularity of the available AVL data was rather high, this distance could be accurately estimated and easily incorporated into the proposed model.

**Assumption 3.6** (Data provision). The only information at our disposal for estimating the car space-mean speed using the AVL data is the bus
distance and time traveled from stop to stop (including the dwell and acceleration/deceleration times at each stop).

Since public transport lines operate only along some links $L_{AVL} \subseteq L$, with a predefined frequency (e.g. every 10 min), the AVL data might not be available for every part of the network and time interval $\tau$. Therefore, we first define a dynamic set of shared mixed-lane links containing the AVL $L_{AVL}(\tau) = \{l \in L : B_l(\tau) \neq \emptyset\}$, where $B_l(\tau)$ represents a dynamic set of buses operating on link $i$ in time interval $\tau$. Using the trajectory data of each individual bus $b'$, we can further determine the total distance $TDT_b(\tau)$ and time traveled $TTT_b(\tau)$, as well as the total cruising time (excluding the dwell and acceleration/deceleration times) of buses $TCT_b(\tau)$ during time interval $\tau$ as:

$$ TDT_b(\tau) = \sum_{l \in L_{AVL}} \sum_{b' \in B_l(\tau)} d_{b,l,b'}(\tau), $$

$$ TTT_b(\tau) = \sum_{l \in L_{AVL}} \sum_{b' \in B_l(\tau)} t_{b,l,b'}(\tau), $$

$$ TCT_b(\tau) = \sum_{l \in L_{AVL}} \sum_{b' \in B_l(\tau)} (t_{b,l,b'}(\tau) - \Omega_{b,l,b'}(\tau)), $$

where $d_{b,l,b'}(\tau)$, $t_{b,l,b'}(\tau)$, and $\Omega_{b,l,b'}(\tau)$ denote the distance traveled, the time traveled, and the dwell time (including deceleration/acceleration time while approaching/departing to/from the bus stop) of bus $b'$ on link $l$ during time interval $\tau$, respectively.

Notice from Eq. 3.3 that the trajectory data from buses provides two values for the total time traveled, with and without the dwell and acceleration/deceleration times. This, in turn, yields two extreme values for the bus space-mean speed at the network level:

$$ v_b(\tau) = \frac{TDT_b(\tau)}{TTT_b(\tau)}, $$

$$ v'_b(\tau) = \frac{TDT_b(\tau)}{TCT_b(\tau)}. $$

If we assume the equality between the free-flow speeds of cars and buses (as per Assumption 3.2), the space-mean speed of cars should then be between these two extreme values, i.e. $v_c(\tau) \in [v_b(\tau), v'_b(\tau)]$. Therefore, in the following, we quantify the conditions under which the bus dwelling process affects the movement of cars. In other words, we find the weights for $v_b(\tau)$ and $v'_b(\tau)$ for computing the car speed at the network level. Note,
However, that, rather than weighting a single bus speed for the entire network, we estimate the weighted bus speed for each segment separately, as there could be a bias if the average car speed is different between mixed-lane links with and without bus stops. The following parameters are assumed to be relevant for defining the weights on the link-basis (Eq. 3.5a):

- $\text{par}_1$ indicating the fraction of cars affected by bus operations.
- $\text{par}_2$ indicating the fraction of the dwell time experienced by cars.
- $\text{par}_3$ indicating the impact of bus network structure on the car speed.

Based on the assumption that cars are uniformly distributed along a given link (Assumption 3.5), we can approximate the fraction of vehicles affected by bus operations as the fraction of the total dwell time within the observed period of time (i.e. duration of the current time interval $T$). Considering, however, that, due to queue dynamics, the actual number of affected vehicles is higher than that, we use an additional factor for $\text{par}_1$ to account for the impact of the traffic conditions (i.e. demand level). Given the lack of other types of data (as per Assumption 3.6), we approximate this factor as the ratio between the free-flow speed of buses and the bus space-mean speed observed on a given link during a given time interval. Intuitively, the longer the dwell time is (i.e. the higher value of this speed ratio), the longer it will take for the traffic conditions to return to the initial state, as a higher number of vehicles will be affected by the dwell time. It is worth mentioning that $\text{par}_1$ implicitly accounts for the bus frequency, as the total dwell time (hence the weighting factor $w_l(\tau)$) depends on whether a bus is present on a given link with a bus stop in a particular time interval or not. Furthermore, to quantify interruptions in the traffic conditions for the car mode caused by the dwelling process, we use $\text{par}_2$ in Eq. 3.5a. This parameter approximates the fraction of the dwell time experienced by cars on average. Given Assumption 3.5 that cars might arrive at any time during the period when a bus is stopped, we assume $\text{par}_2$ to be $1/2$. Finally, $\text{par}_3$ in Eq. 3.5a implies that in case of a one-lane link ($|N_l| = 1$) with a curbside bus stop ($\delta_l = 1$), when a bus stops at a given station, vehicles cannot overtake that bus, which then becomes the leading vehicle (with the speed that includes the dwell and acceleration/deceleration times), followed by the flow of private cars with the same speed. This phenomenon happens only when a bus line operates along a mixed one-lane link and the bus stops are curbside. Therefore, to account for such influence versus the possibility to overtake a bus while boarding/alighting passengers...
at the bus stop location if the number of lanes on that particular link is $|N_i| \geq 2$ or if the link contains a bus bay ($\delta = 0$), we define $par_3$ to be a function of $|N_i|$ and $\delta$, and use it as an additional weighting factor in Eq. 3.5a.

\[
w_l(\tau) = \min \left\{ 1, \left( \frac{\Omega_{b,l}(\tau)}{T} \right) \cdot \left( \frac{u_b}{v_{b,l}(\tau)} \right) \cdot \left( \frac{1}{2} \right) \cdot \left( \max \{0, 2 - |N_l|\} \delta_l \right) \right\},
\]

with

\[
\delta_l = \begin{cases} 
1, & \text{if link } l \text{ contains a curbside bus stop,} \\
0, & \text{if link } l \text{ contains a bus bay,}
\end{cases}
\]

\[\text{(3.5a)}\]

where link $l$ belongs to a dynamic subset of mixed-lane links with a bus stop containing the AVL data in time interval $\tau$, i.e. $l \in L'_\text{AVL}(\tau) \subseteq L\text{AVL}(\tau)$; $w_l(\tau)$ is the weight given to the bus travel time including the dwell (and acceleration/deceleration) time on link $l$; $\Omega_{b,l}(\tau) = \sum_{b' \in B_l(\tau)} \Omega_{b,i,b'}(\tau)$ and $v_{b,i}(\tau) = \sum_{b' \in B_l(\tau)} d_{b,i,b'}(\tau) / \sum_{b' \in B_l(\tau)} t_{b,i,b'}(\tau)$ are the total dwell time and the bus space-mean speed on link $l$, respectively; $u_b$ stands for the free-flow speed of buses; $|N_l|$ represents the number of lanes on link $l$; $\delta_l$ is a binary variable indicating the type of bus stop on link $l$.

Now that we have quantified the impact of boarding/alighting operations on the car speed, we can estimate the space-mean speed of cars using the link-level AVL data as:

\[
\hat{\delta}_c(\tau) = \theta \cdot \frac{TDT_b(\tau)}{TTT^*_b(\tau)},
\]

\[\text{(3.6)}\]

where $TTT^*_b(\tau)$ is the corrected total time traveled by buses in time interval $\tau$ (Eq. 3.7), used to estimate the average speed of cars at the network level.

\[
TTT^*_b(\tau) = \sum_{l \in L\text{AVL}} \sum_{b' \in B_l(\tau)} w_l(\tau) \cdot t_{b,i,b'}(\tau)
+ \sum_{l \in L\text{AVL}} \sum_{b' \in B_l(\tau)} (1 - w_l(\tau)) \cdot (t_{b,i,b'}(\tau) - \Omega_{b,l,b'}(\tau)).
\]

The reader can verify from Eq. 3.7 that, if all bus stops are bus bays (i.e. $\delta_l = 0$, $\forall l \in L'_\text{AVL}(\tau) \Rightarrow w_l(\tau) = 0$), the estimated speed of cars is equivalent to the estimated cruising speed of buses, i.e. $\hat{\delta}_c = v'_{b}(\tau)$ (assuming $\Omega = 1$), given that the boarding/alighting process does not interrupt the operations in mixed-lane segments. Similarly, for a two(or
more)-lane road network ($|N_l| \geq 2, \forall l \in \mathcal{L}'_{AVL}(\tau) \Rightarrow w_l(\tau) = 0$), we also have that $\hat{v}_{AVL}^c(\tau) = v_b'(\tau)$, as each link contains at least one additional lane that can be potentially used for overtaking.

As in the case of the probes method, the AVL-based car space-mean speed can be further combined with the vehicle counts to obtain the density for the car mode, i.e. $\hat{k}_{AVL}^c(\tau) = \hat{q}_c/\hat{v}_{AVL}^c(\tau)$. It is worth to note that, using the proposed AVL-based Lagrangian method it is possible to estimate the (3D-)MFD even where no FCD nor LDD are available. However, the following requirements need to be fulfilled. First, flow measurements collected by any of the alternative methods ($85, 100$) should be available and should cover the same links where the bus information exists. Second, only the AVL data coming from shared mixed-lane road segments can be used. In the extreme case, where all bus lanes are fully dedicated, the proposed methodology should not be applied. The reason is rather simple. Traffic conditions in a dedicated bus lane and in its adjacent car-only lane might differ significantly.

3.1.4 *Lagrangian method based on the fused FCD and AVL data*

So far, we have only discussed how the space-mean speed of cars can be estimated using observations from a single data source, FCD or AVL, each having certain flaws. The accuracy of the method that uses FCD only depends on the number of trajectories used to determine the space-mean speed of cars for the entire network and the spatial distribution of probe vehicles. On the other hand, the AVL data gives just an approximation of the car speed and is useful only when cars and buses drive in mixed-lanes. To address the limitations of both the FCD- and the AVL-based estimation method, we propose the following fusion algorithm for estimating the space-mean speed of cars, which combines FCD and AVL data sources where both are available. The algorithm is shown in a form of a pseudo code below. The following assumptions are made for the proposed fusion method:

**Assumption 3.7 (Data accuracy).** We reasonably assume that the measurement errors associated with different data sources are not available (plus they might change drastically from place to place). Notice that the AVL data gives just an estimation (although relatively accurate) of the car speed, whereas FCD provides the actual speed of probe vehicles. As a result, we assume that adding the bus data to more accurate information received by probe vehicles on the same link would not necessarily improve the estima-
tion results. For these reasons, the information provided by probe vehicles is considered to be more accurate, and therefore valued more for estimating the traffic conditions (i.e. speed) than the AVL data obtained for the same link.

Algorithm 1: Iteration process

Input: 
\{d_{c,l,p}(τ), t_{c,l,p}(τ) : l \in L_{FCD}(τ), p \in P(τ)\}
\{d_{b,l,b'}(τ), t_{b,l,b'}(τ), Ω_{b,l,b'}(τ) : l \in L_{AVL}(τ), b' \in B_l(τ)\}
\{δ_{l'}, |N_{l'}| : l' \in L_{AVL}(τ) ⊆ L_{AVL}(τ)\}

Output: \(\hat{σ}_{c}^{FUS}(τ), \hat{k}_{c}^{FUS}(τ)\)

1. Compute the network coverage of FCD and AVL:

\[
\phi_{FCD}(τ) = \frac{|L_{FCD}(τ)|}{|L|} = \frac{1}{|L|} \cdot \sum_{l \in L} \bar{ξ}_{FCD,l}(τ), \tag{3.8a}
\]

\[
\phi_{AVL}(τ) = \frac{|L_{AVL}(τ)|}{|L|} = \frac{1}{|L|} \cdot \sum_{l \in L} \bar{ξ}_{AVL,l}(τ), \tag{3.8b}
\]

with

\[
\bar{ξ}_{FCD,l}(τ) = \begin{cases} 1, & \text{if } P_l(τ) \neq \emptyset, \\ 0, & \text{otherwise}, \end{cases} \tag{3.9a}
\]

\[
\bar{ξ}_{AVL,l}(τ) = \begin{cases} 1, & \text{if } P_l(τ) = \emptyset, B_l(τ) \neq \emptyset, \\ 0, & \text{otherwise}. \end{cases} \tag{3.9b}
\]

2. Compute: \(TDT_c(τ), TTT_c(τ), TDT_b(τ), TTT^*_b(τ)\).

3. Compute the fused space-mean speed of cars:

\[
\hat{σ}_{c}^{FUS}(τ) = \begin{cases} \frac{TDT_c(τ)}{TTT_c(τ)}, & \text{if } \phi_{FCD}(τ) = 1 \text{ or } \phi_{AVL}(τ) = 0, \phi_{FCD}(τ) \neq 0, \\ \frac{TDT_b(τ)}{TTT^*_b(τ)}, & \text{if } \phi_{FCD}(τ) = 0, \phi_{AVL}(τ) \neq 0, \\ \frac{\phi_{FCD}(τ) \cdot TDT_c(τ) + φ_{AVL}(τ) \cdot TDT_b(τ)}{φ_{FCD}(τ) \cdot TTT_c(τ) + φ_{AVL}(τ) \cdot TTT^*_b(τ)}, & \text{if } \phi_{FCD}(τ) \in (0, 1), \phi_{AVL}(τ) \neq 0, \\ F(\hat{σ}_{c}^{FUS}(τ - 1), \hat{σ}_{c}^{FUS}(τ - 2), ...), & \text{if } \phi_{FCD}(τ) = 0, \phi_{AVL}(τ) = 0. \end{cases} \tag{3.10}
\]

4. Estimate the average network density of cars: \(\hat{k}_{c}^{FUS}(τ) = \hat{q}_c(τ)/\hat{σ}_{c}^{FUS}(τ)\).
Assumption 3.8 (Probe penetration rate). For generality, we assume that the penetration rate of probe vehicles is not known in advance (i.e. it is not available or is very difficult to infer), making the derivation of an optimal fusion algorithm rather challenging. If this information was provided, then the already proposed LDD-FCD fusion procedure by (44) could be used.

Essentially, for a given time interval $\tau$, the algorithm dynamically divides the entire network into two subnetworks: (i) FCD subnetwork $L_{\text{FCD}}(\tau)$; and (ii) AVL subnetwork $L_{\text{AVL}}(\tau)$. The prerequisite to incorporate link $l$ in time interval $\tau$ in the set of links containing FCD is to have at least one probe vehicle traveling at that time on link $l$. Otherwise, the link becomes a potential candidate for inclusion in the set of AVL links, if there is a public transport line operating along that link in time interval $\tau$. In case of an overlap, when a link contains both FCD and AVL, the link gets included only in set $L_{\text{FCD}}(\tau)$, as the FCD information is considered to be more accurate than the AVL information acquired for the same link (see Assumption 3.7). In other words, rather than using $L_{\text{AVL}}(\tau)$ and $L'_{\text{AVL}}(\tau)$ sets for the AVL subnetwork, which might be overlapping with $L_{\text{FCD}}(\tau)$, the algorithm considers only the subset of shared mixed-lane links, $L^*_{\text{AVL}}(\tau) = L_{\text{AVL}}(\tau) \setminus L_{\text{FCD}}(\tau)$, and the subset of shared mixed-lane links with bus stops, $L'^*_{\text{AVL}}(\tau) = L'_{\text{AVL}}(\tau) \setminus L_{\text{FCD}}(\tau)$, containing no FCD in time interval $\tau$. This ensures that the information received by public transport vehicles contributes only to the subnetwork where no FCD is available.

In the next step, for the same time interval $\tau$, the algorithm computes the network coverage of FCD, $\phi_{\text{FCD}}(\tau)$, and AVL, $\phi_{\text{AVL}}(\tau)$, according to Eq. 3.8. Note that $\phi_{\text{AVL}}(\tau) \leq 1 - \phi_{\text{FCD}}(\tau)$, given that the AVL data is used as a complementary source of information only for the subnetwork where no probe vehicles are registered. Taking into account Assumption 3.1 regarding the homogeneous traffic conditions (i.e. similar level of car accumulation in the two subnetworks, $L_{\text{FCD}}$ and $L^*_{\text{AVL}}$), the fused space-mean speed can then be simply computed as an average of the two types of measurements, weighted by their respective network coverage ($\phi_{\text{FCD}}$ in case of FCD; $\phi_{\text{AVL}}$ in case of AVL) (Eq. 3.10). It is worth mentioning that by fusing the total distances and times traveled by both probe vehicles and buses for computing the car space-mean speed, we implicitly take into account the number of vehicles of each mode, whose trajectory data is used for the fusion process. This, however, is not the case when fusing the space-mean speeds computed from different data sources separately.

Notice the interdependence between the network configuration and the weight given to the AVL(FCD)-based estimated parameters. Intuitively,
having a higher fraction of shared mixed-lanes will increase (decrease) the weight given to the AVL (FCD) data, as the impact of public transport on the private mode is more likely to be higher. Furthermore, if the network coverage of either FCD or AVL is zero, no fusion takes place. In that case, the information from the available data source will be used to estimate the space-mean speed of cars for the entire network. This approach can be extremely valuable for many cities where there is abundant AVL data, but no FCD whatsoever (e.g. Santiago, Chile). Similarly, if probe vehicles are registered on every link, at any given time interval, i.e. $L_{\text{FCD}}(\tau) = L$, the space-mean speed of cars will be estimated using a single-data source, FCD. If, however, none of these two data sources is available for a particular time interval, then some of the short-term prediction algorithms can be applied (herein defined as a function $F(\cdot)$ in Eq. 3.10), such as the one by (101), assuming that the traffic conditions are slow-varying.

In the last step, the algorithm combines the estimated car space-mean speed with the network flow of cars (if available) to estimate the density of cars. Note, however, that in terms of the (3D-)MFD estimation, the proposed Lagrangian methods suffer from the same restriction as the estimation methods based on either LDD or FCD: the resulting (3D-)MFD only represents the subnetwork where observations are available. If we assume that vehicular flows are equally available for different parts on the network, both the AVL- and the fusion-based method will yield a relatively accurate estimation of the (3D-)MFD. However, if flows are given only for congested roads, as an example of an extreme case, the density of cars, therefore the (3D-)MFD, will not be reliable, regardless of how accurate the estimated car space-mean speed is.

3.1.5 Determining the accuracy of the estimated space-mean speed of cars

Here, we introduce the approach for measuring the accuracy of the estimated space-mean speed of cars when the network coverage is below 100%, i.e. $L_{\text{FCD}}(\tau) \subset L$. Let $v_c(\tau)$ and $\hat{v}_c(\tau)$ stand for the full network coverage (i.e. the ground truth) and the estimated car space-mean speed (using incomplete information), respectively. Following the approach by (42), we formulate the estimation error for a given time interval $\tau$ using Eq. 3.11. The average relative error for the entire period of time can then be
computed as $\overline{\Delta R} = \sum_{\tau \in \mathcal{T}} \Delta R(\tau)/|\mathcal{T}|$, where $\mathcal{T}$ stands for the set of time intervals during the analyzed period.

$$\Delta R(\tau) = \Delta R(v_c(\tau), \hat{v}_c(\tau)) = \left| \frac{v_c(\tau) - \hat{v}_c(\tau)}{v_c(\tau)} \right|$$  \hspace{1cm} (3.11)

### 3.1.6 Evaluating the performance of the proposed fusion method

To determine how accurate the proposed fusion algorithm is, we follow the idea of (44) and compare its performance with that of a reference method, providing the best possible estimation given the available, incomplete information. The reference method minimizes the relative error introduced in the previous section. For consistency, we use the same basic idea as for the fusion method: weighting the total distance and total time traveled for the two subnetworks with FCD and AVL data. The solution is obtained using Nelder and Mead heuristic search approach, as this is one of the commonly applied numerical methods for finding the minimum of an objective function in a multi-dimensional space:

1. Weight the FCD- and AVL-based estimated total distance and total time traveled with parameters $\alpha$ and $\beta$, respectively:

$$\hat{v}^{\text{REF}}_c(\tau) = \frac{\alpha \cdot TDT_c(\tau) + \beta \cdot TDT_b(\tau)}{\alpha \cdot TTT_c(\tau) + \beta \cdot TTT^*_b(\tau)}$$  \hspace{1cm} (3.12)

2. Define the starting values for both $\alpha$ and $\beta$. Herein, we set $(\alpha, \beta) = (0.5, 0.5)$.

3. Compute the average relative error, $\overline{\Delta R}$, across all network coverages and define new weights in the first step, $(\alpha, \beta) \leftarrow (\alpha', \beta')$, according to (102) procedure.

4. Repeat the procedure until convergence is reached. Herein, we define convergence as reaching a difference of less than $10^{-8}$ in one iteration step.

Note that the performance of the reference method is a function of the used optimization procedure and the way the method is defined. Coefficients used in the reference case are determined based on the ground true space-mean speed of cars, hence are considered to be the best possible for a given measurement. The results can be interpreted as the lower bound for
the error of the estimated car speed that could be obtained with the fusion algorithm using incomplete data. In addition to the reference method, we will also compare the proposed fusion method with one of the commonly used approaches, the FCD-based method (Eq. 3.2).

3.2 CASE STUDIES

For comparison and validation purposes, we use two types of network: a simulated abstract grid network and the real transport network of the city of Zurich, Switzerland. The simulated abstract network is used to gain insights on the performance of the proposed AVL-based method for estimating the speed of cars, as well as for testing the robustness of the fusion algorithm under various operational characteristics of public transport, network coverages of both FCD and AVL, and the spatial distribution of probe vehicles (homogeneous or inhomogeneous). The real Zurich network is used to empirically validate and demonstrate the applicability of the proposed Lagrangian methods in real-life traffic scenarios.

3.2.1 Abstract network

3.2.1.1 General simulation settings

The abstract network (Fig. 3.1a) was developed using a VISSIM microsimulation platform, as a 10 × 10 grid, with |L| = 180 links and ℓ_l = 120 m, ∀l ∈ L. 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.

To evaluate the proposed AVL-based estimation method for the space-mean speed of cars, we use two network configurations: one-lane and two-lane road networks, both with curbside bus stops. The former type of network is used to demonstrate that, in cases where the level of interactions between cars and buses is the highest, neither the cruising speed of buses nor the actual speed of buses including the dwell (plus acceleration/deceleration) time can yield the best approximation of the space-mean speed of cars. On the other hand, in the latter type of network, where cars can utilize an additional lane for either overtaking a bus while dwelling or not choosing to travel along a mixed-lane segment at all, the cruising speed of buses can actually give a good estimate of the car space-mean speed. In other words, using both types of network we show that it is necessary to
account for the relevant system’s parameter, as described by Eq. 3.5. Note that the latter network configuration is also used to evaluate the performance of the proposed fusion algorithm.

![Study networks](image)

**Figure 3.1:** Studied networks: a) Snapshot of the simulation network layout; b) Zones of interest in the city of Zurich - note that the map also shows residential streets, which are not included in this study.

3.2.1.2 *Simulation scenarios*

Tested traffic scenarios ranged from 5% of both FCD network coverage and probe penetration rate (PPR), to 100%, using increments of 5 percentage points. To account for the randomness, 400 combinations for each of the 400 scenarios were tested. In other words, the combination of links containing the FCD and the combination of vehicles marked to be probes were varied 400 times for each specific traffic scenario. PPR was obtained by randomly selecting the corresponding portion of vehicles for a given scenario, either uniformly across the entire network (homogeneous distribution) or from only one quadrant of the network (inhomogeneous distribution), similarly to (43). On the other hand, the FCD network coverage was acquired by randomly selecting the corresponding portion of links for a given scenario, where the trajectory data of at least one probe vehicle was available.

For each traffic scenario we also varied the coverage of public transport lines (shown as black colored links in Fig. 3.1a) in terms of the percentage of lanes containing the AVL data $\phi_{AVL}$. Initially, we set $\phi_{AVL} = \{40\%, 15\%\}$ in case of a one-lane network, $\phi_{AVL} = \{37.5\%, 12.5\%\}$ in case of a two-lane network, and used a bus frequency of $f_b = 7.5\text{ min}$ in all scenarios. Additional simulation experiments for testing the robustness of the proposed fusion algorithm included lower coverages and frequencies of public transport.
lines. One should note that the bus lines operate in a mixed-lane fashion, i.e. no dedicated lanes are allocated to public transport vehicles.

Finally, for the purpose of collecting the trajectory data, each vehicle in the network was tracked with a simulation resolution of 1Hz, during one hour of the evaluation time (same as in 103). A uniform traffic demand was loaded consecutively within an hour. OD nodes were placed in the middle of each of $|\mathcal{L}|$ links, with the routes dynamically assigned according to the built-in dynamic traffic assignment (DTA) module in VISSIM (see 104, for more details). To account for the stochastic nature of traffic, five different random seeds were executed for each traffic scenario.

### 3.2.2 City of Zurich network

Here, we use empirical data for the real-case-study network of the city of Zurich, Switzerland. Spatially, we focus our analysis on two regions within the city: the City Center and Wiedikon (Fig. 3.1b). These regions were selected as they were used in a previous study (7) that demonstrated the existence of the 3D-MFD. Temporally, we concentrate on the time period between the 26th and 30th of October 2015. This period includes five weekdays and for each day we use the data from 06:00 to 24:00 hrs.

Data on the private mode by is collected by both loop detectors and floating car vehicles. In Zurich, 4852 loop detectors store vehicle counts and occupancy, recorded in a high-resolution time format (0.1 sec) and aggregated on a 15-min basis (for more details see 7, 38). To convert occupancy into density as in Eq. 3.1b, we use an effective vehicle length of $\bar{\ell}_c = 6.3$ m (105).

For the FCD, we matched the GPS trajectories obtained from TomTom navigation devices (27) with the road network. The accuracy level is 10 m and the available TomTom dataset provides aggregated TomTom segmentation-based results. Note that multiple TomTom segments $s$ with the length of $\ell'_{l,s}$ form a link from intersection to intersection, such that $\ell_l = \sum_{s \in S_l} \ell'_{l,s}$, where $S_l$ stands for the set of TomTom segments included in link $l$. For each TomTom segment, the TomTom dataset includes the average speed $v_{c,l,s}$ of all probe vehicles that traveled along that segment during a recording interval of $\tau = 15$ min, the number of unique probe vehicles $N_{c,l,s}$, contributing to this mean value, and the standard deviation of the time-mean speeds. Using this standard deviation, it is possible to transform the time-mean speeds to space-mean speeds, as described by (106). Additionally, given the fact that all probe vehicles detected on a particular TomTom
segment actually traverse its distance during time $\tau$, we can further estimate the total distance $TDT_c(\tau)$ and the total time traveled $TTT_c(\tau)$ by probe vehicles across all links with the TomTom data (Eq. 3.13). Then, the FCD-based car space-mean speed can be derived using Eq. 3.2.

$$TDT_c(\tau) = \sum_{l \in L_{FCD}} \sum_{s \in S_l} N_{c,l,s}(\tau) \cdot \ell'_{l,s}$$  \hspace{1cm} (3.13a)

$$TTT_c(\tau) = \sum_{l \in L_{FCD}} \sum_{s \in S_l} \frac{N_{c,l,s}(\tau) \cdot \ell'_{l,s}}{v_{c,l,s}(\tau)}$$  \hspace{1cm} (3.13b)

Data on public transport is acquired from Zurich’s transit operator (VBZ). The obtained dataset contains information on the dwell time (including acceleration and deceleration times) at each stop and the travel time from stop to stop, for each public transport vehicle (recorded by AVL tracking devices). Given the geo-spatial structure of the TomTom data, each public transport segment may encompass multiple TomTom segments that we have merged into one public transport link for consistency. Based on the known travel (cruising) time, dwell time (including acceleration and deceleration times), distance between each two consecutive stops, and the number of public transport vehicles operating during the analyzed period of time, we can estimate the AVL-based car space-mean speed using Eqs. 3.6–3.7. Note that, in this dissertation, we focus only on the bus mode due to the similar dynamic driving characteristics with the car mode. The tram mode, although can be used to estimate the 3D-MFD (by considering it for the total travel production and accumulation of public transport), should not be used to estimate the speed of cars.

3.3 Results

3.3.1 Simulation outputs

3.3.1.1 Evaluation of the proposed AVL-based estimation method

We first evaluate the accuracy of the method that uses data from public transport to estimate the car space-mean speed. Results are shown in the form of time-series plots (Fig. 3.2), comparing the ground truth space-mean speed of cars $v_c$ with: (i) the proposed AVL-based method, $\hat{v}_{c,AVL}$ (Eqs. 3.6–3.7); (ii) the observed space-mean speed of buses including the dwell (plus acceleration/deceleration) time, $v_{b}$; and (iii) the estimated cruising speed of buses excluding the dwell and acceleration/deceleration times,
$v^\prime_b$. For comparison purposes, two scenarios are considered: one-lane and two-lane road networks, both with curbside bus stops. The first 5 min of the simulation are used as a warm-up period.

Figs. 3.2a and 3.2b show that the proposed AVL-based method yields the best overall estimation of the space-mean speed of cars, whereas the speed of buses including (excluding) the dwell and acceleration and deceleration times substantially underestimates (overestimates) the speed of private vehicles. This highlights the importance of accounting for the relevant system’s parameters, as described by the proposed approach (Eq. 3.5). Quantitatively speaking, for the scenario with 40% (15%) of mixed-lanes in a one-lane road network, the average error on the estimated space-mean speed is 6% (8%), 20% (19%), and 11% (13%) in case of $\hat{v}_c^{AVL}$, $v_b$, and $v^\prime_b$, respectively. Evidently, the higher the network coverage of the AVL data is, the more accurate the estimation of the space-mean speed of cars using the proposed AVL-based method becomes.

![Diagram](image)

**Figure 3.2:** Comparison of the AVL-based estimation methods for the space-mean speed of cars, for a one-lane network with: a) 40% mixed-lanes; b) 15% mixed-lanes; and a two-lane network with: c) 37.5% mixed-lanes; d) 12.5% mixed-lanes.
In case of a two-lane network (Figs. 3.2c and 3.2d), the dwelling process at the bus stop does not impact the movement of private vehicles as in the case of a one-lane road network. In such situations, cruising speed of buses can actually be a good estimator of the space-mean speed of cars. For the scenario with 37.5% and 12.5% of mixed-lanes, the average error on the estimated space-mean speed is 6% and 9%, respectively.

3.3.1.2 Evaluation of the proposed data fusion algorithm

The results are shown in the form of surface plots, combining the contour lines that connect all points with the same average error for the estimated speed of cars. The dataset contains approximately 60 million entries, combining all traffic scenarios and random seeds.

Fig. 3.3 shows the error plot for a two-lane network scenario with 37.5% of lanes being mixed and the bus frequency of 7.5 min, under varying FCD coverages $\phi_{FCD}$ and PPR, for both homogeneous (Fig. 3.3a) and inhomogeneous distribution of probes (Fig. 3.3b). Such network setup allows us to demonstrate how the proposed fusion algorithm performs when the granularity of the AVL data is rather low, i.e. the bus frequency is larger than the aggregation interval of 5 min. Note that for the very low coverage of FCD (i.e. 5%), the proposed fusion algorithm can substantially reduce the error on the space-mean speed of cars (by more than 50%) when compared to the FCD-based estimated speed $\hat{v}_{FCD}$. Furthermore, the errors in the fusion algorithm are almost invariant to changes in the FCD network coverage and PPR, whereas the accuracy of $\hat{v}_{FCD}$ is not. Therefore, the difference between these methods decreases with an increase in the FCD coverage until all the errors converge to the same value: when $\phi_{FCD} = 1$ and no fusion takes place. One should emphasize that for visualization purposes, we only show the FCD network coverage up to 20% (after which the differences between compared methods slowly diminish, especially in case of the homogeneous distribution of probes). In general, the magnitude of the errors in the FCD-based method is significantly higher when probe vehicles are inhomogeneously distributed across the network. This is because the FCD cannot be equally available for all parts of the network, especially when the FCD network coverage is very low. These problems do not appear with the fusion algorithm, which complements the lack of data (low FCD coverage) with the information coming from public transport, substantially reducing the error on the estimated car space-mean speed and outperforming the commonly used FCD-based method.
In terms of the comparison between the proposed fusion algorithm and the reference method, we observe that the reference method is only slightly better (Fig. 3.3). This, in turn, implies that the results of the relatively simple fusion algorithm can emulate a much more complex and computationally demanding optimization procedure, for which the ground truth car space-mean speed has to be known a priori.

![Figure 3.3: Average error on the space-mean speed of cars when $\varphi_{AVL} = 37.5\%$, $f_b = 7.5$ min, and the spatial distribution of probe vehicles is: a) Homogeneous; b) Inhomogeneous.](image)

Given that the overall percentage of mixed-lanes in the previous scenario is relatively high (37.5\%), we have executed additional experiments to assess the impact of low AVL network coverage on the performance of the proposed fusion algorithm. These experiments include 12.5\% of mixed-
lanes and a bus frequency of 7.5 min. The results are shown in Fig. 3.4 and imply similar findings as before: the FCD-based estimation method is always outperformed by the fusion algorithm. Note that for this scenario the error obtained by the fusion method is higher than in the previous case, due to the reduced interactions between different modes of transport, i.e. the coverage of public transport lines. As a result, more weight is given to the FCD-based estimated speed of cars (Eq. 3.10), which by itself has a relatively high error.

Let us now compare the FCD- and AVL-based estimation methods for different FCD network coverages and PPR, so that we can identify the range, i.e. combinations of $\phi^{FCD}$ and PPR, for which the proposed AVL-based method outperforms the FCD-based method. Such frontier is shown as the dashed colored line in Fig. 3.5. There are two lines on each graph, indicating two different network coverages of public transport lines ($\phi^{AVL}_1 = 12.5\%$ and $\phi^{AVL}_2 = 37.5\%$). It can be observed that the improvements from using the bus data can be acquired for both homogeneous (Fig. 3.5a) and inhomogeneous distribution of probes (Fig. 3.5b). Remarkably, as probe vehicles become unequally distributed across the network, the error in the FCD-based method becomes higher, therefore the frontier also becomes higher (Fig. 3.5b). Further notice that, with an increase in the network coverage of public transport lines, the aforementioned frontier increases too.

![Figure 3.5: Contour plot of the range of improvements obtained by the proposed AVL-based estimation method in comparison to the FCD-based method, when the spatial distribution of probe vehicles is: a) homogeneous; b) inhomogeneous.](image)

Although the coverage and availability of the AVL data in the last tested scenario are limited, the improvements generated by the fusion algorithm
Comparison between the fusion and the reference method for:

<table>
<thead>
<tr>
<th>(0+) 8</th>
<th>(I+) 9</th>
<th>(I+) 9</th>
<th>(I+) 9</th>
<th>(I+) 9</th>
<th>(I+) 11</th>
<th>(I+) 11</th>
<th>(I+) 11</th>
<th>(I+) 11</th>
<th>(I+) 11</th>
<th>(I+) 11</th>
<th>(I+) 11</th>
<th>(I+) 11</th>
<th>(I+) 11</th>
<th>(I+) 11</th>
<th>(I+) 11</th>
</tr>
</thead>
<tbody>
<tr>
<td>(I+) 8</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
</tr>
<tr>
<td>(I+) 8</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
</tr>
<tr>
<td>(I+) 8</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
</tr>
<tr>
<td>(I+) 8</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
</tr>
<tr>
<td>(I+) 8</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
</tr>
<tr>
<td>(I+) 8</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
</tr>
<tr>
<td>(I+) 8</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
</tr>
<tr>
<td>(I+) 8</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
</tr>
<tr>
<td>(I+) 8</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
</tr>
<tr>
<td>(I+) 8</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
</tr>
<tr>
<td>(I+) 8</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
<td>(I+) 9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Homogeneous distribution</th>
<th>Pdd[50%]</th>
<th>Pdd[50%]</th>
<th>Pdd[50%]</th>
<th>Pdd[50%]</th>
<th>Pdd[50%]</th>
<th>Pdd[50%]</th>
<th>Pdd[50%]</th>
<th>Pdd[50%]</th>
<th>Pdd[50%]</th>
<th>Pdd[50%]</th>
<th>Pdd[50%]</th>
<th>Pdd[50%]</th>
<th>Pdd[50%]</th>
<th>Pdd[50%]</th>
<th>Pdd[50%]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.20</td>
<td>0.53</td>
<td>0.01</td>
<td>0.05</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Homogeneous distribution

Inhomogeneous distribution

Table 3: Comparison between the fusion and the reference method for:

\[ \phi = \frac{\text{Pdd}}{10} \]
are still substantial. In other words, in the presence of AVL data and incomplete FCD, the fusion algorithm always yields lower errors on the estimated space-mean speed of cars than the FCD-based method. As the network coverage of FCD increases, the difference between the FCD-based and the fusion method decreases, until it completely disappears in the scenario when no AVL data exists or FCD is complete.

For further testing the robustness of the proposed fusion algorithm, we have conducted additional simulation experiments that include less favorable scenarios in terms of the availability of the AVL data, i.e. $\varphi^{AVL} = \{10\%, 5\%\}$ and $f_b = 15$ min. The results are shown in Table 3.2, comparing the average errors on the space-mean speed of cars for the fusion method, $\hat{v}_{FUS}^c$, with the reference case, $\hat{v}_{REF}^c$. Note that, rather than using absolute values to show the difference in errors obtained by both methods, we give the relative difference $\Delta$, i.e. the number of percentage points (ppt) difference between the error for $\hat{v}_{FUS}^c$ and the error for $\hat{v}_{REF}^c$. We exclude the FCD-based method from the comparison, as it is outperformed by the fusion algorithm in all tested scenarios. One can conclude from the obtained results that the fusion algorithm still remains close to the reference method, even if the coverage of public transport data is rather low.

3.3.2 Empirical analysis

3.3.2.1 Cross-comparison of macroscopic traffic models

In this section, we use empirical data from the city of Zurich to validate the findings from the simulation experiments and confirm that they hold true in real-life traffic scenarios. Using the described methodology (see Section 3.1), we estimate the macroscopic traffic variables including flow, speed, and density from all data sources (LDD, FCD, and AVL) and derive the macroscopic fundamental relationships for the car mode. Note that, for brevity, the comparison is done only in terms of the car MFD and not the 3D-MFD, since all considered estimation methods would use the same public transport data when computing the macroscopic traffic variables for the bus mode, hence no differences would be shown for that part.

Fig. 3.6 shows for illustration purposes the MFD for the City Center. Each point in the resulting MFD corresponds to a specific 15-min time interval. There are a total of 360 points for the 5 weekdays analyzed (72 points per day).

When validating the proposed AVL-based method through the comparison with the FCD-based method (Fig. 3.7), we can observe that they are
Figure 3.6: Macroscopic fundamental relationship for the car mode between: a) flow and density; b) speed and density.

Figure 3.7: Empirical validation of the proposed AVL-based Lagrangian method for an average: a) working day; b) weekend day.
well aligned in both regions of interest. Note that, in the City Center, the AVL-based estimated speed of cars $\hat{v}_c^{AVL}$ is almost equivalent to the estimated cruising speed of buses $v'_b$, as there is almost no segment with a bus stop where the bus operations impacts the car speed (Eq. 3.5). On the other hand, in Wiedikon, due to the different bus network configuration layout (i.e. higher percentage of curbside bus stops), we can observe a slightly different pattern: the proposed AVL-based estimated speed of cars $\hat{v}_c^{AVL}$ yields the lowest average error (13%) in contrast with the estimated cruising speed of buses $v'_b$ (15%), as well as the actual speed of buses including the dwell (plus acceleration and deceleration) time $v_b$ (20%), for an average working day. Similar results are also obtained for an average weekend day: the errors in case of $\hat{v}_c^{AVL}$, $v'_b$, and $v_b$ are 14%, 16%, and 23%, respectively. This outcome is in accordance with the findings from the simulation experiments described in the previous subsection.

It is worth mentioning that the example from the city of Zurich presents some unfavorable conditions for evaluating the proposed AVL-based estimation method. 75% of the public transport lanes in the City Center and 60% in Wiedikon are dedicated lanes (7), and public transport enjoys full priority at the vast majority of the intersections in both areas. Nevertheless, the presented results show that Eqs. 3.5–3.7 provide a good estimation, with a relatively low average error compared to the FCD-based speed of cars. In case the network coverage (i.e. percentage of mixed-lanes) and the spatial granularity of the AVL data were higher, or the travel time data were provided in a high-resolution format (e.g. every second), the accuracy(error) of the AVL-based estimated speed of cars would be higher(lower).

3.3.2.2 Practical application of the proposed data fusion algorithm

Using the available empirical FCD and AVL data, in this section, we implement the proposed fusion algorithm on the real-case study network of the city of Zurich. Few things have to be highlighted. First, there was no problem in deriving the fundamental relationship using the proposed fusion algorithm, showing its practical application in real-life traffic scenarios. Second, the (3D-)MFD acquired by combining the fused space-mean speed of cars with the loop detector counts is almost completely aligned with the FCD-based (3D-)MFD. For brevity, we do not show the resulting fusion-based (3D-)MFD here. To unveil the reasons for such an outcome, we compare the network coverages of both FCD and AVL data sources,
for all observed weekdays and all time intervals within each day, for both regions. The results of this comparison are shown in Fig. 3.8.

![Figure 3.8: Network coverages of FCD and AVL data in the city of Zurich, for the analyzed weekdays and time periods.](image)

In each region, we can observe a relatively high coverage of FCD compared to relatively low or, in some time intervals, almost non-existent AVL coverage. The explanation for this is three-fold. First, it is sufficient to have at least one probe vehicle traveling along a certain link to not consider the AVL data on that particular link for a given time interval. Second, for longer time intervals, probe vehicles might traverse multiple links, which are consequently all being included in the set of links containing the FCD. The time interval used in our analysis is set to be 15 min, given that the TomTom (FCD) data are only available in a 15-min time format. Third, the AVL coverage shown in Fig. 3.8 is only related to the coverage of the public transport subnetwork where no FCD exists and which, by itself, does not contain a high portion of lanes reserved for mixed traffic. In other words, the overall coverage of the AVL data is much higher, but for the proposed fusion algorithm we only use the mixed-lane public transport subnetwork containing no FCD. As a result, the weight given to the AVL-based estimated speed of cars is negligible, resulting in practically the same (3D-)MFD as for the FCD-based estimation method, especially for the City Center.

3.3.2.3 Empirical evaluation of the proposed data fusion algorithm

Above, we demonstrated the practical application of the proposed fusion algorithm. However, due to relatively high network coverage of FCD, the (3D-)MFD obtained by fusing the information from FCD and AVL showed
the same shape as the (3D-)MFD derived from FCD only. Hence, it was
difficult to infer the accuracy level that could be achieved by the proposed
fusion method. To address that, we conduct another set of empirical exper-
iments.

First, given the lack of a better estimate, we assume the FCD-derived
(3D-)MFD to be the ground truth (3D-)MFD. This assumption is realistic,
as the analysis of the spatial distribution of TomTom trajectories reveals
that: (i) both regions are almost completely covered by the FCD; and (ii)
the distribution of probe vehicles across the analyzed network seems to be
homogeneous.

We now define the following hypothetical scenarios, considering differ-
ent coverages of FCD, similarly to the simulation experiments from the pre-
vious sections: $\phi_{FCD} = \{5\%, 10\%, 15\%\}$. For all three scenarios, we compare
the performance of the fusion algorithm, $\hat{v}_{FUS}$, and the FCD-based speed of
cars estimated using the limited amount of FCD (i.e. 5%, 10%, and 15%),
$\hat{v}_{FCD-LIM}$, with the FCD-based speed of cars estimated using the full dataset
(100% of the available FCD), $\hat{v}_{FCD}$. We randomly select the corresponding
portion (5%, 10%, and 15%) of the FCD links, so that the homogeneous
distribution of probes is still retained.

![Figure 3.9: Average error on the space-mean speed of cars in case of the limited FCD network coverage (5%, 10%, and 15%).](image)

The resulting errors on the estimated space-mean speed of cars are
shown in the form of box plots (Fig. 3.9), combining the error values of
5 different random combinations of links containing the FCD, for each esti-
mation method. First thing to notice is that, in both regions, the FCD-based
estimated speed using the limited FCD yields the highest error, whereas
the lowest error is obtained by the fusion algorithm, especially in Wiedikon. As we stated before, the reason for such an outcome lies in the fact that the limited FCD used to estimate $\hat{v}_{\text{FCD-LIM}}^c$ might not include any of the mixed-lane segments. On the other hand, by incorporating the AVL data for segments where the trajectories of public transport vehicles are recorded and no FCD is available, we essentially extend our set of observations. This clearly indicates and empirically proves the benefits of the proposed fusion algorithm.

3.4 SUMMARY

In this chapter, we present novel Lagrangian methods for estimating the space-mean speed of cars in bi-modal urban networks: (i) the AVL-based method, which uses only the data provided by public transport vehicles where no FCD is available, and which accounts for the network configuration layout and the configuration of the public transport system; and (ii) the fusion algorithm, which combines the FCD and AVL data sources where both are available, and which does not require a priori knowledge about the probe penetration rate. We use two types of network to evaluate the performance of the proposed methods: a simulated abstract grid network and the real transport network of the city of Zurich.

The simulation results reveal that the AVL-based estimation method can provide a good approximation of the spatially averaged speed of cars at the network level. This further implies that, using the proposed methodology, it is possible to derive the macroscopic bi-modal relationships even in the extreme cases where no FCD nor LDD exist. The only requirement is to have the flow measurements and that the AVL data comes from buses operating in a mixed-lane fashion. On the other hand, in cases where both FCD and AVL are available, the proposed fusion algorithm can substantially reduce the error on the space-mean speed of cars, especially for low FCD network coverages. Furthermore, the results are relatively consistent under different public transport regimes and network configuration layouts, and, in case of the fusion method, are almost invariant to changes in the spatial distribution and the penetration rate of probe vehicles, particularly if the availability of AVL data is rather high.

One should note that other fusion algorithms might also be possible. However, without knowing the distribution of measurement errors associated with different data sources, finding the optimal fusion algorithm is a very challenging (if not infeasible) task. Therefore, we have proposed a
A pragmatic method that can acquire relatively good (i.e., close to optimal) results. In fact, using data from the city of Zurich, Switzerland, we empirically prove and validate the accuracy of the proposed methodology, presenting a cross-comparison with the existing estimation methods. Such empirical comparison is, to the best of our knowledge, the first of its kind. Results unveil that: (i) the proposed AVL-based method can provide a good overall estimation of the average speed of cars, with a relatively low error, even with the limited availability of the public transport data; (ii) there was no problem in deriving the (3D-)MFD using the proposed fusion method, showing that it is applicable in real-life traffic scenarios; and (iii) by fusing the traffic information from different data sources (i.e., FCD and AVL) where available, especially in case of low FCD network coverage, it is possible to obtain a more representative outcome regarding the performance of multi-modal traffic, in comparison to using a single, FCD or AVL data source.
Urban networks are characterized by a multitude of interactions between different transportation modes. Controlling and assessing the capacity of such multi-modal systems is challenging, as many aspects need to be considered, e.g. road space allocation, mode prioritization schemes, traffic signal operations, passenger occupancy, etc. While buses, for example, may carry more passengers, increasing thereby the passenger throughput for a given road, they might also reduce the vehicular capacity due to their: (i) reduced cruising speeds and maneuverability, and (ii) dwell time at public transportation stops. In other words, buses can behave as moving bottlenecks, stationary bottlenecks, or a combination of both. Moreover, even if buses were to move at the same speed of cars and their operations were defined by a timetable, their headways and dwell times would still be subject to stochasticity. Such effects have not been analyzed at the network level.

In this chapter, we close this gap and provide a VT-based approximation of the multi-modal p-MFD. The proposed approximation extends the (3D-)MFD estimation analysis from the previous chapter, allowing one to derive the macroscopic traffic models for the car mode where no data on car traffic exists. This makes the proposed method very appealing for modeling purposes as we show with the application example. The remainder of this chapter is mainly based on the following publication:

4.1 General Methodology

In this section, we first give a brief overview of VT and the method of cuts applied to the MFD estimation in a deterministic environment. Then, we present a probabilistic method used to extend the concept of VT and account for stochastic bus operations (with the introduction of the stochastic shortest path) and passenger dynamics. Finally, we explain how to address the impact of the traffic conditions using an iteration-based algorithm. For the readers’ convenience, Table 4.1 provides the list of the most important notation used in this chapter.

4.1.1 Uni-modal MFD in a deterministic environment

The variational approach for traffic flow theory uses a shortest path (or least cost) formulation to solve complex kinematic wave problems given a concave flow-density relationship. A complete description of VT and the method of cuts can be found in (18, 19), (107), and (13), respectively. For the reader’s convenience, we review some fundamental properties.

The basic concept of VT can be explained with the moving observer principle. Let $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ be a directed VT graph with nodes denoted as $i \in \mathcal{V}$ and edges represented as $(i, j) \in \mathcal{A}$. Index time intervals by $\tau \in \mathcal{T}$, where $|\mathcal{T}|$ is the total number of time intervals. The time and space resolution are defined by $\varepsilon$ and $\delta$, respectively. The position of each node $i$ is given by its coordinates $(t(i), x(i))$. Let $\{N(o) : o \in \mathcal{O}\}$ be the cumulative vehicle count on the boundary nodes $o \in \mathcal{O}$. To determine the cumulative number of vehicles that can pass a given node $m \in \mathcal{V} \setminus \mathcal{O}$, $N(m)$, we have to evaluate all valid paths\(^1\) of an observer moving with trajectory $x(t)$ from any boundary node $o$ to node $m$ in the VT graph $\mathcal{G}$. Each of these paths indexed by $p \in \mathcal{P}_m$, where $|\mathcal{P}_m|$ is the total number of valid paths from $o$ to $m$, has a total cost $z_p(m)$. This total cost is just the sum of the individual cost $C(i,j)$ of each edge belonging to path $p$. The individual cost of each edge depends on the road’s cost function $r(v, t)$. This cost function is a property of the road, same as the fundamental diagram. Physically, it represents the maximum rate at which traffic can pass an observer moving with speed $v$ at time $t$. The cumulative vehicle number $N(m)$ at any given

---

\(^1\) A path is called valid if it is a piece-wise differentiable curve in the time-space plane with slopes $v$ being within the range of extremal speeds, i.e. $v \in [w, u]$. $u$ and $w$ stand for the free-flow and the backward wave speed, respectively.
VT network

- $\mathcal{G}$: directed VT graph with nodes indexed by $i \in \mathcal{V}$ and edges indexed by $(i,j) \in \mathcal{A}$
- $\mathcal{O}$: set of boundary nodes indexed by $o$
- $\mathcal{T}$: set of time intervals indexed by $\tau$
- $\delta/\epsilon$: space/time resolution of graph $\mathcal{G}$
- $L$: length of the hyperlink
- $\eta$: number of lanes along the hyperlink
- $\bar{\ell}$: average distance between bus stops along the hyperlink

(Macroscopic) fundamental diagram

- $k$: vehicular density
- $r$: vehicular cost function
- $u/w$: free-flow/backward wave speed
- $q^c_*/Q^*$: maximum car/passenger flow seen by a stationary observer
- $R(\vartheta)$: passing rate of moving observer with long-term average speed $\vartheta$

Stochastic shortest path

- $\mathcal{I}_m$: set of predecessor nodes for node $m$
- $\mathcal{P}_m$: set of valid paths from the boundary $\mathcal{O}$ to node $m$
- $z_p(m)$: total cost of path $p$ from the boundary $\mathcal{O}$ to node $m$
- $N(m)$: cumulative number of vehicles that can pass node $m$ (i.e. label on node $m$
- $C(i,m)$: state-cost vector for edge $(i,m)$
- $\Upsilon(i,m)$: state-probability vector for edge $(i,m)$
- $Z(m)$: path-cost vector for node $m \in \mathcal{V}$
- $\Phi(m)$: path-probability vector for node $m \in \mathcal{V}$

Bus operations

- $\mathcal{S}$: set of bus stops indexed by $s$ along the hyperlink
- $B_s$: set of buses indexed by $b$ using stop $s$
- $v_b$: bus space-mean speed
- $n_b$: number of buses in operation (i.e. bus accumulation on the hyperlink)
- $\bar{\xi}$: average number of stops a bus travels during a given interval
- $f_{a,b}/f_{DW}$: conditional probability density function for the arrival/dwell time of bus $b$ at stop $s$
- $\mu_{HW}/\mu_{DW}$: average bus headway/dwell time
- $\sigma_{HW}/\sigma_{DW}$: standard deviation of the bus headway/dwell time

Passenger dynamics

- $\Gamma(\tau)$: total number of on-board passengers at the end of interval $\tau$
- $\Lambda(\tau)$: total number of boarding passengers during interval $\tau$
- $\Omega(\tau)$: total number of alighting passengers during interval $\tau$
- $\gamma_b(\tau)$: average bus occupancy at the end of interval $\tau$
- $\lambda(\tau)$: arrival rate of bus passengers during interval $\tau$
- $\omega(\tau)$: total number of passengers who cannot board the bus by the end of interval $\tau$
- $\bar{l}'$: average trip length of bus passengers

Table 4.1: Nomenclature.
node \( m \) is then given by the most restricting moving observer path, i.e. the path with the least cost:

\[
N(m) = \min_{o \in O, p \in \mathcal{P}_m} \{N(o) + z_p(m)\}, \tag{4.1}
\]

with

\[
z_p(m) = \sum_{(i,j) \in \mathcal{A}_{p \rightarrow m}} \mathcal{C}(i,j) \tag{4.2}
\]

\[
= \sum_{(i,j) \in \mathcal{A}_{p \rightarrow m}} \int_{t(i)}^{t(j)} r(v(i,j), t) \, dt, \tag{4.3}
\]

where \( \mathcal{A}_{p \rightarrow m} \subset \mathcal{A} \) is the set of edges along a valid path \( p \in \mathcal{P}_m \) from \( o \) to \( m \); \( v(i,j) \) is the slope of a given edge. \((107)\) showed that, in the absence of moving bottlenecks, every path \( p \) with slope \( v_p \notin \{w, u\} \) can be substituted with a combination of sub-paths that have slope \( v'_p \in \{w, u\} \). This property reduces the number of paths that need to be considered to a finite, yet still large number. Furthermore, paths having some overlap with a shortcut (i.e. bottleneck) always have lower costs than paths that do not overlap with a bottleneck if connecting the same nodes. In other words, the moving observer in the sense of VT tries to minimize the cost between two nodes, which can also be formulated as a maximization of the time spent on shortcuts \((107)\).

One application of VT is a uni-modal MFD estimation of a hyperlink, defined as a series of successive links with traffic signals and homogeneous traffic conditions \((15)\). Red signal phases correspond to point bottlenecks with fixed location and limited time duration. They are modeled in VT as shortcuts with zero cost. In other words, horizontal edges \((i,j) \in \mathcal{A} \) overlapping with red signal phases in the VT graph are assigned the cost \( \mathcal{C}(i,j) = 0 \). To obtain the full relationship between the average flow and density for a given hyperlink, \((13)\) considered an initial value problem with a periodic initial density with average \( k \). They proved that an MFD function exists and is concave, and is given by:

\[
q = \inf_{\bar{\sigma}} \{ k\bar{\sigma} + R(\bar{\sigma}) \}, \tag{4.4}
\]

with

\[
R(\bar{\sigma}) = \lim_{T_p \to \infty} \inf_p \{ z_p(m) : v_p = \bar{\sigma} \} / T_p. \tag{4.5}
\]

The MFD function \( q = Q(k) \) according to Eq. 4.4 consists of two parts for each \( \bar{\sigma} \): (i) the \( y \)-intercept \( R(\bar{\sigma}) \) on the flow-density plane, representing
the minimum cost (i.e. passing rate) of the moving observer with long-term average speed \( \bar{v} \); and (ii) the density-dependent part \( k\bar{v} \), representing a line with slope \( \bar{v} \). These two parts are commonly referred to as a cut in the flow-density plane. By generating these cuts for different average speeds \( \bar{v} \in [\bar{w}, \bar{u}] \) and taking the lower envelope of all the generated cuts, we can obtain a concave MFD. However, this formally requires that the time duration of paths \( T_p \) approaches infinity. An alternative approach to considering an infinite periodic hyperlink was proposed in (15), which switches the observer path to the upstream end of the hyperlink once it reaches the downstream end (as in a ring road). The authors showed that an accurate \( R(\bar{v}) \) can be obtained by: (i) estimating \( R(\bar{v}) \) on a sufficient number of well distributed time windows (in most cases lower than ten) whose duration is \( L/\bar{v} \), where \( L \) is the length of the hyperlink; and (ii) taking the mean value for \( R(\bar{v}) \) across all time windows. In this dissertation, a value of ten is also adopted for the minimum number of time windows.

### 4.1.2 Multi-modal p-MFD in a stochastic environment

So far, we have only discussed how VT can be applied for the MFD estimation in uni-modal networks with deterministic signal settings, where red signal phases are considered as point bottlenecks with fixed location and limited time duration. To obtain a multi-modal MFD, in particular a multi-modal p-MFD, we have to account for the impact of the public mode of transport. In case of deterministic bus operations, this impact can be easily incorporated by adding edges (between bus stops in the VT graph) corresponding to the speed of buses acting as moving bottlenecks, and treating bus stop locations also as fixed point bottlenecks during the dwell time. However, in most cases, both the bus arrival time and the duration of the dwell time are subject to stochasticity. In addition, bus arrivals might be dependent on each other. Therefore, it is necessary to extend the concept of VT to account for the stochastic nature of bus operations, dependency of bus arrivals, and passenger dynamics. We show in this section how this can be done by incorporating a stochastic shortest path and a macroscopic passenger model into the VT framework.

#### 4.1.2.1 Incorporating stochastic bus operations: Stochastic shortest path

In the past few decades, numerous methods have been proposed to solve a shortest path problem, which has received significant attention in the operation research community. The pioneering work was conducted by (108),
who investigated the shortest path between a source node and all other nodes in a graph using a labeling approach. In contrast to path-based approaches, labeling does not require calculation of every possible path separately. Thus, the labeling concept is still widely used and represents the base for many shortest path applications. Aiming to provide more realistic formulations, some researchers have applied the shortest path algorithms to stochastic environments (109–113), providing a framework to incorporate the uncertainty and correlation in the arc costs, as well as the probability distributions for successive links (114–116). Recent work by (117) analyzed the solution for the stochastic shortest path (SSP) problem by splitting it into two parts: transforming the SSP problem into a set of deterministic problems and solving them using the (108) approach. Around the same time, (118) proposed an algebraic method to solve the shortest path problem for networks with uncertain arcs. The authors showed that the method provides the exact distribution for the SSP problem, as well as the upper and lower bounds for approximating distributions. The method assigns a distinctive probability distribution and the corresponding costs to each arc in a graph. As such, it can be extended and applied for the purpose of computing the SSP within the context of VT. In the following, we explain how this can be done with the proposed SSP algorithm. For more details on the algebraic structure of the problem, readers are referred to (119).

In a given VT graph \( \mathcal{G} \), we can select a boundary node \( o \in \mathcal{O} \), iterate through all nodes \( m \in \mathcal{V} \setminus \mathcal{O} \) that do not lie on the perimeter of the graph in a topological order, and store the cost of reaching node \( m \) in label \( N(m) \). If all immediate predecessors for a selected node \( m \) are already processed and have a correct label \( N \), a topological order of nodes is given. The topological ordering of nodes along an urban corridor assumes that nodes are indexed in such a way that if \((i, m) \in A\) then \( i < m \). Therefore, once we determine the label for a given node \( m \), we do not go back to any node indexed lower than \( m \). In other words, the label of a node depends only on the labels of the predecessor nodes and the costs of the connecting edges. The set of immediate predecessors for node \( m \) is defined as \( I_m = \{ i : (i, m) \in A \} \), where \( i \) defines the predecessor node and \((i, m)\) is the connecting edge. It is worth mentioning that, similarly to (107), we use a lopsided network for our VT graph \( \mathcal{G} \) and assume that space rows (hence horizontal edges) exist only at traffic signal and bus stop locations. In other words, each node in the graph has three (four in the presence of moving bus bottlenecks) incoming and outgoing edges, whereas nodes on
the perimeter of the graph can have fewer than three (four) edges, both incoming and outgoing.

In a deterministic environment, the label $N(m)$ of node $m$ would be determined as $\min_{i \in I_m} \{N(i) + C(i, m)\}$. This gives a single value for the cost of reaching destination node $m$, although this can be achieved through multiple paths with the same (minimum) cost. This, however, is not the case in a stochastic graph, given that each edge $(i, j)$ can have different costs associated with the corresponding probabilities of occurrence. For the problem analyzed in this chapter, an edge can be only in one of two different states: (i) bottleneck state, with the cost $c(i, j)$ and the probability $\psi(i, j)$; and (ii) non-bottleneck state, with the cost $c'(i, j)$ and the probability $\psi'(i, j) = 1 - \psi(i, j)$.

To simplify notation, we divide edges according to three types: (i) horizontal edges $A_{\text{STOP}}$ in space rows corresponding to bus stop locations, with $\psi(i, j) \geq 0$; (ii) edges $A_{\text{MOV}}$ of the VT graph corresponding to moving bus bottlenecks (if present), with $\psi(i, j) \geq 0$ and slope $v_{\text{MOV}} \in (0, u)$; and (iii) deterministic edges $A_{\text{DET}}$ (including horizontal edges in space rows corresponding to traffic signal locations and non-horizontal edges that do not represent moving bus bottlenecks), with $\psi(i, j) = 0$ or $\psi(i, j) = 1$ depending on the slope of the edge and the signal settings. Overall, sets $A_{\text{STOP}}$, $A_{\text{MOV}}$, and $A_{\text{DET}}$ are mutually exclusive and collectively exhaustive.

Therefore, to calculate the cost of reaching any given destination node in such a stochastic graph, it is necessary to use the following state-cost vector $C(i, j)$ and state-probability vector $\Psi(i, j)$ for any given edge:\footnote{The notation $[\cdot]^T$ in this chapter is used to indicate the transpose of a matrix.}

\begin{align}
C(i, j) &= \begin{cases} 
[c(i, j), c'(i, j)]^T, & \text{if } (i, j) \in (A_{\text{STOP}} \cup A_{\text{DET}}), \\
[c(i, j), c'(i, j)]^T, & \text{if } (i, j) \in A_{\text{MOV}}, 
\end{cases} 
\tag{4.6a} \\
\Psi(i, j) &= \begin{cases} 
[\psi(i, j), \psi'(i, j)]^T, & \text{if } (i, j) \in (A_{\text{STOP}} \cup A_{\text{DET}}), \\
[\psi(i, j), \Psi^T(i, j)]^T, & \text{if } (i, j) \in A_{\text{MOV}}.
\end{cases} \tag{4.6b}
\end{align}

Notice that the non-bottleneck cost $c'(i, j)$ of a moving bus shortcut $(i, j) \in A_{\text{MOV}}$ corresponds to the minimum cost of traveling from $i$ to $j$ as if edge $(i, j)$ did not exist. In case all valid paths $P_{ij}$ from $i$ to $j$ contain only deterministic edges, the non-bottleneck cost of a moving bus shortcut $(i, j)$ can be obtained as $c'(i, j) = \min_{p \in P_{ij}} \{\sum_{(m_1, m_2) \in A_{p ightarrow j}} \int_{t(m_1)}^{t(m_1)} r(v(m_1, m_1), t) \, dt\}$. In other words, it is the solution of the deterministic shortest path between
i and j. If, however, some of the edges along any path \( p \in \mathcal{P}_{i,j} \) are stochastic, then all possible combinations of the path cost across all possible state combinations (bottleneck or non-bottleneck) of these stochastic edges need to be evaluated. In other words, the non-bottleneck cost of a moving bus shortcut \((i, j)\) is the solution of the SSP between \( i \) and \( j \), in which case it becomes a vector \( C'(i, j) \), containing all combinations of the path cost from \( i \) to \( j \). Consequently, the non-bottleneck probability also becomes a vector \( \Psi'(i, j) \), containing the corresponding probabilities of occurrence for all combinations of the path cost in vector \( C'(i, j) \).

Before we explain the procedure for the proposed SSP algorithm, let us first derive the bottleneck probability \( \psi(i, j) \) for bus shortcuts. For a given hyperlink, index stops by \( s \in \mathcal{S} \), where \(|\mathcal{S}|\) is the total number of stops\(^3\). Let \( \mathcal{B}_s \) be the set of buses using stop \( s \), indexed by \( b \in \mathcal{B}_s \). Denote the subset of bus stop shortcuts and the subset of moving bus shortcuts that start at the location of bus stop \( s \) by \( \mathcal{A}_{STOP,s} = \{(i, j) \in \mathcal{A}_{STOP,s} : x(i) = x_s\} \) and \( \mathcal{A}_{MOV,s} = \{(i, j) \in \mathcal{A}_{MOV} : x(i) = x_s\} \), respectively. If we further denote by \( M_{s,b}(i,j) \) the event that bus \( b \) is present on edge \((i,j) \in (\mathcal{A}_{STOP,s} \cup \mathcal{A}_{MOV,s})\), we can compute the bottleneck probability of that edge using the following formulation:

\[
\psi(i, j) = \Pr \left( \bigcup_{b \in \mathcal{B}_s} M_{s,b}(i,j) \right) = 1 - \Pr \left( \bigcap_{b \in \mathcal{B}_s} M_{s,b}'(i,j) \right),
\]

with

\[
M_{s,b}(i,j) = \begin{cases} 
\{t_{s,b}^{ARR} \leq t(i), T_{s,b}^{DW} \geq t(j) - t_{s,b}^{ARR}\}, & \text{if } (i,j) \in \mathcal{A}_{STOP,s}, \\
\{t_{s,b}^{ARR} \leq t(m), t(m) - t_{s,b}^{ARR} \leq T_{s,b}^{DW} \leq t(i) - t_{s,b}^{ARR}\}, & \text{if } (i,j) \in \mathcal{A}_{MOV,s},
\end{cases}
\]

where \( t_{s,b}^{ARR} \) and \( T_{s,b}^{DW} \) stand for the arrival time and the duration of the dwell time of bus \( b \) at stop \( s \), respectively; \( \{m \in \mathcal{I}_i : x(m) = x(i) = x_s\} \) is the predecessor node for node \( i \), positioned at location \( x_s \); \( M_{s,b}'(i,j) \) is the complement of event \( M_{s,b}(i,j) \). Fig. 4.1 illustrates the variability in both the bus arrival time and the duration of the dwell time. Given that the dwell time starts only when a bus arrives at a bus stop location, for every arrival time we have a distribution of the dwell time (shown as gray-colored line in Fig. 4.1). Then, the bottleneck probability for a bus stop shortcut is defined as the probability that a bus arrives any time before the starting time coordinate and departs any time after the ending time coordinate of that shortcut.

\(^3\) The notation \(|\cdot|\) in this chapter is used to denote the number of elements in a vector.
Similarly, we define the bottleneck probability for a moving bus shortcut (i.e. moving bus shortcut is active) as the probability that a bus arrives any time before the starting time coordinate and departs any time before the ending time coordinate of the bus stop shortcut that precedes the considered moving bus shortcut (Fig. 4.1b). Within an empirical context, the aforementioned distributions can be obtained, e.g. using historical automatic vehicle location data from public transport or by recording the arrival time and the dwell time of buses at a given stop during a sufficiently long period of time. In this dissertation, we use the former approach for the purpose of conducting the empirical analysis (Section 4.2).

![Diagram of bottleneck probability for bus stop shortcuts](image1)

![Diagram of bottleneck probability for moving bus shortcuts](image2)

**Figure 4.1:** Modeling the bottleneck probability in case of: (a) bus stop shortcuts; (b) moving bus shortcuts.

To compute the bottleneck probability \( \psi(i,j) \), we have to account for all possible combinations of the arrival and dwell times across all buses that serve a given stop. Taking into account that the travel times of buses are
related to the dwell times, which are, on the other hand, a function of the bus headway, the arrival times of buses might be dependent. Therefore, for the purpose of computing $\psi(i,j)$, we use conditional probability density functions for both the arrival time and the duration of the dwell time of bus $b$ at stop $s$, denoted as $f_{s,b}^{\text{ARR}}(t_{s,b}^{\text{ARR}} | t_{s,1}^{\text{ARR}}, ..., t_{s,b-1}^{\text{ARR}}, T_{s,b-1}^{\text{DW}}, ..., T_{s,b}^{\text{DW}})$ and $f_{s,b}^{\text{DW}}(T_{s,b}^{\text{DW}} | t_{s,1}^{\text{ARR}}, ..., t_{s,b}^{\text{ARR}}, T_{s,1}^{\text{DW}}, ..., T_{s,b-1}^{\text{DW}})$, respectively. Note that these conditional probability density functions capture the correlation, i.e. dependency of bus arrival times. Then, the bottleneck probability of a bus shortcut $(i,j) \in (A_{\text{STOP},s} \cup A_{\text{MOV},s})$ that starts at a bus stop location can be obtained using Eq. 4.9.

$$\psi(i,j) = 1 - \left( \int_{-\infty}^{\infty} f_{s,1}^{\text{ARR}}(t_{s,1}^{\text{ARR}}) \int_{0}^{\infty} g_{s,1}(T_{s,1}^{\text{DW}} | t_{s,1}^{\text{ARR}}) \right) \cdots$$

$$\times \int_{-\infty}^{\infty} f_{s,b|B_s}(t_{s,b|B_s}^{\text{ARR}} | t_{s,1}^{\text{ARR}}, ..., t_{s,b-1}^{\text{ARR}}, T_{s,1}^{\text{DW}}, ..., T_{s,b-1}^{\text{DW}})$$

$$\times \int_{0}^{\infty} g_{s,b|B_s}(T_{s,b|B_s}^{\text{DW}} | t_{s,b|B_s}^{\text{ARR}}) \cdots dT_{s,b|B_s}^{\text{DW}} dt_{s,b|B_s}^{\text{ARR}} \cdots dT_{s,1}^{\text{DW}} dt_{s,1}^{\text{ARR}}, \quad (4.9)$$

with

$$g_{s,b}(\cdot) = \begin{cases} 0, & \text{if } t_{s,b}^{\text{ARR}} \leq t(i), T_{s,b}^{\text{DW}} \geq t(j) - t_{s,b}^{\text{ARR}}, (i,j) \in A_{\text{STOP},s}, \\ 0, & \text{if } t_{s,b}^{\text{ARR}} \leq t(m), t(m) - t_{s,b}^{\text{ARR}} \leq T_{s,b}^{\text{DW}} \leq t(i) - t_{s,b}^{\text{ARR}}, (i,j) \in A_{\text{MOV},s}, \\ f_{s,b}^{\text{DW}}(T_{s,b}^{\text{DW}} | t_{s,1}^{\text{ARR}}, ..., t_{s,b}^{\text{ARR}}, T_{s,1}^{\text{DW}}, ..., T_{s,b-1}^{\text{DW}}), & \text{otherwise.} \quad (4.10) \end{cases}$$

Similarly to Eq. 4.7, Eq. 4.9 is computed as the complementary probability of the intersection of non-bottleneck states across all buses. Recall that the state of a bus shortcut is defined by the combination of the bus arrival and dwell times (see Eq. 4.8). Thus, to exclude combinations in Eq. 4.9 under which buses act as bottlenecks, we set function $g_{s,b}(\cdot)$ (Eq. 4.10) to be zero if the conditions given by Eq. 4.8 for a given edge are satisfied.

Finally, the bottleneck probability for all moving bus shortcuts that do not start at a bus stop location is the same as the one of the preceding moving bus shortcut, i.e. $\{\psi(i,j) = \psi(m,i) : (m,i) \in A_{\text{MOV}}, (i,j) \in A_{\text{MOV}}, x(i) \neq x_s, \forall s \in S\}$.

The computed (non-)bottleneck probabilities are now used as an input to the SSP algorithm for calculating the labels of all nodes. The proposed SSP procedure is shown in a form of a pseudo code below (Algorithm 2). Note that, in contrast to the deterministic case where each node label has a unique value, in the proposed SSP algorithm a node label is a matrix.
\( N(m) = [Z(m), \Phi(m)] \), containing the information about all possible combinations of the path cost \( Z(m) \), along with the corresponding probabilities of occurrence \( \Phi(m) \):

\[
Z(m) = [z_1(m), \ldots, z_{|P_m|}(m)]^T, \quad (4.11)
\]

\[
\Phi(m) = [\phi_1(m), \ldots, \phi_{|P_m|}(m)]^T, \quad (4.12)
\]

where \( P_m \) is the set of all combinations of the path cost to node \( m \). Note that \( \sum_{p \in P_m} \phi_p(m) = 1 \), given that all possible combinations of the path cost are represented in matrix \( N(m) \). One might expect that as the size of the graph increases (i.e. \(|\mathcal{V}|\) increases), the size of \( P_m \) also increases. However, considering that there are only three to four (depending on the presence of moving bus bottlenecks) types of edges with a distinctive cost, at each stage of the topological processing of nodes there are many combinations of the path cost with the same value, whose probabilities can therefore be combined. Additionally, all combinations of the path cost associated with zero probability can be removed. This way, it is possible to limit \(|P_m|\) to a rather small size, drastically reducing the computational complexity of the problem.

The following steps illustrate the procedure of the proposed SSP algorithm shown below. Each step has the corresponding line number in the pseudo code.

---

**Algorithm 2: Stochastic shortest path (SSP)**

**Input:** Acyclic directed VT graph \( G = (\mathcal{V}, \mathcal{A}) \)

**Output:** Label vector \( N(m), \forall m \in \mathcal{V}\setminus\mathcal{O} \)

1. \( N(o) = \{[z_1(o), \phi_1(o)] : z_1(o) = 0, \phi_1(o) = 1\}, \forall o \in \mathcal{O} \)

   for \( m \in \mathcal{V}\setminus\mathcal{O} \) do

   2. \( \mathcal{I}_m = \{i : (i, m) \in \mathcal{A}\} \)

   for \( i \in \mathcal{I}_m \) do

   3. \( N_i^*(m) = [Z_i^*(m), \Phi_i^*(m)] \)

4. \( N^{**}(m) = \{[Z^{**}(m), \Phi^{**}(m)] = f(\{N_i^*(m) : i \in \mathcal{I}_m\})\} \)

5. \( N(m) = [Z(m), \Phi(m)] = f(N^{**}(m)) \)

---

**Step 1. Initialize** - Given all the inputs, the label of any boundary node \( o \in \mathcal{O} \) is defined to have the probability \( \phi_1 = 1 \) with zero cost \( (z_1 = 0) \), i.e. \( N(o) = [0, 1] \), as it has no predecessor. In other words, as in the case of (13), we assume to start from an empty hyperlink. Note that all our boundary
nodes have time coordinate \( t(o) = 0, \forall o \in O \).

**Step 2. Find the set of predecessor nodes** - For each node \( m \in V \setminus O \), the set of predecessor nodes is generated. As explained before, the size of this set is between one and four, depending on the location of \( m \) within the VT graph and the presence of moving bus bottlenecks.

**Step 3. Compute dummy (temporary) labels of the current node based on each predecessor node** - Once the incoming edges \( \{(i,m) : i \in I_m\} \) for a given node \( m \) are identified, the path-cost vector \( Z(i) \) and the path-probability vector \( \Phi(i) \) of each predecessor node \( i \) are combined with the state-cost vector \( C(i,m) \) and the state-probability vector \( \Psi(i,m) \) of the corresponding connecting edge \((i,m)\), respectively. This operation computes the dummy (temporary) labels of node \( m \) based on each predecessor individually, and stores the results as \( \{N^*(i)(m) = [Z^*(i)(m),\Phi^*(i)(m)] : i \in I_m\} \). The mathematical formulation for this process is given below:

\[
Z^*(i)(m) = (Z(i) \otimes 1_{\pi(C(i,m)) \times 1}) + (1_{|P_i| \times 1} \otimes C(i,m)),
\]

\[
\Phi^*(i)(m) = \Phi(i) \otimes \Psi(i,m),
\]

where \( \otimes \) is the Kronecker product of two matrices; \( 1_{|P_i| \times 1} \) denotes an \( |P_i| \times 1 \) all-ones matrix; \( \pi(C(i,m)) \) is the number of elements in the state-cost vector \( C(i,m) \). As we stated before, for moving bus bottlenecks \( \pi(C(i,m)) \geq 2 \); for all other edges of the VT graph \( \pi(C(i,m)) = 2 \).

**Step 4. Determine the intermediate label of the current node** - Once all the dummy labels of a given node \( m \) are computed, they need to be integrated into the intermediate node label \( N^{**}(m) \). This process depends on the number of predecessor nodes. If there is a single predecessor node (i.e. \( |I_m| = 1 \)), the intermediate label of the current node corresponds to the dummy label computed based on that predecessor node, i.e. \( N^{**}(m) = N^*_i(m) \). If there is more than one predecessor node (\( |I_m| > 1 \)), all possible combinations \( P^*_m = \{p : p = 1,...,\prod_{i \in I_m} |P_i|\} \) of the path cost across all dummy labels should be considered. For this purpose, we create the following dummy (temporary) path-cost matrix \( Z^*(m) = (Z^*_p(m)) \in \mathbb{R}^{|P^*_m| \times |I_m|} \), where the number of rows corresponds to the number of all possible combinations of the path cost (i.e. \( p = 1,...,|P^*_m| \)), whereas the number of columns corresponds to the number of the predecessor nodes (i.e. \( j = 1,...,|I_m| \)). Note that Eq. 4.15 includes also the case of a single predecessor node for completeness.
Then, to obtain entries $z^*_p(m)$ of the path-cost vector $Z^*(m) = (z^*_p(m)) \in \mathbb{R}^{|P^*_m| \times 1}$ contained in the intermediate node label, we find the minimum cost across all columns $j \in [1, |I_m|]$ of matrix $Z^*(m)$, for each combination $p$ of the path cost (Eq. 4.16).

$$z^*_p(m) = \min_j \{z^*_j(m)\}.$$

$(4.16)$

$Z^*(m)$ gives us the set of possible minimum costs to reach a given node $m$. Each combination $p$ of the path cost has the corresponding probability of occurrence $\phi^*_p(m)$. The calculation of the path-probability vector $\Phi^*(m) = (\phi^*_p(m)) \in \mathbb{R}^{|P^*_m| \times 1}$ contained in the intermediate node label is done analogously to the previously described Step 3, i.e.:

$$\Phi^*(m) = \begin{cases} \Phi_i^*(m), & \text{if } |I_m| = 1, \\ \otimes_{i \in I_m} \Phi_i^*(m), & \text{if } |I_m| > 1, \end{cases}$$

$(4.17)$

where $\otimes_{i \in I_m} \Phi_i^*(m) = \Phi_{i_1}^*(m) \otimes \Phi_{i_2}^*(m) \otimes \cdots \otimes \Phi_{i_{|I_m|}}^*(m)$. 

61
Step 5. Determine the final label of the current node - Now that we have computed the intermediate label of a given node \( m \), we first find the path-cost vector \( Z(m) \) (Eq. 6.2) for the final node label, containing only the unique path costs from vector \( Z^{**}(m) \), i.e. \(|P_m| \leq |P^*_m|\). Then, we add up the probabilities of all combinations from vector \( \Phi^{**}(m) \) that have the same cost in vector \( Z^{**}(m) \), and further obtain the path-probability vector \( \Phi(m) \) (Eq. 4.19) for the final node label.

\[
Z(m) = (z_p(m)) \in \mathbb{R}^{|P_m| \times 1} = \text{unique}(Z^{**}(m)),
\]

\[
\Phi(m) = \left\{ (\phi_p(m)) \in \mathbb{R}^{|P_m| \times 1} : \phi_p(m) = \sum_{p^* \in P^*_m: z_{p^*}(m) = z_p(m)} \phi^{**}_{p^*}(m) \right\},
\]

where function "unique" returns the distinctive elements of a given vector. As explained before, the size of a node label would grow tremendously by individually storing every combination of the path cost and its corresponding probability. Therefore, without loss of information, we undertake Step 5 to substantially reduce the memory requirements and the number of combinations of the path cost that need to be considered for the remaining nodes in the time-space plane.

The solution of the proposed SSP algorithm for a given directed acyclic graph \( \mathcal{G} \) returns the labels of all nodes in the graph. As stated above, these labels contain the information about all possible path costs to each respective node along with the corresponding probabilities. Given that graph \( \mathcal{G} \) is stochastic, there is no unique solution for the shortest path problem as in a deterministic case. Nevertheless, using the proposed approach we can choose, e.g. the expected path cost. More information about this will be given later in Section 4.2.

4.1.2.2 Incorporating passenger dynamics: Macroscopic passenger model

In order to improve mobility in a given network, we are interested in maximizing the passenger throughput. Common applications of VT for unimodal traffic consider, up to now, only the cost function that is based on the vehicular flow (see 13, for more details). This is because the number of passengers served can essentially be approximated with the number of served vehicles, i.e. the average car occupancy is \( \approx 1 \). Considering, however, that in most metropolitan areas traffic is multi-modal and that different modes can carry different number of passengers, one should consider the cost function not only as a property of the road/network, but also as a function of the transportation mode. In this section, we explain the derivation
of the passenger cost function used to estimate p-MFDs for multi-modal hyperlinks.

Let \( Q^s_*(t) \) and \( Q^s_*'(t) \) be the maximum passenger flow during the bottleneck and non-bottleneck state at stop \( s \) and time \( t \), respectively. Without loss of generality, we can also assume that the maximum passenger flow during a non-bottleneck state is independent of stop \( s \), i.e. \( Q^s_*' = Q^s_*', \forall s \in S \). Then, the state-cost vector of a given edge can be expressed in terms of the passenger units as in Eq. 4.20. Recall that the first element of the cost vector corresponds to the cost associated with the bottleneck state, whereas the second element gives the cost for the non-bottleneck state. As we stated before, the non-bottleneck cost for moving bus shortcuts is the solution of the (stochastic) shortest path between the corresponding nodes (SSP\(_{i \rightarrow j}\)).

\[
C(i, j) = \begin{cases} 
\infty, & \text{if } \{(i, j) \in A_{DET} : v(i, j) = u\}, \\
0, & \text{if } \{(i, j) \in A_{DET} : v(i, j) = 0, x(i) = x', \text{red}\}, \\
0, & \text{if } \{(i, j) \in A_{DET} : v(i, j) = 0, x(i) = x', \text{green}\}, \\
\int_{t(i)}^{t(j)} Q^s_*(t) \, dt, & \text{if } \{(i, j) \in A_{STOP} : v(i, j) = 0, x(i) = x_s\}, \\
\int_{t(i)}^{t(j)} Q^s_*'(t) \, dt, & \text{if } \{(i, j) \in A_{DET} : v(i, j) = w\}, \\
\int_{t(i)}^{t(j)} Q^s_*'(t) \frac{u + |w|}{u} \, dt, & \text{if } \{(i, j) \in A_{MOV} : v(i, j) = v_{MOV}\}, \\
\int_{t(i)}^{t(j)} Q^s_*'(t) \frac{u - v_{MOV}}{u} \, dt, & \text{if } \{(i, j) \in A_{MOV} : v(i, j) = v_{MOV}\}, \\
\end{cases}
\]

(4.20)

where \( x' \) denotes traffic signal location; \textit{green} and \textit{red} indicate the signal phase; infinity stands for the cost of deterministic edges \( A_{DET} \) during the state (bottleneck or non-bottleneck) associated with a zero probability\(^4\).

\(^4\) We use infinity only for the purpose of the SSP algorithm. However, any other value can be used as well, as the probability of this state is zero.
These passenger costs come from the vehicular cost function. To estimate the vehicular cost function, we follow the approach by (107) and (13), and assume a triangular fundamental diagram. In that case, the cost function turns out to be linear and characterized by the following parameters: free-flow speed $u$, backward wave speed $w$, and the maximum car flow $q_c^*$. Considering that there are only few examples of urban roads having more than one bus lane (either mixed or dedicated), we will restrict our formulation for the maximum passenger flow to a case of a single bus lane on a road, but different types of lane allocation. Nevertheless, it is trivial to incorporate a larger number of bus lanes into the proposed passenger cost function.

Let $n_b(t)$ stand for the accumulation of buses at time $t$, as defined by a public transport operator. Denote by $\beta$ a parameter that quantifies by how much one bus operating along a mixed lane reduces the maximum car flow $q_c^*$. This parameter can be computed analytically (6, 120) or estimated empirically (121). $\gamma_b$ and $\gamma_c$ denote the average vehicle occupancy for the bus and the car mode, respectively. We assume that $\gamma_c$ is relatively constant, but $\gamma_b$ can change over time, i.e. $\gamma_b = \gamma_b(t)$. More details are given later in this section. Then, $Q^*_s(t)$ and $Q^*'(t)$ for a given lane allocation layout at time $t$ can be formulated as:

$$Q^*_s(t) = \begin{cases} 
(\eta - 1)q_c^*\gamma_c, & \text{if dedicated,} \\
(\eta - \alpha_s)q_c^*\gamma_c - (1 - \alpha_s)\beta q_b(t)\gamma_c, & \text{if mixed,}
\end{cases} \quad (4.21)$$

$$Q^*'(t) = \begin{cases} 
(\eta - 1)q_c^*\gamma_c + q_b(t)\gamma_b(t), & \text{if dedicated,} \\
(\eta - 1)q_c^*\gamma_c + q_b(t)\gamma_b(t) + (q_c^* - \beta q_b(t))\gamma_c, & \text{if mixed,}
\end{cases} \quad (4.22)$$

where $q_b(t)$ is the bus flow at time $t$; $\eta$ stands for the number of lanes along the hyperlink. Notice that the maximum passenger flow during the bottleneck state at any given stop $s$ (Eq. 4.21) consists of (i) the maximum car (passenger) flow in car-only lane(s), and potentially (ii) the reduced car (passenger) flow in a mixed lane, depending on the type of bus stop used (curbside or bus bay). To account for the impact of bus operations on car traffic in a mixed lane as a function of the type of bus stop, we incorporate a binary variable $\alpha_s$ into Eq. 4.21. It takes the value of 1 in case of a curbside bus stop and zero value in case of a bus bay. On the other hand, during the non-bottleneck state, the maximum passenger flow (Eq. 4.22) consists of: (i) the maximum car (passenger) flow in car-only lane(s); (ii) the bus passenger flow in (either mixed or dedicated) bus lane; and (iii)
the reduced car (passenger) flow in a mixed lane (this term is zero for a dedicated bus lane).

The flow of buses \( q_b(t) \) can be computed using Eq. 4.23, where \( \mu_{iHW}(t) \) is the average bus headway, and \( v_b(t) \) stands for the space-mean speed of buses (including the duration of the dwell time) at time \( t \).

\[
q_b(t) = 1/\mu_{iHW}(t) = n_b(t)v_b(t)/L
\]

Note that \( q_b(t) \) might change or not as a function of time. In other words, it is possible that the public transport operator may adjust the number of buses depending on its bus and driver stock, and the traffic conditions. On one hand, the operator might want to operate a headway based schedule, i.e. \( \mu_{iHW}(t) = \mu_{iHW} \), where the number of buses running increases when traffic conditions worsen. On the other hand, the operator might be restricted by the maximum number of buses in stock, where the number of buses running remains constant, i.e. \( n_b(t) = n_b \), independent of traffic conditions. As a matter of fact, bus operators might also operate a constant headway up to certain point, and keep a constant number of buses (i.e. constant spacing) afterwards.

It is worth mentioning that using the proposed passenger cost function it is possible to examine the impact of different bus lane layouts on the shape of multi-modal p-MFDs, as well as to determine the speed threshold when a mixed lane should be switched to a dedicated bus lane and vice-versa, depending on the prevailing traffic conditions. We will elaborate more on this later in the application part, within Section 4.2.

As mentioned before, while, in most cases, the average car occupancy is assumed to be constant (e.g. 1.2 pax/veh) and invariant over time, allowing one to model car passenger dynamics as vehicular traffic dynamics, the bus occupancy is subject to the operating regime of the public transport system. We now show how the bus occupancy \( \gamma_b(t) \) can be modeled at the macroscopic level and, as such, can be incorporated into the proposed passenger cost function.

Let \( \Lambda(\tau) \) and \( \Omega(\tau) \) stand for the total number of boarding passengers and the total number of alighting passengers along a given hyperlink during time interval \( \tau \), respectively. Denote by \( \Gamma(\tau) \) the total number of on-board passengers at the end of time interval \( \tau \) (i.e. at time \( \tau\varepsilon \)). The evolution of the total bus passengers over time can be described as:

\[
\Gamma(\tau) = \Gamma(\tau - 1) + \Lambda(\tau) - \Omega(\tau),
\]

where the total number of boarding passengers is bounded by the two parameters in Eq. 5.37: the total number of (accumulated) passengers \( \Lambda_{ACC} \).
that want to enter the bus, and the total number of passengers \( \Lambda_{\text{MAX}} \) that can enter the bus:

\[
\Lambda(\tau) = \min\{\Lambda_{\text{ACC}}(\tau), \Lambda_{\text{MAX}}(\tau)\}.
\] (4.25)

Before we derive these two parameters, let us first introduce the evolution of the total number of passengers who cannot board the bus along the hyperlink by the end of time interval \( \tau \):

\[
\omega(\tau) = \omega(\tau - 1) + \lambda(\tau)\epsilon - \Lambda(\tau),
\] (4.26)

where \( \lambda(\tau) \) denotes the average arrival rate of bus passengers during time interval \( \tau \). In this chapter, we assume that this parameter is exogenously given as a result of mode choice for the associated equilibrium state (i.e. modal split). We do not explicitly consider mode choice because our focus is on the use of VT and its extensions related to incorporating stochasticity in bus operations, stochastic moving bus bottlenecks, and the correlation in bus arrival times. Nonetheless, it is worth mentioning that to account for both short- and long-term effects that the proposed dynamic allocation of a particular bus lane (see Section 4.2) might have on the modal split, there should be a feedback between the multi-modal p-MFD, travel demand, and mode choice. This could actually be incorporated into the proposed stochastic VT framework, where we can assess the preferred mode of transport of passengers based on the generalized travel costs of the two modes. For that, we could use a simple Logit model as in (35), (15), and (51).

Also note that we ignore changes in passenger arrival within a single interval, given that the interval size \( \epsilon \) is rather small, i.e. \( \lambda(\tau) \approx \lambda(\tau\epsilon) \). Thus, \( \lambda(\tau)\epsilon \) represents the total number of passengers that arrive during an interval of length \( \epsilon \). Moreover, as we do not consider mode choice, we are not interested in the waiting time of passengers who cannot board the bus, only on their number, as it might impact the bus operations in subsequent periods. The total number of passengers that want to enter the bus during time interval \( \tau \) can be obtained as follows:

\[
\Lambda_{\text{ACC}}(\tau) = (\omega(\tau - 1) + \lambda(\tau)\epsilon) \frac{n_b(\tau - 1)\bar{\zeta}(\tau)}{|\mathcal{S}|} \left[ \frac{n_b(\tau - 1)\bar{\zeta}(\tau)}{|\mathcal{S}|} \right]^{-1}, \tag{4.27}
\]

where \( \bar{\zeta}(\tau) = v_b(\tau)\epsilon/\bar{\ell} \), with \( \bar{\ell} \) being the average distance between bus stops, is the number of stops a bus travels during a given time interval\(^5\).

\(^5\) The notation \([\cdot]\) in this chapter is used to indicate a ceiling function.
\( \Lambda_{\text{ACC}} \) consists of two parts: the total demand of bus passengers waiting along the hyperlink (term 1), and the fraction of this total demand that can potentially be served (term 2 and term 3). The last term (term 3) denotes the inverse value of the maximum number of times that any given stop is visited. When combined with term 2, it accounts for the fact that a stop might be visited multiple times, and that the number of visits does not have to be identical for all stops. Therefore, only a fraction of the total generated demand can be served.

On the other hand, the total number of passengers that can enter the bus is determined by the total available capacity across all buses operating along the hyperlink (Eq. 4.28), where \( \gamma^*_b \) stands for the average bus passenger capacity.

\[
\Lambda_{\text{MAX}}(\tau) = n_b(\tau - 1)\gamma^*_b - \Gamma(\tau - 1) + \Omega(\tau).
\]  

(4.28)

The total number of passengers alighting during time interval \( \tau \), \( \Omega(\tau) \), is defined similarly to (35):

\[
\Omega(\tau) = \Gamma(\tau - 1) \left( 1 - \left( 1 - \min \left\{ 1, \frac{\bar{\ell}}{\bar{\ell}'} \right\} \right) \xi(\tau) \right),
\]  

(4.29)

where \( \bar{\ell}' \) is the average trip length of bus passengers. Note from Eq. 5.21 that every on-board passenger from the previous time interval has the probability \( 1 - \min\{1, \bar{\ell}/\bar{\ell}'\} \) of reaching its destination at each stop a bus traverses along a given hyperlink during the current time interval (35).

Now that we have derived all parameters for modeling the evolution of the total bus passengers over time, we can compute the average bus passenger occupancy at the end of a given time interval using Eq. 4.30.

\[
\gamma_b(\tau) = \frac{\Gamma(\tau)}{n_b(\tau)}.
\]  

(4.30)

We then assume that this occupancy remains constant for the whole time interval \( \tau \), i.e. \( \gamma_b(t) = \gamma_b(\tau), \forall t \in [(\tau - 1)\varepsilon, \tau\varepsilon] \). Moreover, the average dwell time for a given interval can be obtained as follows:

\[
\mu_{\text{DW}}(\tau) = T_{\text{DOOR}} + \max \left\{ \frac{\Lambda(\tau)T_{\text{BOAR}}}{n_b(\tau)}, \frac{\Omega(\tau)T_{\text{ALIG}}}{n_b(\tau)} \right\} \frac{1}{\xi(\tau)},
\]  

(4.31)

where \( T_{\text{DOOR}} \) denotes the time spent on opening and closing the doors; \( T_{\text{BOAR}} \) and \( T_{\text{ALIG}} \) are the boarding and alighting time per passenger, respectively.
In this chapter, we use a dwell time model that assumes a simultaneous boarding and alighting process (Eq. 4.31). Alternatively, one can use a dwell time model based on a sequential boarding and alighting process, in which case Eq. 4.31 becomes \( \mu_{DW}(\tau) = T_{DOOR} + (\Lambda(\tau)T_{BOAR}/n_b(\tau) + \Omega(\tau)T_{ALIG}/n_b(\tau))/\zeta(\tau) \).

Last, note that only few variables are needed to apply the proposed macroscopic passenger model (i.e. bus passenger demand, the average trip length of bus passengers, bus space-mean speed, bus accumulation, and the bus passenger capacity), making it pragmatic.

### 4.1.3 Traffic conditions and their effects on bus stop shortcuts

In general, there are two types of effects that need to be analyzed when calculating the solution of the LWR (123, 124) model for multi-modal hyperlinks: (i) the effects that the bus stops and moving bus bottlenecks have on traffic; and (ii) the effects that the traffic conditions have on the temporal distribution of bus stop bottlenecks. To capture the former effects, we use VT, i.e. the shortest path or the stochastic shortest path, depending on whether bus operations are considered deterministic or stochastic. For the latter effects, however, the (stochastic) shortest path cannot be directly applied, as the expected bus arrival time at any given stop (and potentially the dwell time) is affected by the speed of buses. In other words, the temporal distribution of bus stop shortcuts in the VT graph depends on the current traffic state. Therefore, to accurately estimate the p-MFD for such multi-modal hyperlinks, especially in case of stochastic bus operations, we have to account for the impact of the traffic conditions described by the bus cruising speed parameter. For this purpose, we propose to use the following iteration-based algorithm shown in a form of a pseudo code below (Algorithm 3). It can be seen as a feedback between the bus stops and the traffic conditions (i.e. bus speed).

The goal of the proposed algorithm is to use the stochastic VT approach shown before to find the LWR (123, 124) solution for every possible traffic state. We do this by defining a range of bus cruising speeds \( v_{CRUS} \) that have an impact on the temporal distribution of bus stop shortcuts. Each of these cruising speeds corresponds to a particular traffic state scenario indexed by \( \rho \in \mathcal{R} \), where \( |\mathcal{R}| \) is the total number of scenarios. The number of scenarios depends on the speed of the moving bus bottlenecks. Scenarios should encompass speeds \( \{ v_{CRUS}(\rho) \in [0, v_{MOV}] : \rho \in \mathcal{R}, v_{CRUS}(1) = 0, v_{CRUS}(|\mathcal{R}|) = v_{MOV} \} \) for which moving bus bottlenecks are not active, i.e.
buses are affected by traffic conditions and travel with the same speed as cars. The number of scenarios is the largest for the case where buses are not considered moving bottlenecks ($v_{\text{MOV}} = u$). Then, for each scenario, we execute the following steps, each having the corresponding line number in the pseudo code below.

**Algorithm 3: Iteration process**

```
Input: \{\nu_{\text{CRUS}}(\rho) : \rho \in \mathcal{R}\} 
Output: Q 

\text{for } \rho \in \mathcal{R} \text{ do} 
\hspace{1cm} 1 \text{ Develop } \mathcal{G}(\rho) = \mathcal{G}(\nu(\rho)) 
\hspace{1cm} 2 \quad Q(\rho) \leftarrow Q(\nu(\rho)) 
\hspace{1cm} 3 \quad \text{p-MFD} = f(\{Q(\rho) : \rho \in \mathcal{R}\})
```

**Step 1. Develop a new VT graph for a given traffic state scenario** - Considering that the temporal distribution of bus stop shortcuts depends on the bus cruising speed, a new VT graph $\mathcal{G}(\rho)$ is developed for each traffic state scenario. For each graph, we use the corresponding bus cruising speed $\nu_{\text{CRUS}}(\rho)$ to model bus stop shortcuts. Recall that for all scenarios with $\nu_{\text{CRUS}} \leq v_{\text{MOV}}$, buses are affected by congestion and travel with the same speed of cars (i.e. they do not act as moving bottlenecks). In other words, the passing rate is zero; thus the upstream and downstream traffic states with respect to the buses are the same and located on the original fundamental diagram of each link. As a result, only the temporal distribution of shortcuts at the bus stops changes across VT graphs.

**Step 2. Execute the proposed SSP algorithm to find the LWR solution for a given traffic state scenario** - Once the VT graph $\mathcal{G}(\rho)$ is developed for a given traffic state scenario $\rho$, we execute the proposed SSP algorithm to estimate the multi-modal p-MFD and find the LWR solution $Q(\rho)$ for the considered traffic state, i.e. the point on the resulting bi-modal p-MFD with the slope equivalent to $\nu(\rho)$.

**Step 3. Developing a multi-modal p-MFD that accounts for the impact of the traffic conditions** - Once the LWR solution is found for each traffic state, a multi-modal p-MFD is generated by combining LWR solutions $\{Q(\rho) : \rho \in \mathcal{R}\}$ across all traffic state scenarios.
It is worth mentioning that the proposed methodological framework allows to model complex stochastic bus operations and passenger dynamics within the VT framework, incorporating their effects on the traffic states. To the best of the authors’ knowledge, this is the first attempt to build a complete VT-based bi-modal traffic framework that accounts for all aforementioned dependencies.

4.2 Implementation and Discussion

This section discusses and shows potential implementations of our proposed analytical approach for multi-modal p-MFDs. First, we demonstrate the applicability of the proposed VT extensions using a multi-modal corridor in Zurich, Switzerland, and thereby validate, at the same time, the proposed approach with empirical data and with a Monte-Carlo simulation. Second, we investigate the value of our stochastic VT approach by comparing it with results from a deterministic VT framework. Third, we present an application for evaluating the operations of different bus lane layouts.

4.2.1 Empirical implementation and validation

In the following, we apply our proposed analytical approach to an urban multi-modal corridor in Zurich, Switzerland, using empirical data. In addition, we compare the approach with a Monte-Carlo simulation to analyze the congested branch of the multi-modal p-MFD and the performance of the SSP algorithm. Before we present the results, let us first introduce the empirical data set, the derived modeling assumptions, and the mechanisms of the Monte-Carlo simulation.

We use 5 weekdays of empirical data from October 2015. The data corresponds to a corridor in the city of Zurich, Switzerland, with one mixed lane, two signalized intersections, and three curbside bus stops (Fig. 4.2). Data on the private mode is collected by both loop detectors and floating car vehicles, whereas data on public transport is acquired from Zurich’s transit operator (VBZ). Chapter 3 contains a more comprehensive overview of the used data sets. Following the approach in (7), data is aggregated using a 30 min moving average in order to reduce some of the noise due to the overall low number of buses (8 scheduled buses per hour). Based on these data sets, we make the following four assumptions. Note that these assumptions are only used for the case study, i.e. they are independent
of the theoretical model formulations given in the methodological section. Consequently, they can be altered and adapted to the empirical conditions observed.

**Figure 4.2**: Schematic representation of the study site for empirical analysis.

**Assumption 4.1** (Arrival and dwell time distributions). For the purpose of modeling stochastic bus operations, it is assumed that both the bus arrival time and the dwell time follow a Gaussian distribution. Nevertheless, the proposed method is not restricted to a Gaussian distribution, i.e. any probability distribution can be used.

**Assumption 4.2** (Dependency of bus arrival times). Here, to emulate the specific conditions from the study site, we assume that the bus arrival times at any given stop are independent. The public transport system in the city of Zurich operates under a schedule-based control strategy, with no bus holding mechanism implemented on the investigated corridor, leading to independent arrivals. Moreover, as we will later show, the system is mostly uncongested, which also limits the dependencies on the arrivals. Given this assumption, we can further simplify Eq. 4.9 in case of bus stop shortcuts and moving bus shortcuts as Eqs 4.32a and 4.32b, respectively.

\[
\psi(i, j) = 1 - \prod_{b \in B} \left( 1 - \int_{-\infty}^{t(i)} f_{s,b}^{\text{ARR}}(t_{s,b}^{\text{ARR}}) \int_{t(j) - t_{s,b}^{\text{ARR}}}^{\infty} f_{s,b}^{\text{DW}}(T_{s,b}^{\text{DW}} | t_{s,b}^{\text{ARR}}) \text{d}T_{s,b}^{\text{DW}} \text{d}t_{s,b}^{\text{ARR}} \right) .
\]

\[
(4.32a)
\]

\[
\psi(i, j) = 1 - \prod_{b \in B} \left( 1 - \int_{-\infty}^{t(m)} f_{s,b}^{\text{ARR}}(t_{s,b}^{\text{ARR}}) \int_{t(m) - t_{s,b}^{\text{ARR}}}^{t(j) - t_{s,b}^{\text{ARR}}} f_{s,b}^{\text{DW}}(T_{s,b}^{\text{DW}} | t_{s,b}^{\text{ARR}}) \text{d}T_{s,b}^{\text{DW}} \text{d}t_{s,b}^{\text{ARR}} \right) .
\]

\[
(4.32b)
\]

**Assumption 4.3** (Bus passenger demand function). We model the bus passenger demand function according to empirical findings from the city of
Zurich, Switzerland, which indicate that the demand for public transport is strongly correlated to the car travel demand (7). In other words, the public transport and passenger accumulations exhibit a high degree of co-movement. Here, we further investigated the correlation between the two accumulations and found high correlation indices between 0.87 and 0.94 (for different zones and days in the city of Zurich). Assuming that such accumulations can be used as proxies for passenger demand, we can deduce that the car and bus demands are similar. In addition, we observe that an increase in the car demand is generally associated with higher accumulation levels, which, in turn, results in lower average speeds. Therefore, when combining these empirical observations, we model the bus passenger demand under mixed traffic conditions as inversely proportional to the space-mean speed on the hyperlink. This assumption has important and simplifying implications for the passenger model derived in Section 4.1.2.2. It allows to adjust the passenger model in a way that it relies on $\rho$, i.e. the traffic state scenario (see Section 5.1.4). This results in a time invariant demand, i.e. $\lambda(t) = \lambda(v_{crus}) = \lambda(\rho)$ for every traffic state scenario $\rho$. Thus, for every $\rho$, we can first run the passenger model and determine the average bus passenger occupancy and the average dwell time. The obtained values are then used as inputs to the passenger cost function (Eqs. 4.21–4.22) and the scenario-specific VT graph (developed according to Section 5.1.4). This simplifies the derivation of the multi-modal p-MFD.

It is clear that this assumption might not be true for all cities and eventually needs an individual validation. Nonetheless, we believe that this relatively simple first order approximation is a reasonable assumption under mixed traffic conditions. Alternatively, the passenger demand could also be defined exogeneously.

**Assumption 4.4 (Moving bus bottlenecks).** As we do not find evidence of moving bus bottlenecks on the investigated corridor, we refrain from considering this phenomenon.

We observe that the investigated bus line 31 follows a rigid timetable, which leads to a constant headway. In other words, the public transport system in this case displays a constant headway ($\mu_{HW}(t) = \mu_{HW}$) as discussed in Section 5.1.4. From empirical data, we observe an average headway of 450 s with a standard deviation of 150 s. The negative relationship between the bus passenger demand and average speeds postulated in Assumption 4.3 is estimated with a linear regression on the observed average speeds. Other important variables are: free flow speed $u = 50$ km/h; jam density of 135 veh/km; green time of 36 s; red time of 21 s; offset of 2 s;
bus passenger capacity $\gamma_b^* = 60 \text{ pax/veh}$; car occupancy $\gamma_c = 1.2 \text{ pax/veh}$; $\beta = 3$; and $\alpha_s = 1, \forall s \in \mathcal{S}$.

Fig. 4.3 shows the comparison of our stochastic VT approach, empirical data, Monte-Carlo simulation, and the deterministic VT method. The proposed stochastic VT approach is shown using a solid black line, empirical data as dots, Monte-Carlo simulation is represented with a dashed black line, whereas the deterministic VT approach is presented using a dotted black line. In addition, we fitted to the empirical data a simple polynomial regression of fourth order, shown as a gray solid line.

![Comparison of the estimated and empirical multi-modal MFD for a real corridor in the City of Zurich, Switzerland.](image)

**Figure 4.3**: Comparison of the estimated and empirical multi-modal MFD for a real corridor in the City of Zurich, Switzerland.

First thing to notice is the lack of the congested branch on the empirically derived multi-modal p-MFD. There are multiple potential reasons for this phenomenon. First, the chosen temporal aggregation of the (3D-)MFD (30 min moving average) is rather large. It is justified, however, because of the frequency of the bus service along this corridor. Given that there is only one bus line with a relatively tight schedule, an aggregation period lower than the chosen one would yield very unstable bus flows and densities along the corridor. Second, the traffic control system in Zurich is highly adaptive and well connected (for a detailed overview, we refer to (125)). It tries to cope with increasing demand by extending red times at the entries (or extending green times at the exits) of the city gates - in the spirit of a perimeter control. Last, loop detectors (our data source for the car data) have been deployed with the aim to control traffic efficiently. Thus, there exists a certain selection bias: the sensors used to detect congestion are located in places where congestion should be mitigated.
We also observe that our stochastic VT approach matches the empirical multi-modal p-MFD well. In fact, we compute a normalized root mean squared error (NRMSE) of 8.9%, which is relatively close to the NRMSE yielded by a simple regression using a polynomial of fourth order (8.1%). Such result not only validates our proposed stochastic VT approach, but it also shows its applicability on a real corridor. This suggests that the proposed framework can successfully be applied to estimate the impact of stochastic bus operations on car traffic. Furthermore, it also validates the proposed passenger model and the assumptions made therefor.

Since there is no empirical data available for the congested branch, as we stated before, we also compare the proposed stochastic VT approach with a Monte-Carlo simulation. The simulation consists of a random sampling performed for each of the pre-defined scenarios characterizing different bus arrival times and dwell times. In each scenario, the Monte-Carlo module draws the arrival time and the dwell time duration for each bus at every stop from the corresponding distributions. Hence, bus stop shortcuts can be treated the same way as the deterministic shortcuts corresponding to red signal phases, turning the SSP problem into a deterministic shortest path problem, which can be solved with the existing VT framework and the standard shortest path algorithm. This Monte-Carlo simulation serves as a benchmark for the congested regime of the multi-modal p-MFD and our SSP implementation. We observe that the stochastic VT framework and the Monte-Carlo simulation are almost perfectly aligned, both in the uncongested and the congested part of the multi-modal p-MFD. In fact, for the Monte-Carlo multi-modal p-MFD, we quantify a NRMSE of 9.4% (in relation to the empirical multi-modal p-MFD), close to (but slightly worse than) the NRMSE of the stochastic VT approach (8.9%). Moreover, we can see that the shape of the multi-modal p-MFD is reproduced by the Monte-Carlo simulation and is thus not an artifact of our SSP, but is based on the link’s short length and the reduced number of intersections. It is important to note that the Monte-Carlo simulation is computationally very expensive and inefficient compared to the proposed approach. In other words, with the proposed methodology we can obtain the same results as with a large number of Monte-Carlo simulations that can be computationally demanding, especially for long hyperlinks and/or simulation periods. For this specific example, we used 1000 runs for the Monte-Carlo simulation. This required a total of 2470 s of computational time, whereas the SSP algorithm required only 121 s of computational time (a time reduction of
Experiments were run on a 4-core Intel i7 processor with 2.6 GHz and 20 GB RAM.

It is worth mentioning that another way to address the lack of the congested branch in the MFD would be to apply the recently proposed symmetry in the kinematic wave models \((17, 126)\). By performing a simple linear transformation of density, \((53)\) showed that the congested branch becomes a mirror image of the free-flow branch. However, such symmetry procedure requires a good estimation of the speed associated with the critical density, as well as the backward wave speed, which is not available in our empirical case.

Finally, one potential question regarding the empirical analysis is whether the same solution could have been obtained with the existing (deterministic) VT approach, which does neither account for traffic conditions nor for stochasticity. We cannot, however, make a direct comparison between approaches for the empirical example, as the existing VT methodology cannot account for passenger dynamics either. Therefore, to obtain the multimodal p-MFD in case of the deterministic VT approach, we use a constant average bus passenger occupancy, which gives a NRMSE of 15.7\%, substantially higher than the one yielded by the proposed stochastic VT framework (8.9\%).

### 4.2.2 Value of the proposed analytical framework

Here, we further elaborate on the differences between the proposed stochastic framework and the deterministic VT solution for vehicular MFD. We do this by comparing two different bus scenarios and demonstrating the sensitivity of the results to the frequency and the stochasticity of bus operations. In addition, we evaluate the framework’s robustness by reporting the range of flows obtained by our stochastic extension of VT. Recall that the SSP can return for any given density not only the expected flow, but also all other potential flows including their probabilities (see the path-probability vector in Section 4.1.2.1).

The considered abstract hyperlink is composed of six intersections, each connected by a street segment. The length of the hyperlink is set to be \(L = 2000\) m. To simplify the problem, the lengths of the street segments are chosen to be identical (400 m each) with five stops, one located in the middle of each link, leading to a constant bus stop spacing of \(\bar{\ell} = 400\) m. Two scenarios with different levels of stochasticity and frequency in terms of bus operations are analyzed; each with a different combination of
the mean free-flow headway and its coefficient of variation CV\textsubscript{HW}: scenario I characterizing a short headway, high CV\textsubscript{HW}: \( \mu_{\text{HW}} = 150 \text{ s}, \sigma_{\text{HW}} = 75 \text{ s} \) (CV\textsubscript{HW} = 0.5); and scenario II characterizing a long headway, low CV\textsubscript{HW}: \( \mu_{\text{HW}} = 600 \text{ s}, \sigma_{\text{HW}} = 75 \text{ s} \) (CV\textsubscript{HW} = 0.125). In order to avoid any distortion related to passenger dynamics, we focus on the vehicular multi-modal MFD. In addition, we fix the number of running buses \((n_b(t) = n_b)\) and keep the same average dwell time with \( \mu_{\text{DW}} = 30 \text{ s} \) and \( \sigma_{\text{DW}} = 10 \text{ s} \).

Fig. 4.4 shows a comparison of the vehicular multi-modal MFDs for the car mode obtained using the proposed stochastic VT approach (including the effects of the traffic conditions and uncertainty) and the deterministic VT approach for both scenario I and scenario II. Note that for the deterministic case \( \sigma_{\text{HW}} = \sigma_{\text{DW}} = 0 \). For the MFD obtained with the proposed methodology, we also show the uncertainty range. This uncertainty as well as the discrepancy with the deterministic MFD is, not surprisingly, higher for scenario I. In both scenarios, however, the uncertainty range is neither evenly distributed over the densities nor the flows. The uncongested branch of the MFD exhibits a higher uncertainty than the congested branch, and the maximum uncertainty is not observed at capacity. More importantly, the uncertainty seems to be larger below the MFD curve. These findings have relevant implications for the modeling of the error distribution, as they are often used as inputs to estimation methods of the MFD (e.g. 44). We observe that our stochastic approach yields higher jam density and flows than the deterministic one. However, it remains unclear with which magnitude each of the proposed VT extensions (stochasticity and accounting for traffic conditions) contributes to the observed differences.

![Figure 4.4](image)

**Figure 4.4**: Differences between the stochastic and the deterministic VT approaches.

Therefore, in Fig. 4.4b, we analyze the impacts of each extension separately. The difference due to the stochasticity is obtained by comparing
the deterministic and the stochastic approach, while not adjusting for the traffic conditions in either of the two. The difference due to traffic conditions is obtained by comparing two deterministic approaches, one accounting and one not accounting for the traffic conditions. Each scenario’s box plot stands for the absolute relative difference $\Delta$ in flows, obtained over the entire range of density values. Not surprisingly, the differences due to stochasticity are higher for scenario I with a higher $CV_{m/v}$, confirming the value of our stochastic VT graph. These differences are partly explained by the fact that the deterministic approach may exacerbate certain interactions between the bus stops and the traffic signals, whereas the proposed approach is capable of relaxing such issue by introducing some variability into these interactions. Similarly, the differences generated by taking into account traffic conditions are also larger for scenario I. This confirms that, as the number of bus stop bottlenecks increases (due to the short headways), the role of traffic conditions becomes more significant. In other words, when traffic conditions worsen, the average bus headway increases due to a constant number of buses in operation (which is not captured by the deterministic VT approach), resulting in almost zero bus flow in over-saturated conditions, i.e. $q_b(t) \rightarrow 0$ for $v_b(t) \rightarrow 0$ (see Eq. 4.23). Thus, capturing the impact of the traffic conditions by adjusting the VT graph accordingly, the proposed approach reduces the number of bus stop shortcuts in the stochastic VT graph, resulting in lower cut costs (i.e. higher flows for a given car density). Interestingly, we observe that the impact of accounting for traffic conditions (related to the number of bus stop shortcuts in the VT graph) is more significant than the impact of accounting for stochasticity (related to the bottleneck probability of bus stop shortcuts), at least in these examples. Overall, we observe that if stochasticity is not taken into account, the average difference in flows can be up to 9% compared to the existing VT-based deterministic method of cuts, and if the traffic conditions are not taken into account, the average difference in flows can be up to 17%. That being said, it is clear that the chosen probability distributions (here Gaussian) have a considerable impact on such analysis. In the future, an in-depth sensitivity analysis, as that proposed in (127), could further shed light on the bounds of our stochastic VT approach, i.e. in which cases it would be reasonable to apply the deterministic VT framework instead.

It is worth noting that due to the use of exogenous bus arrival and dwell time distributions, similarly to (128), the proposed modeling framework cannot capture bus bunching dynamics. Nevertheless, the proposed macroscopic passenger model can be extended to relate the number of passen-
gers gathered at bus stops (i.e. the number of boarding passengers and further the dwell time) to the bus headway. This is a topic of future works.

4.2.3 Application example: Evaluation of different bus lane layouts

In the previous section, we established the importance of incorporating the stochasticity and the effects of traffic conditions when deriving multi-modal p-MFDs. Here, we focus on a potential application of the proposed stochastic VT framework.

In general, the implementation of a dedicated bus lane is a very effective strategy to prioritize public transport. However, it requires a substantial allocation of space (120). Therefore, it is important to investigate in which cases such prioritization scheme becomes justifiable. In the following, we analyze and compare different bus lane layouts. For this purpose, we use the previously introduced empirical corridor (Fig. 4.2) with the same settings and passenger demand, but one additional lane. Moreover, we assume a public transport operator that keeps the number of buses constant, independently of traffic conditions, i.e. $n_b(t) = n_b$.

![Figure 4.5: Comparison of the p-MFDs from a real corridor with a mixed or dedicated bus lane.](image)

Using our stochastic VT extension, we can define the multi-modal p-MFD for two layouts: cars may drive on the bus lane (i.e. mixed lane), and cars may not drive on the bus lane (i.e. dedicated bus lane). In addition, we investigate the effects of different mean headways: scenario III reflecting the existing headway in the Zurich corridor $\mu_{HW} = 450$ s, $\sigma_{HW} = 150$ s, and scenario IV reflecting a shorter headway (half the real one) with the same $CV_{HW}$, i.e. $\mu_{HW} = 225$ s, $\sigma_{HW} = 75$ s. Fig. 4.5 shows the resulting multi-modal p-MFDs for both scenarios. Not surprisingly, we observe that
for uncongested traffic conditions, a mixed lane always yields a higher passenger throughput. However, as the congestion level increases, the dedicated lane starts outperforming the mixed lane in terms of the passenger throughput. This is reasonable, as the dedicated lane can guarantee free flowing conditions for buses, which carry a higher number of passengers, even when the car lane gets into a complete gridlock. For this real corridor in Zurich, we can see that an exclusive bus lane improves passenger flows only for high car densities, which are not observed with empirical data shown in Fig. 4.3. Notice that the intersection point between the multi-modal p-MFDs represents the traffic state ($k_{CRIT}$) at which the layout should be switched from a mixed lane to a dedicated bus lane in order to improve bus operations. This is a relatively simple example of how our proposed methodology can be used to evaluate different types of bus operations and use of the infrastructure to ultimately improve the overall passenger flow in a system. That being said, further improvements could be achieved if a limited number of cars were still allowed to drive on the bus lane even for traffic conditions with densities $k < k_{CRIT}$.

Note that an accurate estimation of the multi-modal p-MFD is essential. A short analysis has shown that an under- or over-estimations of already 10% in the flows of the multi-modal p-MFD can result in a passenger flow reduction of around 10%.

4.3 SUMMARY

Recent advances in the field have shown that the multi-modal p-MFD is a powerful tool to quantify traffic dynamics in urban networks. In this chapter, we propose a methodology to estimate the p-MFD diagram for multi-modal urban corridors while accounting for the stochastic nature of bus operations. The proposed framework extends the existing VT approaches by: (i) introducing a probabilistic VT graph, where the costs are computed using an efficient stochastic shortest path algorithm; (ii) capturing the effects of stochastic moving bus bottlenecks and the correlation of bus arrival times; (iii) incorporating a macroscopic passenger model that reflects the passenger dynamics for the different modes; (iv) accounting for the effects that the traffic conditions might have on bus operations; and (v) introducing an innovative application example for the evaluation of different bus lane layouts, aiming to maximize the passenger throughput along a multi-modal urban corridor. Such application would not have been possible without the proposed VT extensions. In addition to these
contributions, we evaluate the framework’s robustness and quantify the importance of each proposed VT extension.
Public transport is often seen as a key component of a sustainable urban development, as it allows more passengers to efficiently travel across an urban area at low environmental and economic costs. Multiple factors can influence the public transport level of service. All take root in the network structure and the operating regime, i.e. how the public transport lines, in particular those for the bus system, are arranged atop the street network and how the service frequency is adjusted to meet urban mobility patterns. This is known as the bus network design problem and has been the subject of several studies. The problem is so challenging that most studies until now resort to strong assumptions such as a static description of the peak hour demand, homogeneous user behavior, and equal trip lengths. Potential effects of different types of user behavior and trip lengths patterns on the user and/or operator cost have not been investigated whatsoever. Moreover, the existing studies have neither considered the effects of complex interactions between the modes on the bus network structure nor the effects of bus network design on private car users.

In this chapter, we close this gap and provide a general framework considering multiple trip length patterns, two types of user behavior, and the effects that the bus network structure might have on the traffic performance and passenger mode choice. The remainder of this chapter is mainly based on the following publication:

5.1 General Methodology

In this section, we first describe general network settings for the bus network design problem. Then, we formulate the objective function used to determine the optimal design parameters. Finally, we derive the operating speed of each mode while taking into account the complex modal interactions, as well as the effects of network topology. For the readers’ convenience, Table 5.1 provides the list of the most important notation used in this chapter.

5.1.1 General network settings

We consider here a multi-modal, bi-directional urban network with an average trip generation rate $\Lambda$ during the loading time of the peak period, and an average trip generation rate $\lambda < \Lambda$ during the off-peak periods and unloading time of the peak period. Notice that, unlike the previous studies on the bus network design problem, we do not consider a mean value for the passenger demand over the full peak period, but distinguish between the loading and unloading phases. This is because we are interested in analyzing how congestion emerges across the network, allowing more realistic modeling of traffic conditions. That being said, the cost-terms (defined in the following subsection) are time-independent, i.e. they are estimated on an hourly base during the peak period and are given as an average of the loading and unloading phase. Also note that in this chapter we look at the bus network design problem under the steady state, when the traffic conditions are undesaturated or saturated, with no disruptions causing the system to enter a very congested state, i.e. there are no spillover effects. This is common in design applications, where one assumes that the network performs in an (under)saturated state (see e.g. 59–62, 64, 66, 129).

Let $\mathcal{M} = \{b, c\}$ be the set of modes, indexed by $m \in \mathcal{M}$, such that $b$ denotes the bus and $c$ denotes the car mode. Index cardinal directions of travel by $p \in \mathcal{P} = \{eb, wb, nb, sb\}$, where $eb$, $wb$, $nb$, and $sb$ stand for the eastbound, westbound, northbound, and southbound directions, respectively. Similarly to previous studies, we assume that the network has a grid-like structure. The reason for focusing on a grid-like configuration is because it is the most common network structure used in the literature, not only for the bus network design problem (see e.g. 59, 60, 62, 64–66, 129–131), but also for many traffic-related problems (see e.g. 20, 27, 35, 45, 47, 48, 104, 132–137). To make a grid representation as realistic as possible, we
\(\mathcal{M}\) set of modes indexed by \(m\)
\(\mathcal{A}\) set of possible lane allocations indexed by \(a\)
\(\mathcal{P}\) set of cardinal directions of travel indexed by \(p\)
\(\Lambda\) trip generation rate during loading time of the peak hour
\(\lambda\) trip generation rate during unloading time of the peak hour and off-peak periods
\(w\) average walking speed
\(\varphi\) expected number of transfers
\(\alpha\) average value of time
\(\theta\) parameter of the Logit model
\(\delta\) fixed penalty for transfers expressed in terms of an equivalent walking distance
\(\omega\) time lost per stop due to door operations and deceleration/acceleration time
\(\omega'\) time added per boarding/alighting passenger
\(\Omega\) average dwell time in the network
\(C\) bus capacity
\(L_a\) infrastructure length of lane allocation \(a\)
\(B\) size of the bus fleet
\(V\) total vehicular distance traveled by buses per hour of operation
\(\phi_x/\phi_y\) network length along E-W/N-S directions
\(\psi_x/\psi_y\) street spacing along E-W/N-S directions
\(s_x/s_y\) stop spacing along E-W/N-S directions
\(l_x/l_y\) line spacing along E-W/N-S directions
\(H_x/H_y\) bus headway along E-W/N-S directions
\(N_x/N_y\) number of corridors along E-W/N-S directions
\(N_{x,b}/N_{y,b}\) number of bus lines along E-W/N-S directions
\(z_{x,z}/z_{x,x}\) number of street spacings included in the stop spacing along E-W/N-S
\(z_{x,z}/z_{i,x}\) number of stop spacings included in the line spacing along E-W/N-S
\(\xi_{x,m,a}/\xi_{y,m,a}\) fraction of corridors with lane allocation \(a\) used by mode \(m\) along E-W/N-S directions
\(\ell_x/\ell_y\) trip length realization along E-W/N-S
\(f(\ell_x, \ell_y)\) joint probability density function of the user trip lengths
\(A_m/W_m/u_m\) access (including the egress) time/waiting time/free-flow speed of mode \(m\)
\(\Pr(m)\) probability of choosing mode \(m\)
\(\Pr(m|\ell_x, \ell_y)\) probability of choosing mode \(m\) for a given combination of \(\ell_x\) and \(\ell_y\)
\(IVTT_m|\ell_x, \ell_y\) in-vehicle time traveled for mode \(m\) and combination of \(\ell_x\) and \(\ell_y\)
\(O_p/Q_p\) maximum bus occupancy/maximum number of on-board passengers in direction \(p\)
\(\eta_p\) number of lanes along a corridor in direction \(p\)
\(v_{p,m}\) expected speed of mode \(m\) in direction \(p\)
\(v_{p,m,a}\) operating speed of mode \(m\) for lane allocation \(a\) in direction \(p\)
\(\tau_{p,c,a}\) experienced car delay along a corridor with lane allocation \(a\) in direction \(p\)
\(H_{p,c,a}\) maximum car outflow for lane allocation \(a\) in direction \(p\) given by the 3D-MFD
\(\Lambda_{p,c,a}\) generated car demand on corridor with allocation \(a\) in direction \(p\) during loading time

Table 5.1: Nomenclature.
assume a rectangular shape with sides $\phi_x$ and $\phi_y$, as depicted in Fig. 5.1. Such a rectangular form is inspired by many cities (e.g. Barcelona, Manhattan, Buenos Aires, Oslo, Helsinki, Miami, and Washington D.C) that are elongated in shape \cite{60}. The spacings between the streets along E-W (including $eb$ and $wb$) and N-S (including $nb$ and $sb$) directions are $\psi_x$ and $\psi_y$, respectively. Atop this street structure is the bus network. The bus network consists of two types of bus lines that cross each other but do not overlap: (i) horizontal lines, with stop spacing $s_x$ and line spacing $l_y$; and (ii) vertical lines, with stop spacing $s_y$ and line spacing $l_x$ (see Fig. 5.1).

Note that, although we focus on a grid bus network layout (as in e.g. \cite{65, 66, 130, 131}), we could still apply the proposed framework for combined network structures such as the hybrid one. This could be done by determining the grid bus network design in the city center using the proposed approach, and then distributing the bus lines across the peripheral part in a radial fashion (see \cite{59} and \cite{60} for more details). For radial or ring street layouts, however, we would need to extend our formulations. The formulations presented here can be considered as the first building block towards the design of public transport networks that consider for the first time the interactions with the traffic system, mode choice behavior, and

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure5_1.png}
\caption{City structure and the corresponding bus network.}
\end{figure}
heterogeneous demand patterns and types of user behavior, as elaborated further in this chapter.

Following the approach by (60) and (130), and for simplification purposes, we assume that the stop spacing includes an integer number of street segments, such that:

\[
s_x = z_{s,x} \psi_x, \quad z_{s,x} \in \mathbb{N}, \tag{5.1a}
\]

\[
s_y = z_{s,y} \psi_y, \quad z_{s,y} \in \mathbb{N}. \tag{5.1b}
\]

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

\[
l_x = z_{l,x} s_x = z_{l,x} z_{s,x} \psi_x, \quad z_{l,x}, z_{s,x} \in \mathbb{N}, \tag{5.2a}
\]

\[
l_y = z_{l,y} s_y = z_{l,y} z_{s,y} \psi_y, \quad z_{l,y}, z_{s,y} \in \mathbb{N}. \tag{5.2b}
\]

Let \( \ell_x \) and \( \ell_y \) be the user trip lengths along the E-W and N-S directions, respectively. Denote by \( \{ f(\ell_x, \ell_y) : \ell_x \in [-\phi_x, \phi_x], \ell_y \in [-\phi_y, \phi_y] \} \) the probability density function of the user trip lengths, such that: \( \ell_x \geq 0 \) for \( eb \); \( \ell_x < 0 \) for \( wb \); \( \ell_y \geq 0 \) for \( nb \); and \( \ell_y < 0 \) for \( sb \). This distribution can be estimated by enumerating and grouping all OD pairs with the same \( \ell_x \) and \( \ell_y \), using taxi GPS data (138), using mobile phone traces (139, 140), or sampling other types of vehicle records (141). We consider the estimation of \( f(\ell_x, \ell_y) \) to be out of the scope for this dissertation, thus assume that \( f(\ell_x, \ell_y) \) is exogenously given.

It should be noted here that by using \( f(\ell_x, \ell_y) \) as an intermediate level of abstraction we lose the information about the locations of the trip initialization, which can affect the estimation of the maximum bus occupancy. That being said, the proposed framework is valid for all distributions of the user origins that are slowly-varying in space. This level of abstraction, however, allows to achieve not only a more realistic modeling of urban mobility patterns compared to the uniform distribution of user origins and destinations that comes with a unique trip distance, but also the analytical tractability of the problem. Nevertheless, it is worth noting that the proposed framework could also be applied for spatially-heterogeneous demand distributions. This could be done by clustering the passenger demand into different levels (e.g. \textit{level 1} and \textit{level 2}). Then, using the proposed approach, we would determine the optimal design parameters for \textit{level 1} passenger demand. Afterwards, following the approach by (66), we would
increase the bus frequency in regions with higher demand densities (i.e. with level 2 demand) by adding local bus routes parallel to the main ones (designed for level 1 demand) in those regions (see (66) for more details).

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

In this chapter, we consider two types of user behavior: users adjust the walking distance at the origin and the destination to minimize the number of transfers while respecting a maximum threshold for the walking distance (Type 1); and users adjust the number of transfers to minimize the walking distance at the origin and the destination (Type 2). In each direction of travel, the bus demand is served by buses operating with a constant headway (\( H_x \) along E-W; \( H_y \) along N-S) and identical passenger capacity \( C \), such that \( C_x = C_y = C \) (i.e. all N-S lines have the same headway which is not necessarily the same as that of the E-W lines; and all buses are the same size). Furthermore, there are three types of corridors considered in this chapter: (i) corridors with car-only lanes; (ii) corridors with a dedicated bus lane; and (iii) corridors with a mixed lane. The share of the number of corridors of each type within the total number of corridors in the network directly affects the level of service provided by each individual mode, thus the passenger mode choice.

Let \( A_m \) be the set of types of corridors, indexed by \( a \in A_m \), along which mode \( m \) can operate. Assuming that cars can drive along all types of corridors in the network (i.e. the number of lanes along a corridor in any given direction \( p \) is \( \eta_p \geq 2 \), we set \( A_b = \{db, mb\} \) and \( A_c = \{co, db, mb\} \), where \( co, db, \) and \( mb \) stand for corridors with car-only lanes, corridors with a dedicated bus lane, and corridors with a mixed lane, respectively. Denote by \( N_{x,b} \) the number of E-W corridors with a bus line and by \( N_{y,b} \) the number of N-S corridors with a bus line, such that:

\[
N_{x,b} = \sum_{a \in A_b} N_{x,b,a} = \lfloor \phi_y / l_y \rfloor + 1, \quad (5.3a)
\]

\[
N_{y,b} = \sum_{a \in A_b} N_{y,b,a} = \lfloor \phi_x / l_x \rfloor + 1, \quad (5.3b)
\]
where \( \lfloor \cdot \rfloor \) indicates the floor function; \( N_{x,b,a} \) and \( N_{y,b,a} \) represent the total number of corridors with a bus line operating with lane allocation \( a \in A_b \) along E-W and N-S directions, respectively. Note that the fraction of dedicated/mixed bus lines actually represents the fraction of bus lines with a dedicated/mixed bus lane. Since there is at most a single bus line per corridor, this fraction can also be interpreted as the fraction of bus lines running on dedicated/mixed bus lanes in a given cardinal direction. Then, assuming that the lane allocation remains the same along the whole corridor and for both travel directions (\( eb \) and \( wb \) along E-W; \( nb \) and \( sb \) along N-S), the fraction of bus lines with each lane allocation type along E-W and N-S directions can simply be obtained as:

\[
\zeta_{x,b,a} = \frac{N_{x,b,a}}{N_{x,b}}, \quad (5.4a)
\]
\[
\zeta_{x,b,a} = \frac{N_{y,b,a}}{N_{y,b}}. \quad (5.4b)
\]

Similarly, the fraction of corridors available for car traffic operating with lane allocation \( a \in A_c \) along E-W (\( \zeta_{x,c,a} \)) and N-S (\( \zeta_{y,c,a} \)), can be computed using Eq. 5.5, where \( N_x \) is the number of E-W corridors and \( N_y \) is the number of N-S corridors in the network (Eq. 5.6).

\[
\zeta_{x,c,a} = \begin{cases} 
1 - \frac{N_{x,b}}{N_x}, & \text{if } a = co, \\
\zeta_{x,b,a} \frac{N_{x,b}}{N_x}, & \text{if } a \in \{db, mb\},
\end{cases} \quad (5.5a)
\]
\[
\zeta_{y,c,a} = \begin{cases} 
1 - \frac{N_{y,b}}{N_y}, & \text{if } a = co, \\
\zeta_{y,b,a} \frac{N_{y,b}}{N_y}, & \text{if } a \in \{db, mb\},
\end{cases} \quad (5.5b)
\]

with

\[
N_x = \phi_y / \psi_y + 1, \quad (5.6a)
\]
\[
N_y = \phi_x / \psi_x + 1. \quad (5.6b)
\]

Notice from Eqs. 5.4–5.5 that we do not specify which corridors have a dedicated bus lane. Instead, we stick to the fraction of dedicated bus lines along E-W and N-S directions as the global parameters for our design problem.
5.1.2 Mathematical formulation of the objective function

Here we formulate the objective function used to determine the optimal design parameters. The decision variables include stop spacings ($z_{s,x}$ and $z_{s,y}$), line spacings ($z_{l,x}$ and $z_{l,y}$), headways ($H_x$ and $H_y$), and fractions of dedicated bus lines ($\xi_{x,b,db}$ and $\xi_{y,b,db}$) along E-W and N-S directions. They are denoted as $D = [z_{s,x}, z_{s,y}, z_{l,x}, z_{l,y}, H_x, H_y, \xi_{x,b,db}, \xi_{y,b,db}]$. The objective function is given by Eq. 6.2 and, similarly to the previous studies, it consists of two components: agency cost $Z_O(D)$; and user cost $Z_U(D)$.

$$\min_D Z(D) = Z_O(D) + Z_U(D), \quad (5.7)$$

s.t. $H_x, H_y \geq H_{\text{min}}, \quad (5.8)$

$$z_{s,x}, z_{s,y}, z_{l,x}, z_{l,y} \in \mathbb{N}, \quad (5.9)$$

$$0 \leq \xi_{x,b,db} \leq 1, \quad 0 \leq \xi_{y,b,db} \leq 1, \quad (5.10)$$

$$O_p \leq C, \quad \forall p \in \mathcal{P}. \quad (5.11)$$

where $O_p$ is the maximum bus occupancy in direction $p$.

Note that all decision variables need to be strictly positive for physical reasons. In addition, the headway in each cardinal direction should not be below the predefined threshold $H_{\text{min}}$ (Eq. 5.8). The stop spacings should include an integer number of street segments, whereas the line spacings should consist of an integer number of stop spacings (Eq. 5.9). Moreover, the number of dedicated bus lines cannot exceed the number of bus corridors (Eq. 5.10). Finally, the last constraint (Eq. 5.11) indicates that the maximum bus occupancy in each cardinal direction $p$ needs to be lower than the bus passenger capacity $C$.

It is also worth noting that there is no need to incorporate constraints related to the connectivity or continuity of bus lanes, given that the bus lines operate along the entire corridor’s length in all cardinal directions. In other words, both the connectivity and continuity of bus lanes are ensured by the general problem setup. They are connected to the level of service by explicitly considering and incorporating all parameters of the bus system in the operator and user cost function, as elaborated further below. Before we define the cost components, let us first introduce the assumptions used for the methodological framework, in terms of the lane allocation along a corridor, the waiting time of bus passengers, the vehicle dynamics, and the delay time experienced during the peak hour.
Assumption 5.1 (Lane allocation along a corridor). We assume that the lane allocation remains the same along the whole corridor and for both travel directions.

Remark. Assumption 5.1 is made to prevent possible disruptions to the bus system induced by varying lane allocation, guaranteeing thereby the same operating conditions in both travel directions (120). Recall that this is a design problem, thus it would not be reasonable from a design point of view to disrupt operations by changing the lane allocation along a given arterial.

Assumption 5.2 (Waiting time of bus passengers). We assume that bus users arrive at the stop independently of the schedule, thus they wait on average half of the bus headway.

Remark. Assumption 5.2 is reasonable for relatively short headways that are considered in this chapter (as in 59, 60, 66, 129).

Assumption 5.3 (Vehicle dynamics). We assume that, along corridors with mixed lane, buses drive at the speed of cars while cruising, i.e. both modes have the same speed either in the free-flow or congested traffic conditions.

Remark. Assumption 5.3 has been used in previous studies (see e.g. 35, 59, 60, 84, 129) and was empirically confirmed (10). Nevertheless, it is only made for simplification purposes, given that the method used to compute the operating speed for each mode (9) allows to treat buses as moving bottlenecks while cruising. More details are given in Section 5.1.4.

Assumption 5.4 (Delays experienced during the peak hour). We assume that the delays experienced during the peak hour in case the demand exceeds the maximum outflow can be approximated using a single bottleneck model.

Remark. Assumption 5.4 also used in (65) is made because we look at the bus network design problem under the steady state, when the traffic conditions are undersaturated or saturated, and there are no disruptions causing the network capacity to be dynamically reduced due to some internal or external factors. In other words, since the oversaturated conditions should not be solved by the design, but a proper control mechanism, we do not consider complex congestion spreading during very congested states. More details are given in Section 5.1.4.

The agency cost metrics include the infrastructure length of each lane allocation type \( \{L_a(D) : a \in A_b\} \), the size of the bus fleet \( B(D) \), and the total
vehicular distance traveled by buses per hour of operation $V(D)$ (Eq. 6.9). The associated parameters $\{\pi_{L,a} : a \in A_b\}$, $\pi_B$, and $\pi_V$ are the corresponding unit costs. To put all the cost components in the objective function (Eq. 6.2) under the same unit (i.e. hours per passenger), we convert agency money into an equivalent user riding time with parameter $\alpha$ denoting an average value of time, and divide it by the total number of bus passengers, $Pr(b)\lambda$. Note that we use lower trip generation rate value $\lambda$ in Eq. 6.9 as is gives us the maximum bound for the operator cost, which corresponds to the worst-case scenario (see 59, for more details).

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

with

$$Pr(m) = \int_{-\phi_y}^{\phi_y} \int_{-\phi_y}^{\phi_y} Pr(m | \ell_x, \ell_y) f(\ell_x, \ell_y) \, d\ell_y \, d\ell_x, \quad (5.13)$$

where $Pr(m | \ell_x, \ell_y)$ stands for the probability of choosing mode $m$ for a given combination of $\ell_x$ and $\ell_y$. This probability can be computed by applying the Logit model (142), as given by Eq. 5.14, where $\theta$ is the parameter of the Logit model, and $TTT(D)_{m | \ell_x, \ell_y}$ denotes the total time traveled for a given mode $m$ and combination of $\ell_x$ and $\ell_y$.

$$Pr(m | \ell_x, \ell_y) = \frac{\exp(-\theta \cdot TTT(D)_{m | \ell_x, \ell_y})}{\sum_{m' \in M} \exp(-\theta \cdot TTT(D)_{m' | \ell_x, \ell_y})}. \quad (5.14)$$

On the other hand, the user cost metric of interest is the expected total time traveled, computed across both modes and the entire set of the user trip lengths (Eq. 6.10).

$$Z_U(D) = \int_{-\phi_x}^{\phi_x} \int_{-\phi_y}^{\phi_y} TTT(D)_{m | \ell_x, \ell_y} Pr(m | \ell_x, \ell_y) f(\ell_x, \ell_y) \, d\ell_y \, d\ell_x. \quad (5.15)$$

In the following, we derive the components of the operator cost function. We start by formulating the infrastructure length (Eq. 5.16). This variable can simply be obtained by multiplying the length associated with one bus line by the total number of bus lines with a given lane allocation along E-W and N-S directions.

$$L_a(D) = \begin{cases} \zeta_{x,b,db} N_x, b \phi_x + \zeta_{y,b,db} N_y, b \phi_y, & \text{if } a = db, \\ (1 - \zeta_{x,b,db}) N_x, b \phi_x + (1 - \zeta_{y,b,db}) N_y, b \phi_y, & \text{if } a = mb. \end{cases} \quad (5.16)$$

The bus fleet consists of buses operating in all cardinal directions (Eq. 5.17). In each direction of travel, the number of required buses can be
approximated by multiplying the number of buses required for one bus line by the total number of bus lines.

\[
B(D) = \sum_{p \in \{eb, wb\}} N_{x,b} \phi_x / (H_x \nu_{p,b}) + \sum_{p \in \{nb, sb\}} N_{y,b} \phi_y / (H_y \nu_{p,b}),
\]

(5.17)

where \( \nu_{p,b} \) stands for the expected bus speed (including the dwell and acceleration/deceleration time) in direction \( p \). This expected bus speed can be determined using Eq. 5.18, where \( \nu_{p,m,a} \) denotes the operating speed of mode \( m \) for lane allocation \( a \) in direction \( p \).

\[
\nu_{p,m} = \begin{cases} \sum_{a \in A_m} \nu_{p,m,a} \zeta_{x,m,a} & \text{if } p \in \{eb, wb\}, \\ \sum_{a \in A_m} \nu_{p,m,a} \zeta_{y,m,a} & \text{if } p \in \{nb, sb\}. \end{cases}
\]

(5.18)

Note that the operating speeds \( \nu_{p,m,a} \) are not predefined, but depend on the network loading and the topology in a given cardinal direction. This will be further elaborated in Section 5.1.4. Following the approach by (59), we formulate the bus operating speed as follows:

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

(5.19)

where \( u_b \) is the bus free-flow speed; \( \nu_{p,c,mb} \) denotes the operating car speed along a mixed lane in direction \( p \); \( \omega \) represents the time lost per stop due to required door operations and deceleration/acceleration maneuvers; \( \Omega \) stands for the average dwell time in the network. To estimate \( \Omega \), we first need to define the last agency cost metric, the total vehicular distance traveled by buses per hour of operation \( V(D) \). Following the approach by (59), we define this distance as the ratio of the length of the network covered by bus lines and the corresponding bus headway across all cardinal directions. It can also be defined as the product of the total number of buses operating in all cardinal directions and the corresponding expected bus speeds (Eq. 5.20). Note that, since the headway for E-W/N-S bus lines is the same in both travel directions (east to west and west to east/north to south and south to north), we multiply the total vehicular distance traveled along E-W/N-S direction by 2.

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

(5.20)
Then, the average dwell time in the network can be obtained by multiplying the time added per boarding/alighting passenger $\omega'$ by the average number of passengers a vehicle collects per unit distance traveled:

$$\Omega = \omega' \Pr(b) \Lambda(1 + \varphi(D)) / V(D). \quad (5.21)$$

where $\varphi(D)$ is the expected number of transfers. According to (59), the average number of passengers a vehicle collects per unit distance traveled can be approximated as the ratio of the number of boarding passengers generated per hour during the peak period ($\Pr(b) \Lambda(1 + \varphi)$) and the total vehicular distance traveled by buses per hour of operation. To get the maximum bound for the user cost function, we use higher trip generation rate value $\Lambda$ in Eq. (5.21), similarly to (59) and (60).

Finally, the last term in the objective function is the total time traveled for a given mode $m$ and combination of the user trip lengths. It consists of the expected access (including the egress) time $A_m(D)$, the expected waiting time $W_m(D)$, the in-vehicle time traveled $IVTT(D)_{m | \ell_x, \ell_y}$ for a given $\ell_x$ and $\ell_y$, and some additional time components (Eq. 5.22). For the bus mode, these additional components include the transferring time, whereas for the car mode, the additional time traveled is related to the equivalent mileage costs of an auto trip (as in 59).

$$TTT(D)_{m | \ell_x, \ell_y} = A_m(D) + W_{m}(D) + IVTT(D)_{m | \ell_x, \ell_y}$$

$$+ \mathbbm{1}_{\{m=b\}} \varphi(D) \delta / w + \mathbbm{1}_{\{m=c\}} (|\ell_x| + |\ell_y|) \pi_D / \alpha, \quad (5.22)$$

with

$$IVTT(D)_{m | \ell_x, \ell_y} = |\ell_x| / v_{\xi(\ell_x),m} + |\ell_y| / v_{\xi(\ell_y),m}, \quad (5.23)$$

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

The following equations approximate the network-level components of the user cost function for the bus mode (Eq. 5.22), including the access,

---

1 Alternatively, one can also incorporate the fare for using the bus mode and/or congestion tolls, parking, and garaging costs associated with the car mode.
the waiting, and the transferring time, for each type of user behavior. For
Type 1, Eqs. 5.24–5.26 represent an extension of those in (60) for more
general bus network configurations. We say an extension, because the for-
mulations proposed by (60) are only applicable for networks with the same
stop spacing and bus headway along E-W and N-S directions (i.e. \( s_x = s_y \)
and \( H_x = H_y \)), which may not be the case in this dissertation (i.e. design
parameters may vary across cardinal directions).

**Type 1** (Minimizing the number of transfer).

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

\[
\varphi_1(D) = \Pr(\varphi > 0)
\]

\[
= 1 - \Pr(\varphi = 0)
\]

\[
= 1 - (l_y \varphi_x + l_x \varphi_y - l_x l_y) / (\varphi_x \varphi_y),
\]

\[
W_{1,b}(D) = (1 + \varphi_1(D)) \bar{H} / 2,
\] (5.24)

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

**Type 2** (Minimizing the walking distance). Depending on the ratio be-
tween \( l_x \) and \( l_y \), which defines the rectangular shape of the network structure
presented in Fig. 5.1, we distinguish the following two cases:

**Case 1** \( (l_x \geq l_y) \).

\[
A_{2,b}(D) = (6l_x l_y + 6l_x s_x - 2l_x^2 - 3l_y s_x + 3l_y s_y) / (12l_x w).
\]

\[
\varphi_2(D) = \Pr(\varphi > 0)(4l_x^2 l_y^2 - 2l_x l_y^3 + l_x^4) / (2l_x^2 l_y^2).
\]

\[
W_{2,b}(D) = \Pr(\varphi = 0) \bar{H} / 2 + \Pr(\varphi > 0)(H_x (8l_x^2 l_y^2 - 4l_x l_y^3 + l_y^4)
\]

\[
+ H_y (4l_x^2 l_y^2 + l_x^4)) / (8l_x^2 l_y^2).
\] (5.27)

**Case 2** \( (l_x < l_y) \).

\[
A_{2,b}(D) = (6l_x l_y + 6l_y s_y - 2l_x^2 - 3l_x s_y + 3l_x s_x) / (12l_y w).
\]

\[
\varphi_2(D) = \Pr(\varphi > 0)(4l_x^2 l_y^2 - 2l_x l_y^3 + l_x^4) / (2l_x^2 l_y^2).
\]

\[
W_{2,b}(D) = \Pr(\varphi = 0) \bar{H} / 2 + \Pr(\varphi > 0)(H_x (4l_x^2 l_y^2 + l_x^4)
\]

\[
+ H_y (8l_x^2 l_y^2 - 4l_x l_y^3 + l_x^4)) / (8l_x^2 l_y^2).
\] (5.29)
5.1.3 Derivation of the maximum bus occupancy

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

To compute $O_p$, we have to evaluate the maximum number of on-board passengers $Q_p$ across all points in the corresponding cardinal direction. For this purpose, we use vertical and horizontal cordons denoted as $\beta \phi_x$ and $\beta \phi_y$, respectively, such that $\beta \in [0,1]$ (see Fig. 5.2). Intuitively, vertical cordons are used for $Q_{eb}$ and $Q_{wb}$, whereas the horizontal ones are used for $Q_{nb}$ and $Q_{sb}$. Recall that due to the use of the trip length distribution as an intermediate level of abstraction, we do not know the locations of the trip initialization. Hence, for each cordon, we have to determine the probability that the bus users cross the cordon for a given $\ell_x$ (in case of a vertical cordon) or $\ell_y$ (in case of a horizontal cordon), in which case they are considered as part of the total number of on-board passengers in the corresponding cardinal direction. These probabilities are denoted as $\Pr(J_{\beta}(\ell_x))$ and $\Pr(J_{\beta}(\ell_y))$, where $J_{\beta}(\ell_x)$ is the event that the bus users travel across vertical cordon $\beta \phi_x$ for a given $\ell_x$, and $J_{\beta}(\ell_y)$ is the event that the bus users travel across horizontal cordon $\beta \phi_y$ for a given $\ell_y$. Given the assumption regarding the slowly-varying spatial distribution of the user origins, the aforementioned probabilities can be computed as:

$$
\Pr(J_{\beta}(\ell_x)) = \begin{cases} 
|\ell_{x,b}| / (\phi_x - |\ell_{x,b}|), & \text{if } |\ell_{x,b}| \leq \phi_x^{\text{MIN}}, \\
\phi_x^{\text{MIN}} / (\phi_x - |\ell_{x,b}|), & \text{if } \phi_x^{\text{MIN}} < |\ell_{x,b}| \leq \phi_x^{\text{MAX}}, \\
1, & \text{if } |\ell_{x,b}| > \phi_x^{\text{MAX}},
\end{cases}
$$

$$
\Pr(J_{\beta}(\ell_y)) = \begin{cases} 
|\ell_{y,b}| / (\phi_y - |\ell_{y,b}|), & \text{if } |\ell_{y,b}| \leq \phi_y^{\text{MIN}}, \\
\phi_y^{\text{MIN}} / (\phi_y - |\ell_{y,b}|), & \text{if } \phi_y^{\text{MIN}} < |\ell_{y,b}| \leq \phi_y^{\text{MAX}}, \\
1, & \text{if } |\ell_{y,b}| > \phi_y^{\text{MAX}}.
\end{cases}
$$

with

$$
\phi_x^{\text{MIN}} = \min\{\beta \phi_x, (1 - \beta) \phi_x\},
$$

$$
\phi_x^{\text{MAX}} = \max\{\beta \phi_x, (1 - \beta) \phi_x\},
$$

$$
\phi_y^{\text{MIN}} = \min\{\beta \phi_y, (1 - \beta) \phi_y\},
$$

$$
\phi_y^{\text{MAX}} = \max\{\beta \phi_y, (1 - \beta) \phi_y\}.
$$
\[ \phi_y^{\text{MAX}} = \max \{ \beta \phi_y, (1 - \beta) \phi_y \}, \]  

where \( \phi_x - |\ell_x| \) and \( \phi_y - |\ell_y| \) represent the domains of valid user origins for the corresponding \( \ell_x \) and \( \ell_y \) (see Fig. 5.2).

**Figure 5.2:** A city and its vertical and horizontal cordons.

Then, the maximum bus occupancy can simply be obtained by dividing the maximum number of on-board passengers across all bus lines and all cordons in a given cardinal direction by the total number of bus lines and the respective bus frequency (Eq. 6.27). In other words, by moving the cordons in a given cardinal direction, we enumerate all possible occupancy rates, based on which we find the maximum one for the corresponding direction of travel. Note that we use the higher trip generation rate in Eq. 5.36, i.e. the value of \( \Lambda \) observed during the loading time of the peak period, to get the maximum number of on-board passengers.

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

(5.35)

with

\[
Q_{eb} = \max_{\beta \in [0,1]} \left\{ \int_0^{\phi_x} \int_{-\phi_y}^{\phi_y} \Pr( J_{\beta}(\ell_x) ) \Lambda \Pr(b | \ell_x, \ell_y) f(\ell_x, \ell_y) \, d\ell_y \, d\ell_x \right\},
\]  

(5.36a)

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

(5.36b)
\[ Q_{nb} = \max_{\beta \in [0,1]} \left\{ \int_0^{\Phi_y} \int_{-\Phi_x}^{\Phi_x} \Pr(J_\beta(\ell_y)) \Lambda \Pr(b \mid \ell_x, \ell_y) f(\ell_x, \ell_y) \, d\ell_x \, d\ell_y \right\}, \]

(5.36c)

\[ Q_{sb} = \max_{\beta \in [0,1]} \left\{ \int_{-\Phi_y}^{0} \int_{-\Phi_x}^{\Phi_x} \Pr(J_\beta(\ell_y)) \Lambda \Pr(b \mid \ell_x, \ell_y) f(\ell_x, \ell_y) \, d\ell_x \, d\ell_y \right\}. \]

(5.36d)

Notice from Eq. 6.27 the importance of providing different bus headways along E-W and N-S directions. In other words, a unique bus headway for the entire network might lead to an efficient bus service (measured by the ability to serve the total generated demand with the available bus capacity) in some directions. However, other directions might experience a lower level of service, i.e. significantly higher bus occupancy, due to its less favorable design parameters (e.g. less number of bus lines, lower fraction of dedicated bus lines, or longer stop spacing), constrained by the geometrical characteristics of the network. In other words, a unique bus headway, as it has been previously proposed in most of the literature (see e.g. 59, 60, 65, 66), might not lead to the optimal design parameters in all cardinal directions of travel. Therefore, by considering different bus headways, we balance the bus level of service provided in each direction, preventing such scenario.

To verify Eqs. 5.33–5.36, we show that the solution for the maximum bus occupancy in case of uniformly and independently distributed origins and destinations is the same to that of (59) (see the Appendix for more details).

5.1.4 Derivation of the operating car speed for each lane allocation type

As mentioned before, in the problem investigated in this chapter, there are three types of corridor in each cardinal direction of travel \( p \) for which the operating speed \( v_{p,m,a} \) needs to be estimated: (i) corridors with car-only lanes; (ii) corridors with a dedicated bus lane; and (iii) corridors with a mixed lane. For corridors with a dedicated bus lane, the operating bus speed \( v_{p,b,db} \) is not affected by the traffic conditions, i.e. interactions with the car mode (assuming that a dedicated bus lane operates with a transit signal priority, as in 143). Hence, it can simply be modeled as the free-flow bus speed \( u_b \) (including the dwell and acceleration/deceleration time) (Eq. 5.19). Likewise, given Assumption 5.3 that buses drive at the speed of cars while cruising, the operating bus speed \( v_{p,b,mb} \) for corridors with a mixed lane can be modeled as the operating car speed \( v_{p,c,mb} \) along a corridor.
with a mixed lane in the corresponding direction of travel (including the dwell time and acceleration/deceleration time) (Eq. 5.19). As a result, for each cardinal direction and type of corridors, we only need to estimate the operating car speed. This operating car speed depends on the accumulation of both buses and cars.

Recall from the previous chapters that one of the commonly used approaches to determine the operating speed for each mode at the network level, while capturing the modal interactions and the network topology, is the 3D-MFD (6–9, 137). Note that by the modal interactions we refer to how cars and buses influence each other in terms of traffic dynamics. In particular, such interactions include the impact of bus operations (e.g. accumulation, headway, dwell times at bus stops along mixed lanes) on the speed of cars, as well as the impact of car accumulation on the operation of buses (e.g. reduced speed along mixed lanes).

In the classical framework, the 3D-MFD is assumed to be a function of the total production in the network with respect to the accumulation of both buses and cars. Such 3D-MFD is resourceful for practitioners to understand the impact of each mode in the network. Nevertheless, for modeling purposes, it is more appealing to segregate the 3D-MFD into two partial 3D-MFDs, one for each mode of transport (137). This segregation yields a more accurate estimation of the operating speed for each mode. Therefore, in this chapter, we apply the partial car 3D-MFD to determine the operating car speed and ultimately the optimal design parameters. It is worth noting that we could also apply the partial bus 3D-MFD to determine the operating bus speed along corridors with a mixed lane. However, this would not yield a very accurate bus speed estimation for a given level of bus passenger demand, given that the partial bus 3D-MFD does not explicitly account for the dwell time that depends on the bus passenger demand level. For that reason, we compute the bus speed along corridors with a mixed lane using the operating car speed (determined from the partial car 3D-MFD) and the analytical expression for the dwell time as a function of the bus passenger demand (Eq. 5.21).

The partial car 3D-MFD is estimated by generating multi-modal car MFDs for a range of bus headways using the framework developed in Chapter 4. Each multi-modal car MFD represents a slice of the partial car 3D-MFD, covering a range of traffic conditions for a particular bus headway. Hence, to change the operating regime of buses with the decision variables (i.e. bus headways \( H_x \) and \( H_y \)), we use different slices of the obtained partial car 3D-MFD, as long as the network topology (defined by the
decision variables $\zeta_{x,b,db}$ and $\zeta_{y,b,db}$) remains the same. Once the network topology changes, we derive the corresponding partial car 3D-MFD by repeating the process mentioned above. This way, we are able to determine the effects that the design variables have on the interactions between the modes. We do so by quantifying the operating car speed (and hence the operating bus speed) for any given bus network configuration scenario (as defined by the decision variables).

Note, however, that most analytical studies for approximating the (multi-modal) MFD using variational theory treat the entire network as a single, one dimensional corridor (13). This may not lead to a very accurate multi-modal MFD approximation for networks whose topological (intersection and stop spacing) and operational characteristics (signal settings, lane allocation, and bus headway) vary significantly across cardinal directions, as it may be the case in the considered problem. Therefore, to more accurately quantify the performance of each mode, we estimate the partial car 3D-MFD for each cardinal direction and corridor type, i.e. lane allocation layout (corridors with car only lanes, a dedicated bus lane, or a mixed lane), and then combine the results similar to (48). It is important to note that combining directional partial car 3D-MFDs is reasonable for the considered bus network design problem, given that we examine the best case scenario when there are no disruptions causing the network capacity to be dynamically reduced due to some internal or external factors. In other words, the traffic control operates well-enough to remove local capacity drops. Consequently, the outflow in one direction can smoothly go in the other. This is a regular process when discussing the design problem. Once the partial car 3D-MFD is estimated for each travel direction and type of corridor, we feed the derived partial car 3D-MFD with the corresponding travel demand, as elaborated further below. This way, we implicitly take into account the interactions between cardinal directions. Interactions across the two cardinal directions include signal timing allocation (which we do take into account), demand distribution (which we also take into account), and spillovers (which are not relevant for a design problem, as discussed in Section 5.1.1).

To compute $v_{p,c,a}$, we first need to determine whether the total generated car demand $\Lambda_{p,c,a}$ (Eq. 5.37) along a corridor with allocation $a$ in direction $p$ during the loading time $T$ of the peak period (Fig. 5.3) is smaller than the corresponding corridor capacity $\mu_{p,c,a} = P_{p,c,a}^{*}/\bar{l}_p$. We define the corridor capacity as the maximum outflow, i.e. ratio of the maximum travel production $P_{p,c,a}^{*}$ and the average trip length for a given cardinal direction.
\( \bar{\ell}_p \) (Eq. 5.38). Notice from Eq. 5.37 that we assume that the car capacity for a corridor with a dedicated bus line is reduced by one lane compared to other lane allocation types.

\[
\Lambda_{eb,c,a} = \int_0^{\phi_x} \int_{-\phi_y}^{\phi_y} \left( \Delta / N_X(\eta_p - 1_{\{a=db\}}) \right) \Pr(c \mid \ell_x, \ell_y) f(\ell_x, \ell_y) \, d\ell_y \, d\ell_x,
\]

(5.37a)

\[
\Lambda_{wb,c,a} = \int_0^{\phi_x} \int_{-\phi_y}^{\phi_y} \left( \Delta / N_X(\eta_p - 1_{\{a=db\}}) \right) \Pr(c \mid \ell_x, \ell_y) f(\ell_x, \ell_y) \, d\ell_y \, d\ell_x,
\]

(5.37b)

\[
\Lambda_{nb,c,a} = \int_0^{\phi_y} \int_{-\phi_x}^{\phi_x} \left( \Delta / N_Y(\eta_p - 1_{\{a=db\}}) \right) \Pr(c \mid \ell_x, \ell_y) f(\ell_x, \ell_y) \, d\ell_x \, d\ell_y,
\]

(5.37c)

\[
\Lambda_{sb,c,a} = \int_{-\phi_y}^{\phi_y} \int_{-\phi_x}^{\phi_x} \left( \Delta / N_Y(\eta_p - 1_{\{a=db\}}) \right) \Pr(c \mid \ell_x, \ell_y) f(\ell_x, \ell_y) \, d\ell_x \, d\ell_y.
\]

(5.37d)

\[
\bar{\ell}_{eb} = \frac{\int_0^{\phi_x} \int_{-\phi_y}^{\phi_y} \ell_x | f(\ell_x, \ell_y) \, d\ell_y \, d\ell_x}{\int_0^{\phi_x} \int_{-\phi_y}^{\phi_y} f(\ell_x, \ell_y) \, d\ell_y \, d\ell_x},
\]

(5.38a)

\[
\bar{\ell}_{wb} = \frac{\int_{-\phi_x}^{\phi_x} \int_{-\phi_y}^{\phi_y} \ell_x | f(\ell_x, \ell_y) \, d\ell_y \, d\ell_x}{\int_{-\phi_x}^{\phi_x} \int_{-\phi_y}^{\phi_y} f(\ell_x, \ell_y) \, d\ell_y \, d\ell_x},
\]

(5.38b)

\[
\bar{\ell}_{nb} = \frac{\int_0^{\phi_y} \int_{-\phi_x}^{\phi_x} \ell_y | f(\ell_x, \ell_y) \, d\ell_x \, d\ell_y}{\int_0^{\phi_y} \int_{-\phi_x}^{\phi_x} f(\ell_x, \ell_y) \, d\ell_x \, d\ell_y},
\]

(5.38c)

\[
\bar{\ell}_{sb} = \frac{\int_{-\phi_y}^{\phi_y} \int_{-\phi_x}^{\phi_x} \ell_y | f(\ell_x, \ell_y) \, d\ell_x \, d\ell_y}{\int_{-\phi_y}^{\phi_y} \int_{-\phi_x}^{\phi_x} f(\ell_x, \ell_y) \, d\ell_x \, d\ell_y}.
\]

(5.38d)

Depending on the ratio between \( \Lambda_{p,c,a} \) and \( \mu_{p,c,a} \), we can distinguish the following two cases: \( \Lambda_{p,c,a} \leq \mu_{p,c,a} \); and \( \Lambda_{p,c,a} > \mu_{p,c,a} \). The first case implies uncongested traffic conditions, in which the corridor capacity is sufficient to serve the total generated demand, hence no delays due to congestion are experienced in the system. Consequently, there is a direct match between the total generated demand and the outflow along a corridor. In the second case, however, such a direct match between the demand and the outflow can no longer be established, as the outflow might be limited to
the corridor capacity during the time when congestion is active. Therefore, to tackle this problem and given Assumption 5.4, we use a simple bottleneck model that considers the entire corridor as a single bottleneck with a mean capacity (see (65) for more details). Note that this model is applied to account for possible delays during the peak hour in case the demand exceeds the maximum outflow. It is used because we look at the bus network design problem under the steady state, when the traffic conditions are (under)saturated and there are no disruptions causing the network capacity to be dynamically reduced due to some internal or external factors, as we stated before. Oversaturated conditions should not be solved by the design, but a proper control mechanism (e.g. 120). Hence, we do not consider complex congestion spreading during very congested states. In the following, we summarize how the operating car speed is computed for each case separately.

\[ P_{p,c,a} = \Lambda_{p,c,a}\ell_p. \]  

**Case 1** \( (\Lambda_{p,c,a} \leq \mu_{p,c,a}) \). The network operates in the uncongested regime. The operating car speed can be determined by finding the point on the multi-modal car MFD in which the total travel production is equivalent to that obtained by applying the queuing formula of (144):

**Case 2** \( (\Lambda_{p,c,a} > \mu_{p,c,a}) \). The network operates at capacity. Due to congestion, cars experience certain delay \( \tau_{p,c,a} \), which needs to be taken into account when computing \( v_{p,c,a} \). Following the approach by (65), we can
compute this delay as half of the maximum car delay $\tau_{p,c,a}^{\text{MAX}}$ experienced during the peak hour, i.e. $\tau_{p,c,a} = \tau_{p,c,a}^{\text{MAX}} / 2 = (T(\Lambda_{p,c,a} - \mu_{p,c,a})/\mu_{p,c,a})/2$ (see Fig. 5.3). The operating car speed is then given by Eq. 5.40, where $v_{p,c,a}^*$ denotes the operating car speed corresponding to the maximum outflow along a corridor with allocation $a$ in direction $p$.

$$v_{p,c,a} = (1/v_{p,c,a}^* + \tau_{p,c,a})^{-1}. \quad (5.40)$$

5.2 NUMERICAL RESULTS

5.2.1 Solution algorithm

For the purpose of computing the optimal design parameters, we follow the approach of previous studies and use a brute-force search technique (see e.g. [59–61, 129]). Alternatively, other optimization techniques for solving the problem could also be explored, which may give the solution faster. However, this is out of the scope for this dissertation. To speed up the searching process over the feasible region, we perform a Monte Carlo sampling for a given trip length pattern. Notice that more elaborated sampling techniques could potentially be used to further increase the computational speed ([145]). The result is the set of trip length values along E-W ($L_x$) and N-S ($L_y$) directions. Then, we find the network-level parameters of the bus user cost function according to the selected type of user behavior (i.e. minimizing the number of transfers or the walking distance). To compute the total travel cost for each mode, we need to find the operating speeds. These operating speeds depend on the mode choice that is, on the other hand, a function of the utility cost (i.e. travel time) for each mode. The utility costs are determined based on the bi-modal MFD derived for a given bus network configuration scenario (defined by the decision variables). In other words, the total travel costs are computed by solving a fixed point problem, as shown by the flow chart in Fig. 5.4. This flow chart summarizes the methodological steps.

Note that, rather than using a conventional convergence algorithm to solve the fixed point problem, we discretize (by 1%) potential mode choice values, and for each value we compute its absolute difference to the mode choice value given by the Logit model (Eqs. 5.13–5.14). Then, using a brute-force approach, we find the value of the mode choice that has the minimum absolute difference compared to one obtained from the Logit model. This minimum absolute difference found is always less than 1%, given that we discretize our mode choice values by 1%. Once the fixed point problem
is solved, we compute the operator and user costs, hence the total system cost.

\[ \text{Trip length pattern} \xrightarrow{f(\ell_x, \ell_y)} \text{Monte Carlo sampling} \xrightarrow{\mathcal{L}_x, \mathcal{L}_y} \text{Transfer vs. walking} \]

\[ \text{Mode choice distribution} \xrightarrow{v_{p,m,a}} \text{Total travel costs} \xrightarrow{TTT_{m|\ell_x,\ell_y}} \text{Traffic state} \]

\[ \text{Total system cost} \]

**Figure 5.4**: Schematic representation of the solution algorithm.

### 5.2.2 Case study and analyzed scenarios

Here we investigate the effects of demand intensity, user behavior, and trip length patterns on the optimal bus network configuration and passenger mode choice (i.e. traffic performance). We do this by applying the previously described framework to the Barcelona network. Table 5.2 summarizes the input parameters for this case study. The values are similar to those in the literature (see e.g. 59–61, 129).

### 5.2.3 General results and the effects of demand intensity

To quantify the effects of demand intensity, we analyze two scenarios: SC-I, indicating the base case, with the trip generation rates given in Table 5.2; and SC-II, representing an increase in the trip generation rate (both during the peak and off-peak period) by 33%. Similarly to (66), we explore four distinct trip length patterns, reflecting different spatial demand distributions: (i) a uniform city (Fig. 5.5a), where origins and destinations are uniformly distributed across the entire network; (ii) a mono-centric city (Fig. 5.5c), where most of the origins and destinations are located within the city center, resulting in shorter trip lengths compared to (i) (see Fig. 5.5d); (iii) a commuter city (Fig. 5.5e), where most of the origins and destinations are located in two regions positioned in the opposite corners of the network, resulting in longer trip lengths compared to (i) (see Fig. 5.5f);
and (iv) a twin city (Fig. 5.5g), where most of the origins and destinations are located in two nearly-adjacent regions near the city center.

<table>
<thead>
<tr>
<th>Input parameter</th>
<th>Variable</th>
<th>Units</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network length</td>
<td>$\phi_x/\phi_y$</td>
<td>km</td>
<td>10/4.95</td>
</tr>
<tr>
<td>Street spacings</td>
<td>$\psi_x/\psi_y$</td>
<td>km</td>
<td>0.25/0.15</td>
</tr>
<tr>
<td>Equivalent penalty distance per transfer</td>
<td>$\delta$</td>
<td>km</td>
<td>0.03</td>
</tr>
<tr>
<td>Trip generation rate during the peak hour</td>
<td>$\Lambda$</td>
<td>pax/hr</td>
<td>75000</td>
</tr>
<tr>
<td>Trip generation rate during the off-peak hour</td>
<td>$\lambda$</td>
<td>pax/hr</td>
<td>30000</td>
</tr>
<tr>
<td>Loading time of the peak period</td>
<td>$T$</td>
<td>hr</td>
<td>1</td>
</tr>
<tr>
<td>Bus free-flow speed</td>
<td>$u_b$</td>
<td>km/hr</td>
<td>40</td>
</tr>
<tr>
<td>Average walking speed</td>
<td>$w$</td>
<td>km/hr</td>
<td>2</td>
</tr>
<tr>
<td>Time lost per stop</td>
<td>$\omega$</td>
<td>sec</td>
<td>30</td>
</tr>
<tr>
<td>Boarding and alighting time per passenger</td>
<td>$\omega'$</td>
<td>sec/pax</td>
<td>1</td>
</tr>
<tr>
<td>Unit dedicated-infrastructure cost</td>
<td>$\pi_{L,db}$</td>
<td>$$/km-hr$</td>
<td>90</td>
</tr>
<tr>
<td>Unit mixed-infrastructure cost</td>
<td>$\pi_{L,mb}$</td>
<td>$$/km-hr$</td>
<td>9</td>
</tr>
<tr>
<td>Unit vehicle cost</td>
<td>$\pi_B$</td>
<td>$$/veh-hr$</td>
<td>40</td>
</tr>
<tr>
<td>Unit distance cost</td>
<td>$\pi_V$</td>
<td>$$/veh-km$</td>
<td>2</td>
</tr>
<tr>
<td>Unit mileage cost</td>
<td>$\pi_D$</td>
<td>$$/km$</td>
<td>0.3</td>
</tr>
<tr>
<td>Value of time</td>
<td>$\alpha$</td>
<td>$$/pax-hr$</td>
<td>20</td>
</tr>
<tr>
<td>Minimum headway</td>
<td>$H_{min}$</td>
<td>min</td>
<td>3</td>
</tr>
<tr>
<td>Bus passenger capacity</td>
<td>$C$</td>
<td>pax/veh</td>
<td>150</td>
</tr>
<tr>
<td>Average car waiting time</td>
<td>$W_c$</td>
<td>min</td>
<td>7</td>
</tr>
<tr>
<td>Parameter of the Logit model</td>
<td>$\theta$</td>
<td>$\odot$</td>
<td>1.5</td>
</tr>
</tbody>
</table>

Table 5.2: Experimental settings.

Notice that, due to the rectangular structure of the street network, the E-W traffic has a longer trip length than the N-S one in all considered patterns, as shown by the tails of the trip length distributions (Figs. 5.5b, 5.5d, 5.5f, and 5.5h). Also note that the trip length distribution spreads the most for the uniform trip length pattern, given that the origins and destinations are distributed uniformly all across the whole network. As the origins and destinations become concentrated in a certain region of the network, the range of trip lengths gets reduced.

Table 5.3 shows the optimal design parameters, operator and user costs, as well as the corresponding mode choice, for each demand scenario, type of user behavior, and trip length pattern. The headways are given in min-
Figure 5.5: Concentration of user origins and destinations, and the corresponding trip length pattern in case of: (a-b) uniform city; (c-d) a monocentric city; (e-f) a commuter city; and (g-h) a twin city.
utes, whereas the costs are expressed in hours. The results suggest that for longer trip length patterns and higher demand levels more users tend to choose the bus system. The reason for this is two-fold. First, the mileage cost component (besides the in-vehicle time traveled) in the car user cost function (defined as in \cite{59, 60, 129}) linearly increases with an increase in the trip length (see Eq. 5.22). This, however, is not the case for the bus mode, as all the cost components (except for the in-vehicle time traveled) remain the same, i.e. they are independent of the user trip length. Second, by accounting for the interactions between the modes and their effects on the traffic performance, we observe that the system can benefit if more users select the bus mode, as otherwise the car delay would increase due to a higher car accumulation in the network. This, in turn, would also affect the bus system along mixed lanes, ultimately leading to overall lower network performance. To cope with increasing number of users in the system in scenario SC-II and ensure sufficient passenger capacity, the optimal bus network configuration requires a more frequent service compared to SC-I, as shown in Table 5.3. This holds true for both types of user behavior. Nevertheless, it is worth mentioning that the same pattern might not be observed in real-life scenarios in case the utility functions are significantly different from those used in the numerical experiments.

Notice from Table 5.3 that (i) both the user behavior and the trip length pattern do affect the optimal bus network configuration and passenger mode choice, and that (ii) the optimal design parameters are not constant across cardinal directions. This is an important insight, given that most scientific literature on the bus network design problem assumes, up to now, unique design parameters across the network and one particular type of user behavior and trip length pattern. Therefore, in the following, we demonstrate how important is to relax these assumptions. In particular, we quantify the value of each of the proposed methodological extensions including: (i) modeling of mode choice at the trip length level; (ii) considering heterogeneous types of user behavior; (iii) considering heterogeneous trip length patterns; (iv) accounting for the complex modal interactions; and (v) considering heterogeneous design parameters across cardinal directions.
Table 5.3: Optimal design parameters, operator cost, user cost, and the corresponding mode choice.

<table>
<thead>
<tr>
<th>Type</th>
<th>SC-I</th>
<th>SC-II</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1050</td>
<td>6.460</td>
<td>6.820</td>
</tr>
<tr>
<td>6200</td>
<td>6.540</td>
<td>6.360</td>
</tr>
<tr>
<td>9600</td>
<td>6.350</td>
<td>6.400</td>
</tr>
<tr>
<td>1900</td>
<td>6.510</td>
<td>6.340</td>
</tr>
<tr>
<td>6900</td>
<td>6.250</td>
<td>6.450</td>
</tr>
<tr>
<td>1900</td>
<td>6.370</td>
<td>6.320</td>
</tr>
<tr>
<td>6800</td>
<td>6.280</td>
<td>6.410</td>
</tr>
<tr>
<td>6600</td>
<td>6.340</td>
<td>6.430</td>
</tr>
<tr>
<td>6600</td>
<td>6.340</td>
<td>6.430</td>
</tr>
<tr>
<td>6900</td>
<td>6.260</td>
<td>6.460</td>
</tr>
<tr>
<td>6200</td>
<td>6.360</td>
<td>6.400</td>
</tr>
<tr>
<td>9600</td>
<td>6.350</td>
<td>6.400</td>
</tr>
<tr>
<td>1900</td>
<td>6.510</td>
<td>6.340</td>
</tr>
<tr>
<td>6900</td>
<td>6.250</td>
<td>6.450</td>
</tr>
<tr>
<td>1900</td>
<td>6.370</td>
<td>6.320</td>
</tr>
<tr>
<td>6800</td>
<td>6.280</td>
<td>6.410</td>
</tr>
<tr>
<td>6600</td>
<td>6.340</td>
<td>6.430</td>
</tr>
<tr>
<td>6600</td>
<td>6.340</td>
<td>6.430</td>
</tr>
<tr>
<td>6900</td>
<td>6.260</td>
<td>6.460</td>
</tr>
<tr>
<td>6200</td>
<td>6.360</td>
<td>6.400</td>
</tr>
<tr>
<td>9600</td>
<td>6.350</td>
<td>6.400</td>
</tr>
<tr>
<td>1900</td>
<td>6.510</td>
<td>6.340</td>
</tr>
<tr>
<td>6900</td>
<td>6.250</td>
<td>6.450</td>
</tr>
<tr>
<td>1900</td>
<td>6.370</td>
<td>6.320</td>
</tr>
<tr>
<td>6800</td>
<td>6.280</td>
<td>6.410</td>
</tr>
<tr>
<td>6600</td>
<td>6.340</td>
<td>6.430</td>
</tr>
<tr>
<td>6600</td>
<td>6.340</td>
<td>6.430</td>
</tr>
<tr>
<td>6900</td>
<td>6.260</td>
<td>6.460</td>
</tr>
<tr>
<td>6200</td>
<td>6.360</td>
<td>6.400</td>
</tr>
<tr>
<td>9600</td>
<td>6.350</td>
<td>6.400</td>
</tr>
<tr>
<td>1900</td>
<td>6.510</td>
<td>6.340</td>
</tr>
<tr>
<td>6900</td>
<td>6.250</td>
<td>6.450</td>
</tr>
<tr>
<td>1900</td>
<td>6.370</td>
<td>6.320</td>
</tr>
<tr>
<td>6800</td>
<td>6.280</td>
<td>6.410</td>
</tr>
<tr>
<td>6600</td>
<td>6.340</td>
<td>6.430</td>
</tr>
<tr>
<td>6600</td>
<td>6.340</td>
<td>6.430</td>
</tr>
<tr>
<td>6900</td>
<td>6.260</td>
<td>6.460</td>
</tr>
<tr>
<td>6200</td>
<td>6.360</td>
<td>6.400</td>
</tr>
<tr>
<td>9600</td>
<td>6.350</td>
<td>6.400</td>
</tr>
<tr>
<td>1900</td>
<td>6.510</td>
<td>6.340</td>
</tr>
<tr>
<td>6900</td>
<td>6.250</td>
<td>6.450</td>
</tr>
<tr>
<td>1900</td>
<td>6.370</td>
<td>6.320</td>
</tr>
<tr>
<td>6800</td>
<td>6.280</td>
<td>6.410</td>
</tr>
<tr>
<td>6600</td>
<td>6.340</td>
<td>6.430</td>
</tr>
<tr>
<td>6600</td>
<td>6.340</td>
<td>6.430</td>
</tr>
<tr>
<td>6900</td>
<td>6.260</td>
<td>6.460</td>
</tr>
<tr>
<td>6200</td>
<td>6.360</td>
<td>6.400</td>
</tr>
<tr>
<td>9600</td>
<td>6.350</td>
<td>6.400</td>
</tr>
<tr>
<td>1900</td>
<td>6.510</td>
<td>6.340</td>
</tr>
<tr>
<td>6900</td>
<td>6.250</td>
<td>6.450</td>
</tr>
<tr>
<td>1900</td>
<td>6.370</td>
<td>6.320</td>
</tr>
<tr>
<td>6800</td>
<td>6.280</td>
<td>6.410</td>
</tr>
<tr>
<td>6600</td>
<td>6.340</td>
<td>6.430</td>
</tr>
<tr>
<td>6600</td>
<td>6.340</td>
<td>6.430</td>
</tr>
<tr>
<td>6900</td>
<td>6.260</td>
<td>6.460</td>
</tr>
<tr>
<td>6200</td>
<td>6.360</td>
<td>6.400</td>
</tr>
<tr>
<td>9600</td>
<td>6.350</td>
<td>6.400</td>
</tr>
<tr>
<td>1900</td>
<td>6.510</td>
<td>6.340</td>
</tr>
<tr>
<td>6900</td>
<td>6.250</td>
<td>6.450</td>
</tr>
<tr>
<td>1900</td>
<td>6.370</td>
<td>6.320</td>
</tr>
<tr>
<td>6800</td>
<td>6.280</td>
<td>6.410</td>
</tr>
<tr>
<td>6600</td>
<td>6.340</td>
<td>6.430</td>
</tr>
<tr>
<td>6600</td>
<td>6.340</td>
<td>6.430</td>
</tr>
<tr>
<td>6900</td>
<td>6.260</td>
<td>6.460</td>
</tr>
<tr>
<td>6200</td>
<td>6.360</td>
<td>6.400</td>
</tr>
<tr>
<td>9600</td>
<td>6.350</td>
<td>6.400</td>
</tr>
<tr>
<td>1900</td>
<td>6.510</td>
<td>6.340</td>
</tr>
<tr>
<td>6900</td>
<td>6.250</td>
<td>6.450</td>
</tr>
<tr>
<td>1900</td>
<td>6.370</td>
<td>6.320</td>
</tr>
<tr>
<td>6800</td>
<td>6.280</td>
<td>6.410</td>
</tr>
<tr>
<td>6600</td>
<td>6.340</td>
<td>6.430</td>
</tr>
<tr>
<td>6600</td>
<td>6.340</td>
<td>6.430</td>
</tr>
<tr>
<td>6900</td>
<td>6.260</td>
<td>6.460</td>
</tr>
<tr>
<td>6200</td>
<td>6.360</td>
<td>6.400</td>
</tr>
<tr>
<td>9600</td>
<td>6.350</td>
<td>6.400</td>
</tr>
<tr>
<td>1900</td>
<td>6.510</td>
<td>6.340</td>
</tr>
<tr>
<td>6900</td>
<td>6.250</td>
<td>6.450</td>
</tr>
<tr>
<td>1900</td>
<td>6.370</td>
<td>6.320</td>
</tr>
<tr>
<td>6800</td>
<td>6.280</td>
<td>6.410</td>
</tr>
<tr>
<td>6600</td>
<td>6.340</td>
<td>6.430</td>
</tr>
<tr>
<td>6600</td>
<td>6.340</td>
<td>6.430</td>
</tr>
<tr>
<td>6900</td>
<td>6.260</td>
<td>6.460</td>
</tr>
<tr>
<td>6200</td>
<td>6.360</td>
<td>6.400</td>
</tr>
<tr>
<td>9600</td>
<td>6.350</td>
<td>6.400</td>
</tr>
<tr>
<td>1900</td>
<td>6.510</td>
<td>6.340</td>
</tr>
<tr>
<td>6900</td>
<td>6.250</td>
<td>6.450</td>
</tr>
<tr>
<td>1900</td>
<td>6.370</td>
<td>6.320</td>
</tr>
<tr>
<td>6800</td>
<td>6.280</td>
<td>6.410</td>
</tr>
<tr>
<td>6600</td>
<td>6.340</td>
<td>6.430</td>
</tr>
<tr>
<td>6600</td>
<td>6.340</td>
<td>6.430</td>
</tr>
<tr>
<td>6900</td>
<td>6.260</td>
<td>6.460</td>
</tr>
<tr>
<td>6200</td>
<td>6.360</td>
<td>6.400</td>
</tr>
<tr>
<td>9600</td>
<td>6.350</td>
<td>6.400</td>
</tr>
</tbody>
</table>
5.2.4 Value of the proposed methodological extensions

5.2.4.1 Value of modeling the mode choice at the trip length level

In the previous subsection, we showed that the probability of choosing any given mode depends not only on the demand intensity, but also the type of user behavior. A closer look into the mode choice distribution reveals that this probability is not constant across the user trip lengths, but follows certain distribution (Fig. 5.6). This distribution is not unique, but varies across the trip length patterns, indicating the importance of modeling the mode choice at the trip length level. Interestingly, the distribution for Type 2 behavior shows higher deviation and is shifted towards the right (i.e. higher number of users choose to use the bus system) compared to Type 1. The reason for this is that the bus system designed according to Type 2 behavior provides higher level of service, as elaborated further below. Note that, since the resulting mode choice distributions for both demand scenarios exhibit a similar shape, for brevity, we only show that of SC-I.

![Graphs](image-url)

**Figure 5.6:** Mode choice distribution for the optimal bus network design in scenario SC-I in case of: (a) uniform city; (b) a mono-centric city; (c) a commuter city; and (d) a twin city.
5.2.4.2 Value of considering heterogeneous types of user behavior

For quantifying the importance of considering heterogeneous types of user behavior, we compare the optimal values of the cost functions for Type 1 and Type 2 behavior, such that

\[ \Delta Z_O = \frac{(Z_{2,O} - Z_{1,O})}{Z_{1,O}}, \quad \Delta Z_U = \frac{(Z_{2,U} - Z_{1,U})}{Z_{1,U}}, \quad \text{and} \quad \Delta Z = \frac{(Z_2 - Z_1)}{Z_1}. \]

The results of this comparison are given in Table 5.4. We observe that for all trip length patterns users can benefit (according to the cost functions used here) if they are willing to adjust the number of transfers to minimize the walking distance at the origin and the destination. This is consistent with empirical results from the city of Barcelona (63). Potential improvements can range from 4.2% to 14.8%, depending on the demand level and prevailing trip length pattern (see Table 5.4).

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Trip pattern</th>
<th>( \Delta Z_O )</th>
<th>( \Delta Z_U )</th>
<th>( \Delta Z )</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC-I</td>
<td>(i)</td>
<td>+4.46%</td>
<td>-8.70%</td>
<td>-7.28%</td>
</tr>
<tr>
<td></td>
<td>(ii)</td>
<td>-1.02%</td>
<td>-12.74%</td>
<td>-11.02%</td>
</tr>
<tr>
<td></td>
<td>(iii)</td>
<td>-11.62%</td>
<td>-4.27%</td>
<td>-5.15%</td>
</tr>
<tr>
<td></td>
<td>(iv)</td>
<td>-21.24%</td>
<td>-8.90%</td>
<td>-10.66%</td>
</tr>
<tr>
<td>SC-II</td>
<td>(i)</td>
<td>-5.15%</td>
<td>-7.62%</td>
<td>-7.34%</td>
</tr>
<tr>
<td></td>
<td>(ii)</td>
<td>+6.89%</td>
<td>-14.81%</td>
<td>-11.72%</td>
</tr>
<tr>
<td></td>
<td>(iii)</td>
<td>-3.30%</td>
<td>-4.22%</td>
<td>-4.12%</td>
</tr>
<tr>
<td></td>
<td>(iv)</td>
<td>-26.85%</td>
<td>-6.49%</td>
<td>-9.43%</td>
</tr>
</tbody>
</table>

Table 5.4: Effects of user behavior on the cost functions.

To motivate the users to use the transfers, i.e. provide a well-designed transfer-based network, the user waiting time needs to be minimized. This is indicated by the optimal decision variables in Table 5.3, where we can see that the optimal bus network configuration for Type 2 behavior provides higher bus frequency (i.e. lower headways in the two directions for almost all types of cities in both tested scenarios) compared to Type 1. Consequently, the operator cost might increase for some trip length patterns (e.g. uniform city in SC-I or mono-centric city in SC-II), as shown in Table 5.4. However, since the share of the operator cost within the total system cost is substantially lower than that of the user cost, the total cost function still gets reduced if the network is designed according to Type 2 user behavior. The improvements can range from 5.1% to 11.7% (see Table 5.4).
5.2.4.3 Value of considering heterogeneous trip length patterns

Here we quantify the importance of considering heterogeneous trip length patterns by comparing the costs obtained when the bus network parameters are optimally determined for a given trip length pattern, with those obtained if we were to assume uniformly distributed origins and destinations (Table 5.5). A negative value means that accounting for the actual trip length distribution does lead to lower costs. As expected, such lower total costs are achieved for both types of user behavior, with a similar magnitude of improvements. The improvements tend to increase with an increase in the demand level, especially for a commuter city (i.e. longest trip length pattern) that exhibits the highest improvement (5.9% for Type 2; 5.6% for Type 2) in the total cost function. In contrast, the lowest improvements (0.1% for Type 1; 1.9% for Type 2) are acquired for a twin city.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>User behavior</th>
<th>Trip pattern</th>
<th>$\Delta Z_O$</th>
<th>$\Delta Z_U$</th>
<th>$\Delta Z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC-I</td>
<td>Type 1</td>
<td>(ii)</td>
<td>-5.48%</td>
<td>-0.55%</td>
<td>-1.31%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iii)</td>
<td>+109.93%</td>
<td>-9.89%</td>
<td>-3.30%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iv)</td>
<td>+38.46%</td>
<td>-4.67%</td>
<td>-0.15%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(ii)</td>
<td>-10.44%</td>
<td>-1.22%</td>
<td>-2.86%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iii)</td>
<td>+57.35%</td>
<td>-8.04%</td>
<td>-3.58%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iv)</td>
<td>-1.36%</td>
<td>-1.97%</td>
<td>-1.89%</td>
</tr>
<tr>
<td>SC-II</td>
<td>Type 1</td>
<td>(ii)</td>
<td>-11.68%</td>
<td>+0.39%</td>
<td>-1.52%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iii)</td>
<td>+79.41%</td>
<td>-10.59%</td>
<td>-5.64%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iv)</td>
<td>+31.48%</td>
<td>-4.33%</td>
<td>-0.42%</td>
</tr>
<tr>
<td></td>
<td>Type 2</td>
<td>(ii)</td>
<td>+3.88%</td>
<td>-3.66%</td>
<td>-2.44%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iii)</td>
<td>+66.09%</td>
<td>-10.50%</td>
<td>-5.93%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iv)</td>
<td>-3.87%</td>
<td>-1.60%</td>
<td>-1.87%</td>
</tr>
</tbody>
</table>

Table 5.5: Effects of trip length patterns on the cost functions.

It is worth noting that in some cases the operator cost increases when we use the actual trip length distribution, especially for a commuter city. The reason for this is two-fold. First, the optimal bus network design for a uniform city does not satisfy the constraint for the maximum bus occupancy (Eq. 5.11) when applied to a commuter city, thus cannot be considered to be feasible. From that perspective, an increase in the operator cost (imposed by considering the actual trip length distribution) for a commuter city is justified by ensuring that the bus capacity is sufficient to serve the
maximum bus occupancy during the peak hour. Second, a commuter city requires a bus network with a higher spatial coverage and more frequent bus service compared to the uniform one. Such a requirement, however, ensures that the level of service provided by the bus system is competitive with that of the automobile at a reasonable cost. This is confirmed with the user cost function, which indeed gets reduced when we design the bus network for the actual trip length pattern. Consequently, the total system cost also gets reduced, given that the share of the user costs within the total system cost is substantially higher than that of the operator cost, as we stated before. In other words, increases in the operator cost are compensated by a much more efficient level of service from the user perspective, leading to overall lower costs.

5.2.4.4 Value of accounting for the complex modal interactions

So far, we have discussed how the demand intensity, user behavior, and trip length patterns affect the optimal bus network configuration and quantified the value of considering these factors. To realistically model both, the car and the bus system, we employed the 3D-MFD to take into account the interactions between the modes and their impact on the traffic performance. These interactions, however, are ignored in most scientific literature on the bus network design problem, which typically assumes that the travel times (or equivalently, the bus speeds) are independent of the design parameters. It is therefore the purpose of this subsection to quantify the value of considering the complex modal interactions in the proposed modeling framework.

We do this by comparing the results of the proposed approach to those obtained when the optimal design parameters for the simplified problem (i.e. considering the bus system only and using the average car speed during the peak period as the bus cruising speed along mixed lanes, as in (59) and (60)) are incorporated into the proposed framework. To make a fair comparison between the two approaches, we use the same demand level for the bus mode (determined from the mode choice in the original problem). The results of this comparison are shown in Table 5.6. A negative value means that a proper modeling of the complex bi-modal interactions does lead to lower costs. As expected, such lower total costs are achieved for both types of user behavior and all trip length patterns. Notice that the improvements in the total system cost increase with an increase in the demand level, especially for Type 1 behavior. These improvements come from the significant improvements (up to 18%) in the user cost function.
<table>
<thead>
<tr>
<th>Scenario</th>
<th>User behavior</th>
<th>Trip pattern</th>
<th>$\Delta Z_O$</th>
<th>$\Delta Z_{UI}$</th>
<th>$\Delta Z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC-I</td>
<td>Type 1</td>
<td>(i)</td>
<td>+123.13%</td>
<td>-7.15%</td>
<td>-0.92%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(ii)</td>
<td>+110.91%</td>
<td>-11.17%</td>
<td>-2.90%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iii)</td>
<td>+86.47%</td>
<td>-10.87%</td>
<td>-4.95%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iv)</td>
<td>+106.08%</td>
<td>-11.71%</td>
<td>-3.71%</td>
</tr>
<tr>
<td></td>
<td>Type 2</td>
<td>(i)</td>
<td>+111.59%</td>
<td>-9.76%</td>
<td>-3.01%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(ii)</td>
<td>+89.49%</td>
<td>-13.98%</td>
<td>-5.54%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iii)</td>
<td>+68.76%</td>
<td>-12.04%</td>
<td>-7.09%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iv)</td>
<td>+61.59%</td>
<td>-16.46%</td>
<td>-10.95%</td>
</tr>
<tr>
<td>SC-II</td>
<td>Type 1</td>
<td>(i)</td>
<td>+96.99%</td>
<td>-7.49%</td>
<td>-1.72%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(ii)</td>
<td>+74.14%</td>
<td>-10.64%</td>
<td>-3.98%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iii)</td>
<td>+46.70%</td>
<td>-11.35%</td>
<td>-7.52%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iv)</td>
<td>+94.44%</td>
<td>-12.30%</td>
<td>-4.76%</td>
</tr>
<tr>
<td></td>
<td>Type 2</td>
<td>(i)</td>
<td>+101.26%</td>
<td>-11.47%</td>
<td>-5.47%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(ii)</td>
<td>+102.59%</td>
<td>-18.07%</td>
<td>-8.7%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iii)</td>
<td>+46.66%</td>
<td>-12.87%</td>
<td>-8.98%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iv)</td>
<td>+46.11%</td>
<td>-14.24%</td>
<td>-9.91%</td>
</tr>
</tbody>
</table>

**Table 5.6:** Effects of modal interactions on the cost functions.
That being said, we also observe an increase in the operator cost. This is because the optimal solution of the simplified problem produces a bus network design that operates in a fully mixed (with car traffic) environment across all cardinal directions, with a lower service frequency compared to that of the original problem. Consequently, once the solution of the simplified approach is incorporated into the proposed framework, the user waiting time increase and the occupancy rates become close or equal to the bus capacity (indicating lower level of service) for all demand levels, types of user behavior, and trip length patterns. In contrast, by accounting for the effects of the bus system on private car users, and vice versa, especially for higher demand levels, the proposed framework leads to a bus network design with higher percentage of dedicated bus lines. Due to reduced interactions between the modes, such design allows to increase the bus frequency along dedicated bus segments, improving thereby the level of service (reducing the passenger waiting time and bus occupancy) compared to the simplified approach. This indicates the value of the proposed modeling framework in capturing the necessary factors for determining the optimal bus network configuration, while taking into account multiple transport modes.

5.2.4.5 Value of considering heterogeneous design parameters

The results in Table 5.3 indicate that the optimal bus network configuration for different trip length patterns does not always provide a fully dedicated or mixed (with car traffic) bus system across the entire network (i.e. all cardinal directions), as it was the case in previous studies (see e.g. 59, 60, 65, 129). In fact, most of the decision variables vary along E-W and N-S. This holds true for both types of user behavior. Hence, it is worth quantifying the importance of considering a heterogeneous bus network design across cardinal directions.

We perform this analysis by comparing the results of the proposed approach to those obtained when the optimal design parameters for the simplified problem (i.e. considering the same stop spacing, bus headway, and lane allocation across all cardinal directions, as in 60) are incorporated into the proposed framework. The results of this comparison are shown in Table 5.7. Similarly to the previous analysis, a negative value means that a heterogeneous bus network design does lead to lower costs. We observe that the operator costs can indeed be reduced in both tested scenarios, for both types of user behavior and all trip length patterns, if we allow the decision variables to vary across cardinal directions. Potential improvements
Table 5.7: Effects of heterogeneous design parameters across cardinal directions on the cost functions.

can range from 5.5% to 12.6%. Notice that there are almost no effects on the user cost function, resulting in marginal improvements in the total system cost (see Table 5.7). This shows that the proposed framework provides more flexibility for the bus system to serve the passenger demand while reducing the operator cost compared to the existing approaches.

5.3 Summary

In this chapter, we investigate the effects of demand intensity, user behavior, and trip length patterns on the optimal bus network design, as well as the effects that the bus network structure might have on the traffic performance and passenger mode choice. For this purpose, we introduce several major extensions to the classical problem formulation. First, we consider two types of user behavior when selecting bus routes: (i) minimization of the number of transfers (Type 1); and (ii) minimization of the walking distance (Type 2). Second, we propose a new concept for considering trip length heterogeneity related to different origin-destination trips in the
network. Third, we explicitly account for mixed traffic (i.e. mutual influence of cars and buses in the network) and mode choice at the trip length level. Forth, we use a dynamic description of the peak hour demand by distinguishing between the loading and unloading phase. This way, we are able to achieve a more realistic modeling of network congestion, hence determine a mean speed that depends on the bus network configuration, demand profile, and passenger mode choice.

Numerical experiments performed for different trip length patterns, reflecting different spatial demand distributions, show that all the tested factors do affect the optimal bus network configuration and passenger mode choice. Results reveal that the probability of choosing any given mode is not constant across the user trip lengths, i.e. it follows certain distribution. This distribution is not unique, but varies across the trip length patterns, indicating the importance of modeling the mode choice at the trip length level. Moreover, the results indicate that a well-designed transfer-based bus network, i.e. a bus network designed according to Type 2 behavior, leads to lower costs and attracts more users for the same level of passenger demand compared to that of Type 1. This finding is consistent with empirical results from the city of Barcelona (63). Finally, we demonstrate the significance of addressing simplifications made in previous studies. From that perspective, we show that the bus network parameters optimally determined for a given trip length pattern lead to lower costs compared to those obtained if we were to assume uniformly distributed origins and destinations. More importantly, the optimal bus network design determined for the uniform trip pattern does not satisfy the constraint for the maximum bus occupancy when applied to other trip length patterns, thus cannot be considered to be feasible. In other words, accounting for the actual trip length distribution allows the operator to design the bus network in accordance to the prevailing demand pattern. A comparison with a simplified approach that considers the bus system only and uses the average car speed during the peak period as the bus cruising speed along mixed lanes (as in 59, 60) reveals the value of accounting for the complex modal interactions, especially for higher demand levels. Last, we show that by considering heterogeneous design parameters across cardinal directions we provide more flexibility for the bus system to serve the passenger demand while reducing the operator cost compared to the existing approaches.
In the previous chapter, we developed a methodology to determine the optimal configuration of the public transport system, which provides a level of service competitive to that of the automobile at a reasonable cost. As such, the proposed framework can be very beneficial to both, the policy makers and the practitioners, when constructing new or redesigning the existing public transport network in a given city. However, due to the limited road space in urban areas, it is often not possible to optimally (re)design public transport networks to keep pace with the everlasting increase in travel demand. Therefore, to optimize the performance of multi-modal systems and maximize the utilization of the transportation infrastructure, the focus should be placed on implementing smart network-level management strategies. Such strategies should ensure that public transport continues to serve as the backbone of urban mobility.

For this purpose, in this chapter, we propose a novel flexible bus dispatching system in which a fleet of fully automated modular bus units, together with conventional buses, serves the passenger demand. Such concept offers public transport operators new perspectives and enormous flexibility to better manage the allocation of the vehicle resources and reduce the operating cost. The remainder of this chapter is mainly based on the following publication:

6.1 General Methodology

In this section, we present all the elements of the methodological framework, including the network representation based on the 3D-MFD (Section 6.1.1), the proposed optimization model (Section 6.1.2), and the macroscopic modeling approach from both vehicular (Sections 6.1.3) and passenger (Section 6.1.4) perspective. For the readers’ convenience, Table 6.1 provides the list of the most important notation used in this chapter.

6.1.1 System description and network representation

As in the previous chapters, we consider here a multi-modal urban network consisting of buses and cars denoted as \( b \) and \( c \), respectively. A bus operator manages a set of bus lines \( L \), indexed by \( l \), with two types of buses: (i) conventional buses denoted as \( r \); and (ii) modular buses denoted as \( m \). The bus type is indexed by \( i \in I = \{r, m\} \). Each modular bus can include \( u \in U_{m,l} = \{1, ..., |U_{m,l}|\} \) modular bus units as shown in Fig. 6.1, where \( |U_{m,l}| \) is the maximum number of modular bus units that can be combined as a single modular bus on line \( l \). Note that the modular bus units are fully automated, thus there is no cost for assigning bus drivers to them (similarly to the Next Future Mobility system\(^1\)). To simplify the notation, each conventional bus is considered as an individual conventional bus unit, i.e. \( |U_{r,l}| = 1, \forall l \in L \) (see Fig. 6.1). In addition, all units of the same type \( i \) are considered to be equivalent in terms of operational characteristics (e.g. energy consumption, vehicle dynamics, etc.) and passenger capacity \( C_i \), such that the capacity of the conventional bus unit is \( C_r \) and the capacity of the modular bus unit is \( C_m \leq C_r \). This, in turn, implies that the capacity of the conventional bus is also \( C_r \), whereas the capacity of a modular bus consisting of \( u \) modular bus units is \( u \cdot C_m \). Finally, we assume that each conventional bus is equivalent to \( \zeta \geq 1 \) modular bus units (in terms of capacity), i.e. \( C_r = \zeta \cdot C_m \). The size of the bus fleet \( F \) can therefore be expressed either in terms of the equivalent number of conventional \( (F_r + F_m / \zeta) \) or modular \( (F_r \cdot \zeta + F_m) \) bus units. This gives the following formulation for the penetration rate of each type of bus units, \( p_i \),

\(^1\) www.next-future-mobility.com
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mathcal{L} )</td>
<td>set of bus lines indexed by ( l )</td>
</tr>
<tr>
<td>( \mathcal{S}_l )</td>
<td>set of segments along line ( l ) indexed by ( s )</td>
</tr>
<tr>
<td>( \mathcal{T}_i )</td>
<td>set of types of bus units indexed by ( i )</td>
</tr>
<tr>
<td>( \mathcal{T} )</td>
<td>set of time intervals indexed by ( t )</td>
</tr>
<tr>
<td>( \tau )</td>
<td>length of time interval</td>
</tr>
<tr>
<td>( F_i )</td>
<td>total number of available bus units of type ( i )</td>
</tr>
<tr>
<td>( C_i )</td>
<td>passenger capacity of a bus unit of type ( i )</td>
</tr>
<tr>
<td>( \pi_i )</td>
<td>operating cost of a bus unit of type ( i )</td>
</tr>
<tr>
<td>( \eta_i )</td>
<td>bus unit-car equivalent for a bus unit of type ( i )</td>
</tr>
<tr>
<td>(</td>
<td>U_{i,l}</td>
</tr>
<tr>
<td>( L_b/L_c )</td>
<td>length of the bus/car network</td>
</tr>
<tr>
<td>( \ell_b/\ell_c/\ell_b^t )</td>
<td>average trip length of cars/buses/bus passengers</td>
</tr>
<tr>
<td>( v_{MAX} )</td>
<td>free-flow speed of buses and cars</td>
</tr>
<tr>
<td>( v_b(t)/v_c(t) )</td>
<td>average bus/car speed in the network during ( t )</td>
</tr>
<tr>
<td>( Q_b(t)/Q_c(t) )</td>
<td>circulating flow for the bus/car mode given by the 3D-MFD during ( t )</td>
</tr>
<tr>
<td>( G_b(t)/G_c(t) )</td>
<td>trip completion flow for the bus/car mode given by the 3D-MFD during ( t )</td>
</tr>
<tr>
<td>( \lambda_{INT}(t)/\lambda_{EXT}(t) )</td>
<td>internal/external car demand during ( t )</td>
</tr>
<tr>
<td>( N_b(t)/N_c(t) )</td>
<td>bus/car accumulation in the entire network at the onset of ( t )</td>
</tr>
<tr>
<td>( N_b^{MAX}(t)/N_b^{DED}(t) )</td>
<td>bus accumulation in the mixed/dedicated subnetwork at the onset of ( t )</td>
</tr>
<tr>
<td>( \bar{N}_b )</td>
<td>maximum car accumulation in the car only and mixed subnetworks</td>
</tr>
<tr>
<td>( \ell_{b,i,l}/\ell_{b,i,l}^t )</td>
<td>length of segment ( s )/average distance between bus stops on segment ( s ) of line ( l )</td>
</tr>
<tr>
<td>( \Theta_{i,l}(t) )</td>
<td>dwell time of buses of type ( i ) on segment ( s ) of line ( l ) during ( t )</td>
</tr>
<tr>
<td>( v_{b,i,l}(t) )</td>
<td>bus speed on segment ( s ) of line ( l ) during ( t )</td>
</tr>
<tr>
<td>( \phi_{c,i,l}(t) )</td>
<td>receiving flow of cars to segment ( s ) of line ( l ) during ( t )</td>
</tr>
<tr>
<td>( \psi_{b,i,l}(t)/\Psi_{b,i,l}(t) )</td>
<td>dispatched flow of bus units/buses of type ( i ) of line ( l ) during ( t )</td>
</tr>
<tr>
<td>( \phi_{b,i,l,s}(t)/\Phi_{b,i,l,s}(t) )</td>
<td>transferring flow of bus units/buses of type ( i ) between ( s ) and ( s+1 ) of line ( l ) during ( t )</td>
</tr>
<tr>
<td>( n_{b,i,l}(t)/N_{b,i,l}(t) )</td>
<td>accumulation of bus units/buses of type ( i ) on ( s ) of line ( l ) at the onset of ( t )</td>
</tr>
<tr>
<td>( u_{i,l}(t) )</td>
<td>average number of combined units of type ( i ) on segment ( s ) of line ( l ) at the onset of ( t )</td>
</tr>
<tr>
<td>( \bar{N}_{i,l}(t) )</td>
<td>maximum car accumulation that can be accommodated on ( s ) of line ( l ) at the onset of ( t )</td>
</tr>
<tr>
<td>( N_{i,c,l}(t) )</td>
<td>accumulation of cars on ( s ) of line ( l ) at the onset of ( t )</td>
</tr>
<tr>
<td>( B_{i,l}(t) )</td>
<td>total number of boarding passengers on ( s ) of line ( l ) during ( t )</td>
</tr>
<tr>
<td>( A_{i,l}(t) )</td>
<td>total number of alighting passengers on ( s ) of line ( l ) during ( t )</td>
</tr>
<tr>
<td>( P_{i,s,s'}(t) )</td>
<td>number of on-board passengers on ( s ) with destination on ( s' ) of line ( l ) at the onset of ( t )</td>
</tr>
<tr>
<td>( R_{i,s,s'}(t) )</td>
<td>trip completion flow of bus passengers on ( s ) with destination on ( s' ) of line ( l ) during ( t )</td>
</tr>
<tr>
<td>( D_{i,s,s'}(t) )</td>
<td>number of boarding passengers on ( s ) with destination on ( s' ) of line ( l ) during ( t )</td>
</tr>
<tr>
<td>( O_{i,s,s'}(t) )</td>
<td>outflow of passengers between ( s ) and ( s+1 ) with destination on ( s' ) of line ( l ) during ( t )</td>
</tr>
<tr>
<td>( \omega_{i,s,s'}(t) )</td>
<td>number of unserved passengers on ( s ) with destination on ( s' ) of line ( l ) at the onset of ( t )</td>
</tr>
<tr>
<td>( \lambda_{i,s,s'}(t) )</td>
<td>average arrival rate of bus passengers on ( s ) with destination on ( s' ) of line ( l ) during ( t )</td>
</tr>
<tr>
<td>( \delta_{i,s,s'}(t) )</td>
<td>share of passengers on ( s ) with destination on ( s' ) of line ( l ) during ( t )</td>
</tr>
<tr>
<td>( \rho_{b,i,l,s,s'}(t) )</td>
<td>average bus occupancy of passengers on ( s ) with destination on ( s' ) of line ( l ) during ( t )</td>
</tr>
</tbody>
</table>

**Table 6.1: Nomenclature.**
defined as the share of the number of bus units of a given type \( i \) within the total number of equivalent bus units, i.e.:

\[
p_i = \begin{cases} 
  \frac{(F_r \cdot \zeta)}{(F_r \cdot \zeta + F_m)}, & \text{if } i = r, \\
  \frac{F_m}{(F_r \cdot \zeta + F_m)}, & \text{if } i = m.
\end{cases} \tag{6.1}
\]

\[u = |U_{r,l}| = 1\]

\[u = |U_{m,l}|\]

**Figure 6.1:** Schematic representation of the types of bus units that form the bus fleet in the proposed flexible bus dispatching system.

The bus operator is interested in dynamically determining the bus type and the dispatching frequency for each bus line to optimize its utility (defined in the Section 6.1.2) based on historical demand information. The decision should be made for each time interval \( t \in \mathcal{T} \) of length \( \tau \). We make the following assumption on the demand information to which the operator has access.

**Assumption 6.1** (Data provision). We assume that the bus operator has access to demand and routing information of bus passengers at the bus stop level and car demand at the network level.

**Remark.** We make the following remarks regarding Assumption 6.1. First, the bus operator can predict detailed information (demand and routing) on
bus passengers through historical data (e.g. from smart card data, surveillance cameras inside buses, or passenger surveys). Second, detailed information on car traffic (such as OD pairs and route choices of car users) is typically not available to the bus operator due to privacy concerns (146) and the lack of comprehensive monitoring devices. Thus, only aggregated traffic volumes at the network level are available for the car mode.

Given limited data provision for the car mode (as per Assumption 6.1), we abstract the given multi-modal urban network into a macroscopic multi-modal reservoir system (Fig. 6.2a). This level of abstraction allows us to model the traffic dynamics at the network level using the 3D-MFD (Fig. 6.2b). The 3D-MFD can capture the complex multimodal interactions, which include the impact of bus operations (e.g. accumulation, dispatching frequency, dwell times at bus stops along mixed lanes, transit signal priority along dedicated bus lanes, etc.) on the speed of cars, as well as the impact of car accumulation on the operation of buses (e.g. reduced speed along mixed lanes in congested traffic conditions). Moreover, the 3D-MFD implicitly accounts for the OD pairs and route choices of car users. Recently, researchers have demonstrated using both simulation (137) and empirical data (147) that the 3D-MFD model can be successfully applied as a valid and accurate modeling tool to assess various traffic management strategies, such as the frequency setting problem. In this chapter, we apply the 3D-MFD concept to model the vehicular dynamics for the car mode at the network level. Without loss of generality, we assume that the congestion in the network is homogeneous, i.e. it exhibits a well-defined 3D-MFD, as otherwise the network can be partitioned into several homogeneous regions using well-established partitioning algorithms (see e.g. 26, 27). Note, however, that such network-level modeling cannot be directly applied to the bus mode, as the lane allocation might not remain constant along the whole bus line. This, in turn, may affect the operating regime, i.e. the speed of the bus system along a given line. To account for such effects, we model the vehicular dynamics for the bus mode at the segment level. We do this by splitting each bus line $l$ into multiple segments, indexed by $s \in \mathcal{S}_l = \{1, ..., |\mathcal{S}_l|\}$ (Fig. 6.2a), with similar length $\ell_{b,l,s}$, where $|\mathcal{S}_l|$ is the total number of segments along line $l$. These segments are grouped into two categories, i.e. per lane allocation type (mixed or dedicated): $\mathcal{S}_l = \mathcal{S}_{l}^{\text{MIX}} \cup \mathcal{S}_{l}^{\text{DED}}$. Note that, for any given segment, we assign the lane allocation that is more prevalent along its length. Following the rationale of the cell transmission model (CTM) (148), we set the length of the smallest segment across all bus lines, $\min_{l \in L, s \in \mathcal{S}_l} \ell_{b,l,s}$, to be such
Figure 6.2: Schematic representation of the multi-modal reservoir model (a) used to model traffic dynamics based on the 3D-MFD (b).
that it satisfies the following condition: \( \tau \cdot v_{\text{MAX}} \leq \min_{l \in L, s \in S_l} \{ \ell_{b,l,s} \} \), where \( v_{\text{MAX}} \) denotes the free-flow speed of both modes\(^2\). It is also important to note that the proposed approach for modeling traffic dynamics for the bus mode is CTM-inspired, i.e. it is not the CTM, and is only used to account for varying lane allocation along a bus line. That being said, the modal interactions could potentially be captured better if a detailed CTM is used for both modes. However, this would require a detailed information on OD pairs and route choices of car users, which is not available for the considered problem, as stated in Assumption 6.1.

In the following, we formulate the optimization framework used to determine the optimal number of combined modular bus units and the optimal frequency at which the units should be dispatched.

### 6.1.2 Mathematical formulation of the optimization framework

In this subsection, we present the optimization framework for maximizing the efficiency of the proposed bus dispatching system. The decision variables for the considered optimization problem include: (i) the dispatching flow of buses \( \Psi = [\Psi_{b,i,l}(t) : l \in L, i \in I, t \in T] \), defined as the number of dispatching buses (both conventional and modular) per unit time; and (ii) the dispatching flow of bus units \( \psi = [\psi_{b,i,l}(t) : l \in L, i \in I, t \in T] \), defined as the number of dispatching bus units (both conventional and modular) per unit time, across all bus lines and for each time interval. Recall that, for conventional buses, \( \psi_{b,r,l}(t) = \Psi_{b,r,l}(t) \), as each conventional bus is considered as an individual conventional bus unit. This, however, may not be the case for modular buses, given that each modular bus may include multiple modular bus units, i.e. \( \psi_{b,m,l}(t) \geq \Psi_{b,m,l}(t) \). This way, we can obtain the average number of modular bus units contained in a dispatching modular bus on any given line \( l \) during time interval \( t \) as \( u_{m,l}(t) = \psi_{b,m,l}(t) / \Psi_{b,m,l}(t) \). The objective function is to minimize the total system cost \( Z(\Psi, \psi) \), which includes both the operator, \( Z_O(\Psi, \psi) \), and the user cost, \( Z_U(\Psi, \psi) \) (the specific forms of \( Z_O(\Psi, \psi) \) and \( Z_U(\Psi, \psi) \) will be given in Eq. 6.9 and Eq. 6.10, respectively). Mathematically, the optimization problem can be formulated as:

\[
\min_{\Psi, \psi} Z(\Psi, \psi) = Z_O(\Psi, \psi) + Z_U(\Psi, \psi), \tag{6.2}
\]

\(^2\) We assume that buses and cars have the same free-flow speed, as described later in Assumption 6.4.
\begin{align}
\text{s.t. } \psi_{b,i,l}(t) & \geq 0, \quad \forall l \in \mathcal{L}, t \in \mathcal{T}, i \in \mathcal{I}, \\
\Psi^\text{MIN}_b & \leq \psi_{b,i,l}(t) \leq \Psi^\text{MAX}_b, \quad \forall l \in \mathcal{L}, t \in \mathcal{T}, i \in \mathcal{I}, \\
\psi_{b,i,l}(t) - \psi_{b,i,l}(t) & \leq 0, \quad \forall l \in \mathcal{L}, t \in \mathcal{T}, i \in \mathcal{I}, \\
|\mathcal{U}_{i,l}| \cdot \psi_{b,i,l}(t) & - \psi_{b,i,l}(t) \geq 0, \quad \forall l \in \mathcal{L}, t \in \mathcal{T}, i \in \mathcal{I}, \\
n^\text{TOT}_b(t) & \leq F_i, \quad \forall t \in \mathcal{T}, i \in \mathcal{I},
\end{align}

with

\begin{align}
n^\text{TOT}_b(t) &= \sum_{l \in \mathcal{L}} \psi_{b,i,l}(t) \cdot \tau + \sum_{l \in \mathcal{L}} \sum_{s \in \mathcal{S}_l} n_{b,i,l,s}(t) \\
&\quad - \sum_{l \in \mathcal{L}} \frac{n_{b,i,l,s}(t)}{N_b(t)} \cdot G_b(N_b(t), N_c(t)) \cdot \tau,
\end{align}

where \( n^\text{TOT}_b(t) \) is the total number of bus units required for the operation during any time interval \( t \); \( N_b(t) \) and \( N_c(t) \) stand for the total bus and car accumulation, respectively, at the beginning of time interval \( t \); \( n_{b,i,l,s}(t) \) is the accumulation of bus units of type \( i \) on segment \( s \) of line \( l \) at the beginning of time interval \( t \); \( G_b(\cdot) \) is the trip completion flow for the bus mode given by the 3D-MFD during time interval \( t \). In Eq. 6.8, \text{term 1} calculates the total number of dispatching bus units of type \( i \) across all bus lines during time interval \( t \); \text{term 2} denotes the total accumulation of bus units of type \( i \) across all bus lines at the beginning of time interval \( t \); \text{term 3} computes the total number of bus units of type \( i \) reaching their destination/terminal station across all bus lines during time interval \( t \).

In terms of the constraints, Eq. 6.3 ensures that the dispatching flows of bus units \( \psi \) need to be non-negative. Eq. 6.4 sets bounds for the dispatching flows of buses, where thresholds \( \Psi^\text{MIN}_b \) and \( \Psi^\text{MAX}_b \) are determined by the bus operator as the minimum and maximum bus frequencies, respectively, based on the operator budget, restrictions imposed by traffic authorities, demand patterns, etc. These thresholds could further ensure that the resulting frequencies fall within the range of frequencies observed in the given 3D-MFD, such that the 3D-MFD would be invariant with respect to the dispatching frequencies. Eq. 6.5 represents that the number of dispatching buses along any bus line cannot exceed the number of dispatching bus units. Eq. 6.6 ensures that the number of combined modular units does not exceed the maximum value for a given line. Eq. 6.7 guarantees that the total number of bus units required for the operation during any given time
interval does not exceed the vehicle fleet in case of both conventional and modular bus units.

The operator cost $Z_O(\Psi, \psi)$ is calculated in Eq. 6.9 based on the number of bus units of each type operating during the planning horizon and the corresponding unit cost $\pi_i$. Here, $\pi_i$ includes the unit cost of assigning bus drivers to conventional bus units, as well as the unit cost incurred by the energy consumption and maintenance of both conventional and modular bus units.

$$Z_O(\Psi, \psi) = \sum_{t \in T} \sum_{i \in I} \sum_{l \in L} \sum_{s \in S_l} n_{b,i,l,s}(t) \cdot \pi_i \cdot \tau. \quad (6.9)$$

On the other hand, the user cost $Z_U(\Psi, \psi)$ is related to the total time traveled in the system (including the waiting time), computed across both modes and converted into an equivalent monetary cost with a parameter $\Omega$ denoting the average value of time (Eq. 6.10).

$$Z_U(\Psi, \psi) = \Omega \cdot \sum_{t \in T} \sum_{l \in L} \sum_{s \in S_l} \sum_{s' \in S_l} \left( P_{l,s,s'}(t) + \omega_{l,s,s'}(t) \right) \cdot \tau$$

$$+ \Omega \cdot \sum_{t \in T} \sum_{l \in L} \sum_{s \in S_l} \sum_{s' \in S_l} \frac{1}{2} \cdot \frac{\lambda_{l,s,s'}(t) \cdot \tau \cdot |I|}{\sum_{i \in I} \Phi_{b,i,l,s}(t)} + \Omega \cdot \sum_{t \in T} \rho_c \cdot N_c(t) \cdot \tau,$$

where $P_l(t)$ and $\omega_l(t)$ stand for the total number of on-board passengers and the total number of passengers who cannot board the bus, respectively, on line $l$ at the beginning of time interval $t$; $\lambda_l(t)$ is the average arrival rate of bus passengers on line $l$ during time interval $t$; $\rho_c$ denotes the average car occupancy ($\rho_c \approx 1$). In Eq. 6.10, term 1 calculates the time traveled by bus users and the waiting time of passengers who cannot board the bus; term 2 represents a proxy for the passenger waiting time due to the bus headway defined as half of the average bus headway across all types of bus units, $(\sum_{i \in I} \Phi_{b,i,l}(t)/|I|)^{-1}$; and term 3 denotes the time traveled by car users. This way, the user cost function captures the influence of bus operations on the travel cost of both modes.

Notice that the formulated optimization problem requires input variables from both vehicular and passenger perspective. These variables are estimated based on the 3D-MFD. The input parameters to the 3D-MFD model are the topological layout of bus lines (including the lane allocation, the number of bus stops, as well as the distance between the stops) and
both, the car and the bus passenger demand level. In the following, we
describe the proposed modeling framework based on the 3D-MFD.

We make the following assumptions for the methodological framework,
in terms of passenger preferences, the automated modular vehicle technol-
ogy, and the effects of bus operations on car traffic.

**Assumption 6.2** (Passenger preferences). We make the following assump-
tions regarding the passenger preferences: (i) the demand information of
each mode, as well as the routing information of bus passengers, is exoge-
nously given, i.e. we do not consider a dynamic change in mode choice,
route choice, or departure time as a function of the dispatching policy;
and (ii) bus passengers do not have a preference on which type of units
they choose to board.

**Remark.** We make the following remarks regarding Assumption 6.2. First,
we aim to develop short-term dispatching strategies to improve the overall
operator and passenger performance, rather than making long-term plan-
ning decisions (e.g. adding bus lines, changing the spatial distribution of
bus lanes, changing the bus network structure, etc.) that might affect the de-
mand. This is similar to other concepts related to improving bus operations
such as (multi-modal) perimeter control (32, 149–152), anti-bus bunching
methods (153–158), or other bus dispatching-related studies (79, 83, 159),
which assume that there is no instantaneous mode, route, or departure
time shift due to the implemented control strategy. We will further discuss
this assumption in Section 6.2.2. Second, we assume that bus passengers
do not have a preference on which type of units they choose to board,
which applies to a common scenario where passengers board whichever
bus is first available to them. Nevertheless, the proposed methodological
framework can be extended to account for passenger preference over types
of units, which, however, is out of the scope for this dissertation.

**Assumption 6.3** (Automated modular vehicle technology). We make the
following assumptions regarding the considered modular vehicle technol-
ogy: (i) the composition of modular bus units is determined only at the
terminal station, and that no combining/splitting of modular units occurs
along a bus line once a modular bus is dispatched; (ii) the size of a modular
bus never exceeds the size of the conventional bus, i.e. the value of $|U_{m,l}|$
is chosen such that the length of a modular bus containing the maximum
number of combined modular units corresponds to the length of the con-
tentional bus; and (iii) automated modular bus units have similar vehicle
dynamics as conventional buses.
Remark. We make the following remarks regarding Assumption 6.3. First, although the en-route combining and splitting of modular bus units is technologically feasible, this requires accurate prediction of bus travel times. However, the bus travel times can be highly stochastic on mixed lanes, and thus we assume that the composition of modular bus units is determined only at the terminal station. Moreover, such en-route combining and splitting of units is only helpful if the units were to take different routes or skip stops; and both of these strategies are out of the scope for this dissertation. Second, the main benefit of modular bus units is that we can dispatch buses with lower capacity in scenarios where demand is lower to reduce operational costs, and thus it is not beneficial to consider the operational regime where the size of modular buses is larger than that of conventional buses. Additionally, this constraint ensures that the combined modular bus units never exceed the physical length of the current bus stops. Third, given the assumption of similar vehicle dynamics of modular bus units and conventional buses, we do not consider the impact of reduced reaction times (enabled by the automated vehicle technology) on the shape of the 3D-MFD.

Assumption 6.4 (Effects of bus operations on car traffic). We assume that, along mixed lanes, buses drive at the speed of cars while cruising, i.e. both modes have the same speed either in the free-flow or congested traffic conditions (as in e.g. 10, 35, 59, 60, 84, 129). Consequently, buses do not act as moving bottleneck along mixed lanes, i.e. they affect car traffic only due to their stops (i.e. dwell time). Other effects of bus operations on car traffic, such as the effects of the vehicle size along mixed lanes and transit signal priority along dedicated bus lanes, are incorporated into the parameters of the 3D-MFD model.

Remark. We make the following remarks regarding Assumption 6.4. First, the 3D-MFD implicitly assumes that the vehicle size of buses is comparable to that of conventional buses. Note, however, that the size of a modular bus never exceeds the size of the conventional bus (as per Assumption 6.3). Hence, employing the 3D-MFD could over-estimate the negative impact of modular buses on car traffic. That being said, in this dissertation, we are taking a conservative approach that might under-estimate the benefits of the modular vehicle technology. Second, although buses and cars do not directly interact on dedicated bus lanes, bus operations can still affect car traffic through transit signal priority. These effects are, nevertheless, captured in the 3D-MFD.
6.1.3 Modeling vehicular dynamics

To model car traffic at the network level while capturing the interactions between the modes based on the 3D-MFD, we follow the approach by (7) and (137), and assume a linear model between the car speed and the accumulation of both modes:

\[ v_c(t) = a_{b}^{\text{MIX}} \cdot N_b^{\text{MIX}}(t) + a_{b}^{\text{DED}} \cdot N_b^{\text{DED}}(t) + \alpha_c \cdot N_c(t) + \beta, \quad (6.11) \]

where \( a_{b}^{\text{MIX}} \), \( a_{b}^{\text{DED}} \), and \( \alpha_c \) are parameters that capture the marginal effect of each mode on the average car speed; \( \beta \) characterizes the effects of the bus network topology; \( N_b^{\text{MIX}}(t) \) and \( N_b^{\text{DED}}(t) \) stand for the bus accumulation in the subnetwork with mixed and dedicated bus lanes, respectively, at the beginning of time interval \( t \), such that \( N_b^{\text{MIX}}(t) + N_b^{\text{DED}}(t) = N_b(t) \). Note that \( a_{b}^{\text{MIX}} \) accounts for the effects of the vehicle size of conventional buses on car speed along mixed lanes, whereas \( a_{b}^{\text{DED}} \) captures the effects of transit signal priority along dedicated segments. These parameters can be estimated with real data (see (7) for more details). In this dissertation, we assume that the parameters of the 3D-MFD model are exogenously given.

Due to varying lane allocation along a bus line, we model the bus speed at the segment level as a function of the average (across both types of units) dwell time \( \Theta_{l,s}(t) \) on a given segment, i.e.:

\[ v_{b,l,s}(t) = \begin{cases} (1/v_c(t) + \theta'/\ell_{b,l,s} + \Theta_{l,s}(t))^{-1}, & \text{if } s \in S_{l}^{\text{MIX}}, \\ (1/v_{\text{MAX}} + \theta'/\ell_{b,l,s} + \Theta_{l,s}(t))^{-1}, & \text{if } s \in S_{l}^{\text{DED}}, \end{cases} \quad (6.12) \]

where \( \theta' \) denotes the time lost per stop due to required door operations and deceleration/acceleration maneuvers; \( \ell_{b,l,s} \) is the average distance between the bus stops on segment \( s \) of line \( l \). Notice from Eq. 6.12 that, for dedicated bus segments, the bus speed is modeled independently of other vehicles, i.e. based on the free-flow speed of buses. On the other hand, given Assumption 6.4 that buses drive at the speed of cars while cruising, we model the bus speed along mixed segments based on the average car speed given by the 3D-MFD. Before we formulate the average dwell time \( \Theta_{l,s}(t) \) used in Eq. 6.12, let us first introduce some notation.

Let \( B_{l,s}(t) \) be the total number of boarding passengers and \( A_{l,s}(t) \) the total number of alighting passengers on segment \( s \) of line \( l \) during time interval \( t \). These variables are estimated using the passenger dynamics model, as described in the following subsection. Denote by \( N_{b,i,l,s}(t) \) and \( n_{b,i,l,s}(t) \) the accumulation of buses and the accumulation of bus units, respectively, of type \( i \) on segment \( s \) of line \( l \) at the beginning of time interval.
The dwell time of buses of type $i$ on segment $s$ can be computed as the ratio of the total number of boarding/alighting passengers across all buses of type $i$ on a given segment $s$ (the term inside parenthesis in Eq. 6.13) and the corresponding bus accumulation:

$$\Theta_{i,l,s}(t) = \frac{1}{N_{b,i,l,s}(t)} \cdot \left( \theta \cdot \max\{B_{i,s}(t), A_{i,s}(t)\} \cdot \frac{n_{b,i,l,s}(t) \cdot C_i}{\sum_{i \in \mathcal{I}} n_{b,i,l,s}(t) \cdot C_i} \right),$$

(6.13)

where $\theta$ denotes the time added per boarding/alighting passenger. Note that Eq. 6.13 assumes a simultaneous boarding and alighting process. Alternatively, one can use a dwell time model based on a sequential boarding and alighting process. Also note that, due to Assumption 6.2, the bus occupancy can be uniformly distributed across both types of units. Therefore, the total number of passengers boarding/alighting buses of type $i$ is computed to be proportional to the total capacity of buses of a given type (the fraction inside parenthesis in Eq. 6.13).

The average dwell time on a given segment $s$ can then be computed as an average of the dwell times of both types of buses, weighted by their respective bus accumulations (Eq. 6.14).

$$\Theta_{l,s}(t) = \frac{\sum_{i \in \mathcal{I}} N_{b,i,l,s}(t) \cdot \Theta_{i,l,s}(t)}{\sum_{i \in \mathcal{I}} N_{b,i,l,s}(t)} = \frac{\theta \cdot \max\{B_{l,s}(t), A_{l,s}(t)\}}{\sum_{i \in \mathcal{I}} N_{b,i,l,s}(t)}. \quad (6.14)$$

We can also relate the total bus accumulation in each subnetwork (with mixed and dedicated bus segments) to the bus accumulation across all segments of the same type and across both types of buses (Eq. 6.15).

$$N_{b}^{\text{MIX}}(t) = \sum_{i \in \mathcal{I}} \sum_{l \in \mathcal{L}} \sum_{s \in \mathcal{S}^{\text{MIX}}} N_{b,i,l,s}(t), \quad (6.15a)$$

$$N_{b}^{\text{DED}}(t) = \sum_{i \in \mathcal{I}} \sum_{l \in \mathcal{L}} \sum_{s \in \mathcal{S}^{\text{DED}}} N_{b,i,l,s}(t). \quad (6.15b)$$

Now that we have introduced all the notation and defined all modeling variables, we can formulate the mass conservation equations used to model the evolution of the vehicle accumulation for each mode. Given the assumption of the homogeneous traffic conditions across the network, which assumes that the average car speed in mixed lanes is similar to the average car speed in car only lanes (as in e.g. 10, 35, 84), the system dynamics for the car mode at the network level can be modeled as:

$$N_{c}(t + 1) = N_{c}(t) + \lambda_{c}^{\text{INT}}(t) \cdot \tau + \lambda_{c}^{\text{EXT}}(t) \cdot \tau - G_{c}(N_{b}(t), N_{c}(t)) \cdot \tau, \quad (6.16)$$
where \( \lambda_c^{\text{INT}}(t) \) and \( \lambda_c^{\text{EXT}}(t) \) stand for the internal (originating from within the network) and external (incoming to the network) car demand, respectively, during time interval \( t \); and \( G_c(\cdot) \) is the trip completion flow for the car mode given by the 3D-MFD during time interval \( t \). The trip completion flow for the bus and the car mode includes both the internal outflow (reaching destination inside the network) and the external outflow (exiting the network). (4) showed that \( G_b(\cdot) \) and \( G_c(\cdot) \) are proportional to the total circulating flows, i.e. \( Q_b(\cdot) = v_b(t) \cdot N_b(t) / L_b \) and \( Q_c(\cdot) = v_c(t) \cdot N_c(t) / L_c \), respectively, with a factor that represents the ratio of the network length for a given mode (\( L_c \) for cars; \( L_b \) for buses) and the average vehicular trip length (\( \bar{\ell}_c \) for cars; \( \bar{\ell}_b \) for buses):

\[
G_b(N_b(t), N_c(t)) = Q_b(N_b(t), N_c(t)) \cdot L_b / \bar{\ell}_b = v_b(t) \cdot N_b(t) / \bar{\ell}_b, \quad (6.17a)
\]

\[
G_c(N_b(t), N_c(t)) = Q_c(N_b(t), N_c(t)) \cdot L_c / \bar{\ell}_c = v_c(t) \cdot N_c(t) / \bar{\ell}_c, \quad (6.17b)
\]

where \( v_b(t) \) is the average bus speed in the network (Eq. 6.18). It is worth mentioning that the average vehicular trip length for the bus mode is defined as the average length of a bus line in one direction of travel.

\[
v_b(t) = \frac{\sum_{i \in \mathcal{I}} \sum_{l \in \mathcal{L}} \sum_{s \in \mathcal{S}_i} N_{b,i,l,s}(t) \cdot v_{b,i,l,s}(t)}{\sum_{i \in \mathcal{I}} \sum_{l \in \mathcal{L}} \sum_{s \in \mathcal{S}_i} N_{b,i,l,s}(t)}. \quad (6.18)
\]

On the other hand, due to different operational regimes of buses across different lane allocations along a bus line, as stated before, the system dynamics for the bus mode is modeled at the segment level:

\[
N_{b,i,l,s}(t+1) = \begin{cases} 
N_{b,i,l,s}(t) + \Psi_{b,i,l}(t) \cdot \tau - \Phi_{b,i,l,s}(t) \cdot \tau, & \text{if } s = 1, \\
N_{b,i,l,s}(t) + \Phi_{b,i,l,s-1}(t) \cdot \tau - \Phi_{b,i,l,s}(t) \cdot \tau, & \text{if } 1 < s < |\mathcal{S}_i|, \\
N_{b,i,l,s}(t) + \Phi_{b,i,l,s-1}(t) \cdot \tau - N_{b,i,l,s}^{\text{OUT}}(t), & \text{if } s = |\mathcal{S}_i|, 
\end{cases} \quad (6.19a)
\]

\[
n_{b,i,l,s}(t+1) = \begin{cases} 
n_{b,i,l,s}(t) + \psi_{b,i,l}(t) \cdot \tau - \phi_{b,i,l,s}(t) \cdot \tau, & \text{if } s = 1, \\
n_{b,i,l,s}(t) + \phi_{b,i,l,s-1}(t) \cdot \tau - \phi_{b,i,l,s}(t) \cdot \tau, & \text{if } 1 < s < |\mathcal{S}_i|, \\
n_{b,i,l,s}(t) + \phi_{b,i,l,s-1}(t) \cdot \tau - n_{b,i,l,s}^{\text{OUT}}(t), & \text{if } s = |\mathcal{S}_i|, 
\end{cases} \quad (6.19b)
\]

with

\[
N_{b,i,l,s}^{\text{OUT}}(t) = (N_{b,i,l,s}(t) / N_b(t)) \cdot G_b(N_b(t), N_c(t)) \cdot \tau, \quad (6.20a)
\]
\[ n_{b,i,l,s}^{\text{out}}(t) = \left( n_{b,i,l,s}(t)/N_b(t) \right) \cdot G_b(N_b(t), N_c(t)) \cdot \tau, \]  

(6.20b)

where \( N_{b,i,l,s}^{\text{out}}(t) \) and \( n_{b,i,l,s}^{\text{out}}(t) \) denote the total number of buses and the total number of bus units, respectively, which have completed their trip during time interval \( t \); \( \Phi_{b,i,l,s}(t) \) is the transferring flow of buses and \( \phi_{b,i,l,s}(t) \) is the transferring flow of bus units of type \( i \) between segments \( s \) and \( s+1 \) of line \( l \) during time interval \( t \). These transferring flows are computed using a CTM-inspired approach, as the minimum of the sending flow from segment \( s \) of line \( l \) (the first term in Eq. 6.21) and the receiving flow to segment \( s+1 \) of the same line during a given interval \( t \) (the second term in Eq. 6.21). Note that, for simplicity of presentation, Eq. 6.21 assumes that the bus lines do not overlap as otherwise the sending and receiving flows of buses/bus units should be proportionally distributed on a given overlapping segment based on the number of buses/bus units along the corresponding bus line.

\[
\Phi_{b,i,l,s}(t) = \min \left\{ v_{b,l,s}(t) \cdot \frac{N_{b,i,l,s}(t)}{\ell_{b,l,s}, \frac{\varphi_{c,l,s+1}(t)}{\eta_i \cdot u_{i,l,s}(t)}} \right\}, \quad (6.21a)
\]

\[
\phi_{b,i,l,s}(t) = \min \left\{ v_{b,l,s}(t) \cdot \frac{n_{b,i,l,s}(t)}{\ell_{b,l,s}, \frac{\varphi_{c,l,s+1}(t)}{\eta_i}} \right\}, \quad (6.21b)
\]

where \( \varphi_{c,l,s}(t) \) denotes the receiving flow of cars to segment \( s \) of line \( l \) during time interval \( t \); \( \eta_i \) represents a parameter that quantifies by how much one bus unit of type \( i \) reduces the car flow (i.e. a bus unit-car equivalent); \( u_{m,l,s}(t) = n_{b,m,l,s}(t)/N_{b,m,l,s}(t) \) is the average number of modular bus units contained in a modular bus on segment \( s \) of line \( l \) at the beginning of time interval \( t \). Recall that for conventional bus units \( u_{r,l,s}(t) = 1 \), i.e. \( n_{b,r,l,s}(t) = N_{b,r,l,s}(t) \). Before we formulate \( \varphi_{c,l,s}(t) \), let us first introduce some notation.

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

$$N_{c}^{\text{MIX}}(t) = N_c(t) \cdot \frac{\overline{N}_c^{\text{MIX}}(t)}{N_c} = N_c(t) \cdot \frac{\sum_{l \in L} \sum_{s \in S_{l}^{\text{MIX}}} N_{c,l,s}(t)}{N_c}, \quad (6.22)$$

where $N_{c,l,s}(t)$ stands for the maximum car accumulation that can be accommodated on a given segment $s$ of line $l$ at the beginning of time interval $t$, defined as a function of the bus accumulation on that same segment (Eq. 6.23).

$$\overline{N}_{c,l,s}(t) = \overline{N}_c \cdot \frac{\ell_{b,l,s}}{L_c} - \sum_{i \in I} n_{b,i,l,s}(t) \cdot \eta_i. \quad (6.23)$$

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

$$N_{c,l,s}(t) = N_{c}^{\text{MIX}}(t) \cdot \frac{\overline{N}_{c,l,s}(t)}{\sum_{l \in L} \sum_{s \in S_{l}^{\text{MIX}}} \overline{N}_{c,l,s}(t)} = N_c(t) \cdot \frac{\overline{N}_{c,l,s}(t)}{N_c}. \quad (6.24)$$

Similarly to the CTM, we can now compute the receiving flow of cars based on the maximum available car accumulation, $\overline{N}_{c,l,s}(t)$, and (only in case of a mixed segment) the number of cars present on a given segment at the beginning of time interval $t$, as given by Eq. 6.25, where $w$ is the backward wave speed.

$$\phi_{c,l,s}(t) = \begin{cases} 
    w \cdot (\overline{N}_{c,l,s}(t) - N_{c,l,s}(t)) / \ell_{b,l,s}, & \text{if } s \in S_{l}^{\text{MIX}}, \\
    w \cdot \overline{N}_{c,l,s}(t) / \ell_{b,l,s}, & \text{if } s \in S_{l}^{\text{DED}}.
\end{cases} \quad (6.25)$$

### 6.1.4 Modeling passenger dynamics

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

$$P_{l,s,s'}(t+1) = \begin{cases} 
    P_{l,s,s'}(t) + D_{l,s,s'}(t) - O_{l,s,s'}(t), & \text{if } s = 1, \\
    P_{l,s,s'}(t) + D_{l,s,s'}(t) + O_{l,s-1,s'}(t) - O_{l,s,s'}(t), & \text{if } s > 1, 
\end{cases} \quad (6.26)$$
where $P_{l,s,s'}(t)$ is the total number of on-board passengers on segment $s$ with destination on segment $s'$ of line $l$ at the beginning of time interval $t$; $D_{l,s,s'}(t)$ represents the total number of boarding passengers on segment $s$ with destination on segment $s'$ of line $l$ during time interval $t$; and $O_{l,s,s'}(t)$ denotes the outflow (including the transferring and the trip completion flow, as elaborated further below) of bus passengers between segments $s$ and $s+1$ of line $l$ during time interval $t$, whose destination is on segment $s'$ of the same line (Eq. 6.27).

$$O_{l,s,s'}(t) = \begin{cases} 
\min\{P_{l,s,s'}(t), \sum_{i \in \mathcal{I}} \rho_{b,i,l,s,s'}(t) \cdot \phi_{b,i,l,s}(t) \cdot \tau\}, & \text{if } s \neq s' \land s < |S_l| \\
\min\{P_{l,s,s'}(t), R_{l,s,s'}(N_b(t), N_c(t)) \cdot \tau\}, & \text{if } s = s' \lor s = |S_l| 
\end{cases}$$

(6.27)

where $\rho_{b,i,l,s,s'}(t)$ and $R_{l,s,s'}(\cdot)$ stand for the occupancy of a bus unit of type $i$ and the trip completion flow, respectively, of bus passengers on segment $s$ of line $l$, whose destination is on segment $s'$ of the same line. Given Assumption 6.2 and the implied uniformly distributed bus occupancy across both types of units, we can obtain $\rho_{b,i,l,s,s'}(t)$ by multiplying the number of on-board passengers per unit capacity (the term inside parenthesis in Eq. 6.28) by the corresponding capacity of a given type of unit.

$$\rho_{b,i,l,s,s'}(t) = C_i \cdot \left( \frac{P_{l,s,s'}(t)}{\sum_{i \in \mathcal{I}} n_{b,i,l,s}(t) \cdot C_i} \right).$$

(6.28)

On the other hand, the trip completion flow $R_{l,s,s'}(\cdot)$ is computed by incorporating the average (across both types of units) bus occupancy of passengers on segment $s$ with destination is on segment $s'$ of line $l$, $\rho_{b,l,s,s'}(t)$, into Eq. 6.17b, and substituting the average vehicular distance for the bus mode with the average trip length of bus passengers, $\bar{\ell}_b'$, i.e.:

$$R_{l,s,s'}(N_b(t), N_c(t)) = \rho_{b,l,s,s'}(t) \cdot Q_b(N_b(t), N_c(t)) \cdot L_b / \bar{\ell}_b' = \rho_{b,l,s,s'}(t) \cdot \bar{v}_b(t) \cdot N_b(t) / \bar{\ell}_b'$$

(6.29)

where $\rho_{b,l,s,s'}(t)$ is given as an average of the occupancies of both types of buses (obtained by multiplying the average number of units contained in a bus of each type by the corresponding occupancy of a bus unit), weighted by their respective bus accumulations (Eq. 6.30).

$$\rho_{b,l,s,s'}(t) = \frac{\sum_{i \in \mathcal{I}} N_{b,i,l,s}(t) \cdot u_{i,l,s}(t) \cdot \rho_{b,i,l,s,s'}(t)}{\sum_{i \in \mathcal{I}} N_{b,i,l,s}(t)} = \frac{P_{l,s,s'}(t)}{\sum_{i \in \mathcal{I}} N_{b,i,l,s}(t)}.$$
Note that in case where passengers do not arrive at their destination (i.e. \( s \neq s' \) and \( s < |S| \)), the outflow \( O_{\ell,s,s'}(t) \) represents the transferring flow from segment \( s \) to segment \( s + 1 \) of a given line \( l \) during time interval \( t \). Otherwise (i.e. \( s = s' \) or \( s = |S| \)), the outflow represents the trip completion flow of bus passengers given by the 3D-MFD.

The total number of boarding and alighting passengers on segment \( s \) of line \( l \) during time interval \( t \) can now be determined as:

\[
B_{l,s}(t) = \sum_{s' \in S_l} D_{l,s,s'}(t), \quad (6.31)
\]
\[
A_{l,s}(t) = \sum_{s' \in S_l} 1\{s=s' \lor s=|S|\} \cdot O_{l,s,s'}(t), \quad (6.32)
\]

where \( 1\{\text{condition}\} \) is an indicator function that return the value of 1 if condition is satisfied. The total number of boarding passengers \( D_{l,s,s'}(t) \) is bounded by the two parameters in Eq. 6.33: the total number of (accumulated) passengers \( D_{l,s,s'}^{\text{ACC}}(t) \) that want to enter the bus, and the total number of passengers \( D_{l,s,s'}^{\text{MAX}}(t) \) that can enter the bus:

\[
D_{l,s,s'}(t) = \min\{D_{l,s,s'}^{\text{ACC}}(t), D_{l,s,s'}^{\text{MAX}}(t)\}. \quad (6.33)
\]

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

\[
\omega_{l,s,s'}(t + 1) = \omega_{l,s,s'}(t) + \lambda_{l,s,s'}(t) \cdot \tau - D_{l,s,s'}(t), \quad (6.34)
\]

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

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

\[
D_{l,s,s'}^{\text{ACC}}(t) = \omega_{l,s,s'}(t) + \lambda_{l,s,s'}(t) \cdot \tau. \quad (6.35)
\]

On the other hand, the total number of passengers with destination on segment \( s' \) of line \( l \) that can enter the bus on segment \( s \) of the same line is determined by the total available capacity across both types of bus units operating along line \( l \) during a given time interval (Eq. 6.36).
\[ D_{l,s,s'}^{\text{MAX}}(t) = \delta_{l,s,s'}(t) \cdot \left( \sum_{i \in I} n_{b,i,l,s}(t + 1) \cdot C_i \right) - \sum_{s' \in S_l} \left( P_{l,s,s'}(t) + 1_{\{s > 1\}} \cdot O_{l,s-1,s'}(t) - O_{l,s,s'}(t) \right), \] (6.36)

where \( n_{b,i,l,s}(t + 1) \) is calculated using Eq. 6.19, and \( \delta_{l,s,s'}(t) \) represents the share of passengers on segment \( s \) of line \( l \) during time interval \( t \) whose destination is on segment \( s' \) of the same line (Eq. 6.37).

\[
\delta_{l,s,s'}(t) = \frac{\lambda_{l,s,s'}(t) \cdot \tau + \omega_{l,s,s'}(t)}{\sum_{s' \in S_l} (\lambda_{l,s,s'}(t) \cdot \tau + \omega_{l,s,s'}(t))}.
\] (6.37)

### 6.2 Numerical Experiments and Results

#### 6.2.1 Case study and simulation scenarios

Here we describe the simulation environment for testing the performance of the proposed flexible bus dispatching system. The considered traffic network is inspired by the City of Zurich, Switzerland, and is comprised of five bus lines with varying lane allocation. Dedicated bus segments are placed closer to the center of the network, whereas the mixed lane segments are installed closer to the periphery (Fig. 6.3).

![Figure 6.3: Schematic illustration of the studied network, where solid lines represent dedicated bus segments, whereas dashed lines represent mixed bus segments.](image-url)
The simulated traffic conditions reflect a typical morning-peak period. The 3D-MFD and the demand profiles (both in terms of cars and passengers) during the 3 hour period (Fig. 6.4) are designed to mimic the aggregated traffic features of the city center of Zurich (27), as this one exhibits a well-defined empirical 3D-MFD, as shown in Fig. 6.2b (see 7, for more details). To obtain realistic traffic conditions, we impose lower and upper bounds to the conservation equations (both in terms of vehicles and passengers).

![Figure 6.4](image)

**Figure 6.4:** Simulation settings: (a) car demand profile; and (b) public transport passenger demand profile.

Tested traffic scenarios include three types of car demand (low, medium, and high) and three types of public transport passenger demand (low, medium, and high) (see Fig. 6.4). For each scenario, we vary the penetration rate of modular bus units: $p_m \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 1\}$. The penetration rate of $p_m = 0$ corresponds to the base case (i.e. the case corresponding to existing bus systems with only conventional buses), in which no modular units are incorporated into the bus fleet. In such a base case, the considered bus fleet includes $F_r = 50$ conventional bus units. In other cases with $p_m > 0$, the number of conventional buses depends on the penetration rate of modular bus units, i.e. $F_r = 50 \cdot (1 - p_m)$. We assume that the bus capacity is $C_r = 120$ pax/veh for conventional buses and $C_m = 20$ pax/veh for modular bus units, which gives $\zeta = 6$. We also use the value of $\zeta$ as the maximum number of modular bus units that can be combined, and this value is kept the same for all bus lines, i.e. $|U_{m,l}| = \zeta, \forall l \in L$. The values of $C_m$ and $\zeta$ are determined based on the capacity of a modular bus unit (as in e.g. the Next Future Mobility system), such that the length of a modular bus containing the maximum number of combined modular
units corresponds to the length of an articulated bus. The passenger cost per unit time is assumed to be 20 CHF/hr, comparable to the value of time in the city of Zurich (160). Finally, following recommendations by (161), we approximate the unit cost $\pi_i$ for the bus operator by accounting for the driver cost (only in case of conventional buses), vehicle cost, operational cost (including the energy consumption and maintenance), overhead (e.g. administration and planning cost), and ticketing cost. Using the values of the aforementioned elements of the operating unit cost for Switzerland, we obtain $\pi_r \approx 260$ CHF/hr and $\pi_m \approx 30$ CHF/hr.

The proposed optimization problem is solved using a conventional sequential quadratic programming algorithm (162), which allows to formulate the constraints explicitly, without using the penalty terms in the objective function. To find the optimal solution, we set the number of initial points to 50. Sensitivity analyses show that increasing the number of initial points beyond 50 leads to marginal improvements in the objective function (we omit the results here for brevity) for all scenarios with different demand levels and penetration rates of modular units. In the future, it might be possible to use more in depth sensitivity analysis methods (145, 163) to further improve the efficiency of the solution algorithm. Experiments are run on a 16-core Intel Xeon processor (3.19 GHz) with 256 GB RAM. The computation time for a given scenario and penetration rate of modular units is around 8.5 min for a 3-hour simulation.

6.2.2 Value of considering modular vehicle technology

In this subsection, we demonstrate the value of considering the modular vehicle technology by comparing the total system cost (including the operator cost and the system delay) of the proposed flexible bus dispatching system with different penetration rates of modular units to that of the base case. The absolute (in CHF/hr) and relative (in %) improvements are shown in the form of bar plots (Fig. 6.5). Recall that the base case refers to the existing bus dispatching system consisting only of conventional buses, i.e. with $p_m = 0$. Note that even if the performance is the same, the cost for the base case described here is still not equivalent to the conventional bus cost. This is because we are only dealing with the frequency setting and not vehicle scheduling. The underlying vehicle scheduling problem (deadheading/assigning vehicles to trips or to reposition to merge with other vehicles) for modular bus units has a substantially different cost structure than that for conventional bus (likely much lower cost). Therefore, there
are additional cost savings that are not covered here from that perspective. In Fig. 6.5, traffic scenarios are represented as XY, where X indicates the level of car demand and Y indicates the level of passenger demand for public transport: low (L), medium (M), or high (H). For example, LM stands for the scenario with low car and medium public transport passenger demand.

As demonstrated by Fig. 6.5, the proposed system substantially outperforms the base case with only conventional buses, especially for scenarios with lower car and public transport demand. Potential improvements can be up to 19%, depending on the level of car and passenger demand, as well as the penetration rate of modular bus units. As the level of demand (in particular for the car mode) increases, the improvement is smaller, but still significant. The reason for this is two-fold. First, in case of low bus and car demand, the system sends a group of few combined (if not individual) modular units, optimizing thereby the utilization of the vehicle’s capacity and reducing the operating cost. Second, due to a lower number of units being contained in a modular bus, the system has more units available in stock, allowing to dispatch more modular buses and reduce the passenger waiting time. This, on the other hand, has minimum impact on car traffic in case of low car demand, due to lower interactions between the modes. That being said, dispatching more modular buses for higher car demand levels does not necessarily have to improve the system, as the impact on car traffic may be significant, resulting in lower overall system performance. This is illustrated in Fig. 6.6, where we observe that the average (across all bus lines) number of combined modular units increases as the car demand increases, especially during the peak period (Fig. 6.6b) and

Figure 6.5: Comparison of the improvement in the total system cost for different penetration rates of modular units.
for higher penetration rates of modular units. Consequently, the improvements in the total system delay do not seem to be significant for congested traffic scenarios (Fig. 6.7), as the system has less flexibility in managing the allocation of the vehicle resources compared to other scenarios. In other words, in such congested cases, the proposed system tends to behave similarly to the conventional bus dispatching system, especially since we define that the capacity of the modular bus containing the maximum number of modular units is the same as that of the conventional bus. Nevertheless, the operator cost function still gets significantly improved (Fig. 6.8). Potential improvements in the operator cost function can be up to 58% for simulated scenarios. This shows that the proposed system provides more flexibility for dispatching buses to serve the passenger demand while reducing the operator cost compared to the conventional bus dispatching system, even for congested traffic conditions.

Figure 6.6: Evolution of the average number of units contained in a modular bus for different penetration rates of modular units: (a) before the peak; (b) during the peak; and (c) after the peak.
A closer look into the spatial-temporal distribution of the accumulation of buses and bus units (Fig. 6.9) reveals that the average number of combined modular units varies not only with time, but also across the bus lines, indicating the importance of accounting for the passenger dynamics at the segment level. For brevity, in Fig. 6.9, we only show three bus lines with different levels of bus passenger demand, for one particular penetration rate of modular bus units (10%) and traffic scenario (MM). Lines 2 and 4 exhibit very similar patterns to those of lines 3 and 5, respectively. As expected, with an increase in the bus passenger demand (see e.g. line 3 compared to line 1 in Fig. 6.9), the proposed system tends to dispatch more modular buses with lower headway, especially during the peak period. This is achieved by dispatching modular buses with lower capacity,
i.e. lower number of combined modular units (see the ratio between the dashed line and dark gray area in Fig. 6.9). In case of low penetration rate of modular units (such as that in Fig. 6.9), an increase in the bus passenger demand along a line is also followed by a higher share of regular buses within the total number of buses (see e.g. the light gray area for line 5 compared to that of line 3 in Fig 6.9).

![Spatial-temporal distribution of the accumulation of modular buses, conventional buses, and modular bus units for different bus lines.](image)

**Figure 6.9:** Spatial-temporal distribution of the accumulation of modular buses, conventional buses, and modular bus units for different bus lines.

It is also worth mentioning that the minimum tested penetration rate of modular units (10%) exhibits a substantial reduction in the total cost. This suggests that the total system cost can be significantly reduced even when only a small fraction of vehicles within the bus fleet are modular, indicating the value of the modular vehicle technology in the early deployment stages. As the penetration rate of modular bus units increases, the marginal benefits of the modular vehicle technology decrease. Interestingly, there are only negligible improvements when the penetration rate is increased from 50% to 100%. This holds true for all tested scenarios (see Fig. 6.5). Such finding indicates that operators do not have to replace the whole fleet to benefit from the modular technology, but even a few units can make a difference.

Finally, one potential question regarding the previous analysis is whether the same conclusions could have been made if the mode choice is taken into account (i.e. without making Assumption 6.2). To address this question, we establish a feedback loop between the passenger demand, the optimal dispatching policy, and the mode choice. Such a feedback loop allows to assess the preferred mode of transport as a function of the prevailing traffic conditions, i.e. based on the generalized travel costs for both modes. These generalized travel costs are defined as the average time traveled per user (\(TT_b\) for the bus mode; \(TT_c\) for the car mode), determined by
diving the total (across all users) time traveled for a given mode by the corresponding total number of completed trips (given by the trip completion flow), i.e.:

\[
TT_b = \frac{\sum_{t \in T} \sum_{l \in L} \sum_{s \in S_l} \sum_{s' \in S_l} \left( P_{l,s,s'}(t) + \omega_{l,s,s'}(t) + \frac{\lambda_{l,s,s'}(t) \cdot |I|}{2 \cdot \sum_{i \in I} \Phi_{b,l,s,i}(t)} \right) \cdot \tau}{\sum_{t \in T} \sum_{l \in L} \sum_{s \in S_l} \sum_{s' \in S_l} \mathbb{1}_{\{s=s' \lor s=|S_l|\}} \cdot O_{l,s,s'}(t)},
\]

\[
TT_c = \frac{\sum_{t \in T} N_c(t) \cdot \tau}{\sum_{t \in T} G_c(N_b(t), N_c(t)) \cdot \tau}.
\]

Using the Logit model, we solve the fixed point problem and find the equilibrium mode choice for each penetration rate of modular bus units and traffic scenario. The results of this analysis indicate that there are negligible differences in the equilibrium mode choice values across different penetration rates of modular bus units. This validates our assumption that the short-term dispatching strategies have no influence on the passenger demand, i.e. mode choice (Assumption 6.2). Moreover, it also implies that the conclusions made from the initial experiments (without considering the mode choice) are valid even when the mode choice is taken into account. We omit the results here for brevity.

6.2.3 Value of using the 3D-MFD to capture complex system dynamics

In this subsection, we quantify the value of integrating the 3D-MFD into the modeling framework. Recall that, to realistically model the proposed system, we employed the 3D-MFD to take into account factors such as the complex multimodal interactions and congestion propagation. These factors, however, are ignored in most scientific literature on the frequency setting problem, which typically assumes that the travel times (or equivalently, the bus speeds) are independent of the bus dispatching policy.

The value of employing the 3D-MFD is quantified by comparing the results of the proposed approach to those obtained when the optimal solution for the simplified problem (i.e. considering the bus system only) is incorporated into the proposed modeling framework. In other words, in such a simplified problem, we find the optimal bus dispatching policy \( (\Psi^\ast, \phi^\ast) \) by considering neither the interactions with car traffic nor congestion propagation based on the 3D-MFD. This optimal bus dispatching policy is then used as an input into the proposed modeling framework to compute the total cost \( Z^\ast \). The value of considering multimodal interactions and congestion propagation using the 3D-MFD, \( \Delta Z \), is computed
as the relative difference between the optimal total cost determined using the proposed optimization framework (i.e. from Section 6.2.2), $Z$, and the value of $Z^*$, i.e. $\Delta Z = [Z - Z^*]/Z^*$. The results of this comparison are shown in Fig. 6.10.

![Figure 6.10](image)

**Figure 6.10:** Comparison of the improvement in the system cost made by accounting for the complex multimodal interactions and congestion propagation for different penetration rates of modular units.

We observe that substantial improvements (up to 99%) in the total system cost can be achieved with a proper modeling of the complex multimodal interactions and congestion propagation. As expected, the lowest improvements are obtained for the scenarios with uncongested traffic conditions (i.e. low level of car demand) and zero penetration rate of modular units. The reason for this is two-fold. First, in low car demand scenarios, the interactions between the modes are minimized. As the level of interactions increases, the improvements increase. Second, for zero penetration rate of modular units, the system has fewer number of vehicles available to dispatch, reducing thereby the bus accumulation in the network and the associated negative effects on car traffic compared to the scenarios with non-zero penetration rates. This is illustrated in Fig. 6.11, which shows the evolution of both the bus and the car accumulation for the tested traffic scenarios when the optimal bus dispatching policy from the simplified problem ($\Psi^*$ and $\psi^*$) is incorporated into the proposed modeling framework. Note that, for brevity, we only show the results for the scenarios with the penetration rates of $p_m = 0$ and $p_m = 0.1$, as they exhibit the lowest improvements in Fig. 6.10.
Notice from Fig. 6.11 that, without a proper modeling of multimodal interactions and congestion propagation based on the 3D-MFD, the vehicular accumulation of both modes significantly increases once the modular units are introduced to the system, especially for the congested traffic scenarios. This is because the system has more flexibility in adjusting the number of combined modular bus units to the passenger demand, resulting in higher dispatching flow of buses compared to the scenarios with zero penetration rate. Consequently, the total circulating flow in the network is substantially reduced or, in some cases, even equivalent to zero (see e.g. very congested traffic scenarios HM and HH in Fig. 6.11). In fact, the scenarios in which the total circulating flow reaches zero value are the cases when the improvements tend to 99%. This happens not only for HM and HL scenarios with the penetration rate of $p_m = 0.1$ in Fig. 6.11, but also for all tested scenarios with the penetration rate of modular units that is $p_m > 0.1$, as shown in Fig. 6.10. These problems, however, do not appear in the proposed optimization framework, indicating its value in capturing...
the necessary factors for optimizing the performance of the whole network, while taking into account all transport modes.

6.2.4 Effects of the operating unit cost of modular bus units

In the previous subsections, we used the value of $\pi_m \approx 30$ CHF/hr as the operating unit cost of modular bus units to quantify the performance of the proposed flexible bus dispatching system. This value was determined following the recommendations by (161). Given such value of $\pi_m$, the operating cost of the modular bus consisting of the maximum number of modular units that can be combined along a bus line (in this dissertation set to $|U_{m,l}| = \zeta = 6, \forall l \in L$) is lower than that of the conventional (articulated) bus. This is realistic, considering that the modular bus units are fully automated, i.e. there is no cost for assigning bus drivers to them, which in some cities represents the dominant element in the total operating unit cost of conventional buses (161).

Nevertheless, one can pose the following question: What happens to the performance of the proposed dispatching system if we increase the operating unit cost of modular bus units? This question can be meaningful for two reasons: (i) it is possible that the cost of modular units is high before the technology gets fully mature; and (ii) the analysis of this question can shed light on addressing the scenarios where safety drivers are required for these units in the early stage of deployment. To address this question, in this subsection we conduct additional experiments for the following two cases: (i) $\pi_m = 60$ CHF/hr; and (ii) $\pi_m = 120$ CHF/hr. Note that in the first/second case, four/two combined modular bus units yields the same operating cost as that of the conventional bus. Consequently, the operating cost of the modular bus consisting of the maximum number of modular units is, in both cases, significantly higher than that of the conventional bus. Although these two cases are rather unrealistic, they help to shed light on the robustness of the proposed optimization framework.

The results are shown in Table 6.2, comparing the performance of the proposed bus dispatching system with the penetration rate of $p_m = 0.1$ and the operating unit costs $\pi_m$ of 30 CHF/hr, 60 CHF/hr, and 120 CHF/hr. Note that for the last two cases the system provides a very similar bus dispatching policy for all penetration rates under a given operating unit cost of modular bus units (either 60 CHF/hr or 120 CHF/hr). In other words, the results for the two tested cases with higher $\pi_m$ are almost insensitive to the penetration rate of modular bus units. This, in turn, implies that the
marginal benefits of additional modular units are negligible. Nonetheless, notice that, even when the operating unit cost of modular bus units is significantly increased, the proposed system still substantially outperforms the existing bus dispatching system with only conventional buses. As expected, with an increase in the operating unit cost of modular bus units, the improvements (in particular those from the operator perspective) reduce. Finally, in terms of the average number of combined modular bus units, the analysis reveals that the system recognizes higher costs for combining more modular units, especially in case of $\pi_m = 120$ CHF/hr. Consequently, the capacity of modular buses is reduced compared to the case of $\pi_m = 30$ CHF/hr with $p_m = 0.1$. That being said, in case of $\pi_m = 60$ CHF/hr, the system dispatches up to two combined modular bus units, whereas in case of $\pi_m = 120$ CHF/hr the system operates mostly with individual modular units. We omit those results here for brevity.

<table>
<thead>
<tr>
<th>$\pi_m$</th>
<th>30</th>
<th>60</th>
<th>120</th>
<th>30</th>
<th>60</th>
<th>120</th>
<th>30</th>
<th>60</th>
<th>120</th>
</tr>
</thead>
<tbody>
<tr>
<td>LL</td>
<td>-16.0%</td>
<td>-12.3%</td>
<td>-7.8%</td>
<td>-44.5%</td>
<td>-34.1%</td>
<td>-18.8%</td>
<td>-5.0%</td>
<td>-5.0%</td>
<td>-4.2%</td>
</tr>
<tr>
<td>LM</td>
<td>-13.3%</td>
<td>-10.9%</td>
<td>-6.9%</td>
<td>-41.0%</td>
<td>-30.9%</td>
<td>-16.8%</td>
<td>-4.2%</td>
<td>-4.3%</td>
<td>-3.6%</td>
</tr>
<tr>
<td>LH</td>
<td>-11.7%</td>
<td>-9.5%</td>
<td>-5.9%</td>
<td>-36.9%</td>
<td>-27.6%</td>
<td>-14.7%</td>
<td>-3.6%</td>
<td>-3.7%</td>
<td>-3.1%</td>
</tr>
<tr>
<td>ML</td>
<td>-13.2%</td>
<td>-10.7%</td>
<td>-6.9%</td>
<td>-42.9%</td>
<td>-32.5%</td>
<td>-18.1%</td>
<td>-4.0%</td>
<td>-4.0%</td>
<td>-3.4%</td>
</tr>
<tr>
<td>MM</td>
<td>-11.6%</td>
<td>-9.3%</td>
<td>-5.9%</td>
<td>-39.0%</td>
<td>-29.2%</td>
<td>-16.1%</td>
<td>-3.3%</td>
<td>-3.3%</td>
<td>-2.8%</td>
</tr>
<tr>
<td>MH</td>
<td>-9.9%</td>
<td>-8.0%</td>
<td>-5.0%</td>
<td>-34.2%</td>
<td>-25.8%</td>
<td>-13.9%</td>
<td>-2.7%</td>
<td>-2.8%</td>
<td>-2.3%</td>
</tr>
<tr>
<td>HL</td>
<td>-11.0%</td>
<td>-9.0%</td>
<td>-5.8%</td>
<td>-40.0%</td>
<td>-30.4%</td>
<td>-17.3%</td>
<td>-2.9%</td>
<td>-3.0%</td>
<td>-2.5%</td>
</tr>
<tr>
<td>HM</td>
<td>-9.5%</td>
<td>-7.7%</td>
<td>-4.8%</td>
<td>-35.6%</td>
<td>-26.7%</td>
<td>-14.8%</td>
<td>-2.3%</td>
<td>-2.4%</td>
<td>-2.1%</td>
</tr>
<tr>
<td>HH</td>
<td>-7.9%</td>
<td>-6.4%</td>
<td>-3.9%</td>
<td>-31.3%</td>
<td>-23.2%</td>
<td>-12.5%</td>
<td>-1.6%</td>
<td>-1.8%</td>
<td>-1.6%</td>
</tr>
</tbody>
</table>

Table 6.2: Effects of the operating unit cost of modular bus units on the cost functions ($p_m = 0.1$).

6.2.5 Effects of the size of the bus network and the bus fleet

In the previous analyses, we used a rather larger bus network (in terms of the number of bus lines) inspired by that of the City of Zurich, Switzerland, as it exhibits a well-defined empirical 3D-MFD. Although a higher number of bus lines provides more flexibility in adjusting the number of combined modular units and their dispatching frequencies to the passenger demand, it also adds more constraints to the formulated optimization
problem. Therefore, another interesting question that can be posed is: How well does the proposed system perform when implemented on a smaller bus network?

To address this question, in this subsection we apply the proposed optimization framework on the bus network operating with two bus lines. For the purpose of computing the operator cost, we use the initial operating unit cost of modular bus units of $\pi_m = 30$ CHF/hr. To analyze how sensitive the results are to the size of the bus fleet, we analyze the following two cases indicating the number of conventional buses within the bus fleet for zero penetration rate of modular units: (i) $F = 8$; and (ii) $F = 16$. For brevity, in both cases, we only investigate the penetration rates of modular unit of $p_m \in \{0.1, 0.2, 0.3, 0.4\}$. Recall that for the scenarios with $p_m > 0$, the number of conventional buses is given as $F_r = F \cdot (1 - p_m)$.

The results are shown in Table 6.3, comparing the performance of the proposed flexible bus dispatching system with different penetration rates of modular units to that of the existing dispatching system (i.e. consisting only of conventional buses). Notice that, even in case of smaller bus fleet, the proposed system significantly outperforms the existing one, especially from the operator perspective. As the size of the bus fleet increases, the improvements increase. Potential reductions in the total and the operator cost function can be up to 32% and 75%, respectively, depending on the level of car and passenger demand, penetration rate of modular bus units, and the size of the bus fleet. Interestingly, there seems to be a critical number of modular bus units, after which the marginal benefits of additional modular units are negligible (observe similar improvements in the total system cost for $p_m \geq 0.4$ and $p_m \geq 0.2$ in case of $F = 8$ and $F = 16$, respectively). This indicates the robustness of the proposed optimization framework.

6.3 Summary

In this chapter, we propose a novel concept, called flexible bus dispatching system, which offers new perspectives and enormous flexibility to better manage the dispatching frequencies and the allocation of the vehicle resources, reducing thereby the operating cost. In such a flexible bus dispatching system, the bus fleet consists not only of conventional buses, but also of modular and fully automated bus units that can either operate individually or combined together (forming thereby a single modular unit of a higher passenger capacity). To determine the optimal number of combined modular bus units and the optimal frequency at which the units (both con-
Table 6.3: Effects of the size of the bus network and the bus fleet on the cost functions.

| %   | 9.6% | 9.5% | 9.5% | 8.7% | 7.9% | 7.0% | 6.7% | 6.4% | 6.2% | 6.0% | 5.7% | 5.5% | 5.2% | 5.0% | 4.8% | 4.6% | 4.4% | 4.2% | 4.0% | 3.8% | 3.6% | 3.4% | 3.2% | 3.0% | 2.8% | 2.6% | 2.4% | 2.2% | 2.0% |
|-----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| %   | 2.0% | 1.9% | 1.8% | 1.7% | 1.6% | 1.5% | 1.4% | 1.3% | 1.2% | 1.1% | 1.0% | 0.9% | 0.8% | 0.7% | 0.6% | 0.5% | 0.4% | 0.3% | 0.2% | 0.1% |
| %   | 4.0% | 3.9% | 3.8% | 3.7% | 3.6% | 3.5% | 3.4% | 3.3% | 3.2% | 3.1% | 3.0% | 2.9% | 2.8% | 2.7% | 2.6% | 2.5% | 2.4% | 2.3% | 2.2% | 2.1% |
| %   | 6.0% | 5.9% | 5.8% | 5.7% | 5.6% | 5.5% | 5.4% | 5.3% | 5.2% | 5.1% | 5.0% | 4.9% | 4.8% | 4.7% | 4.6% | 4.5% | 4.4% | 4.3% | 4.2% | 4.1% |
| %   | 8.0% | 7.9% | 7.8% | 7.7% | 7.6% | 7.5% | 7.4% | 7.3% | 7.2% | 7.1% | 7.0% | 6.9% | 6.8% | 6.7% | 6.6% | 6.5% | 6.4% | 6.3% | 6.2% | 6.1% |
| %   | 10.0%| 9.9% | 9.8% | 9.7% | 9.6% | 9.5% | 9.4% | 9.3% | 9.2% | 9.1% | 9.0% | 8.9% | 8.8% | 8.7% | 8.6% | 8.5% | 8.4% | 8.3% | 8.2% | 8.1% |
| %   | 12.0%| 11.9%| 11.8%| 11.7%| 11.6%| 11.5%| 11.4%| 11.3%| 11.2%| 11.1%| 11.0%| 10.9%| 10.8%| 10.7%| 10.6%| 10.5%| 10.4%| 10.3%| 10.2%|
| %   | 14.0%| 13.9%| 13.8%| 13.7%| 13.6%| 13.5%| 13.4%| 13.3%| 13.2%| 13.1%| 13.0%| 12.9%| 12.8%| 12.7%| 12.6%| 12.5%| 12.4%| 12.3%| 12.2%|
| %   | 16.0%| 15.9%| 15.8%| 15.7%| 15.6%| 15.5%| 15.4%| 15.3%| 15.2%| 15.1%| 15.0%| 14.9%| 14.8%| 14.7%| 14.6%| 14.5%| 14.4%| 14.3%| 14.2%|

| %   | 0.0% | 0.3% | 0.7% | 1.1% | 1.5% | 1.9% | 2.3% | 2.7% | 3.1% | 3.5% | 3.9% | 4.3% | 4.7% | 5.1% | 5.5% | 5.9% | 6.3% | 6.7% | 7.1% | 7.5% | 7.9% | 8.3% | 8.7% | 9.1% | 9.5% | 9.9% | 10.3%|

<table>
<thead>
<tr>
<th></th>
<th>system delay</th>
<th>operator cost</th>
<th>total cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>0.0%</td>
<td>0.3%</td>
<td>0.7%</td>
</tr>
</tbody>
</table>
ventional and modular) should be dispatched across different bus lines, while accounting for the traffic dynamics at the network level, we propose an optimization framework based on the 3D-MFD. To the best of our knowledge, this is the first application of the 3D-MFD and modular bus units for the frequency setting problem in the domain of bus operations.

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

If you can’t explain it simply, you don’t understand it well enough.
— Albert Einstein

This doctoral research is devoted to the development of novel macroscopic tools for monitoring, modeling, design, and optimization of multi-modal urban systems based on the recently proposed 3D-MFD. The 3D-MFD offers great potential to model the complex dynamics of urban networks while capturing the interactions between different transportation modes, including those between buses and cars. Therefore, efforts are made to identify and address the challenges associated with the utilization of the 3D-MFD for the aforementioned applications. The main contributions and findings of this dissertation, as well as the future research directions are summarized below.

7.1 SUMMARY OF CONTRIBUTIONS AND FINDINGS

The main contributions and findings are clustered into the four main applications of the 3D-MFD.

1. Monitoring multi-modal systems from the macroscopic perspective with a limited amount of available data.

In Chapter 3, we explore the potential of using the AVL data provided by public transport vehicles for estimating the car space-mean speed - contribution (1.1). This, in turn, allows us to monitor the performance of multi-modal urban systems and determine the current level of traffic congestion with the limited data, i.e. when no LDD nor FCD is available. We further investigate the potential of improving the estimation of the average car speed at the network level by combining the AVL and FCD data sources where both are available, but FCD is not complete - contribution (1.2). As a result, we propose two novel Lagrangian estimation methods for the space-mean speed of cars based on: (i) the AVL data only; and (ii) the fused AVL and
FCD. Both methods account for the network configuration layout and the configuration of the public transport system.

Using empirical data from the city of Zurich, we demonstrate the applicability and validate the accuracy of the proposed methods in real-life traffic scenarios - contribution (1.3). In addition, we provide an empirical cross-comparison with the existing estimation methods - contribution (1.4). Such empirical comparison is, to the best of our knowledge, the first of its kind. The findings show that the proposed AVL-based estimation method can provide a good approximation of the average speed of cars at the network level. On the other hand, by fusing the FCD and AVL data, especially in case of sparse FCD, it is possible to obtain a more representative outcome regarding the performance of multi-modal traffic.

2. **Modeling multi-modal systems from the macroscopic perspective while incorporating stochasticity in bus operations.**

In Chapter 4, we extend the estimation analysis of Chapter 3 and explore the potential of using VT to derive the multi-modal p-MFD while capturing the stochastic nature of bus operations. For this purpose, we extend the existing VT approaches by: (i) introducing a probabilistic VT graph, where the costs are computed using an efficient stochastic shortest path algorithm - contribution (2.1); (ii) capturing the effects of stochastic moving bus bottlenecks and the correlation of bus arrival times - contribution (2.2); (iii) incorporating a macroscopic passenger model that reflects the passenger dynamics for the different modes - contribution (2.3); and (iv) accounting for the effects that the traffic conditions might have on bus operations - contribution (2.4).

Using a Monte-Carlo simulation and empirical data from a multi-modal corridor in Zurich, Switzerland, we not only successfully validate the results yielded by the proposed stochastic VT approach, but also show its applicability on a real corridor. A comparison with a deterministic VT approach reveals the value of the proposed framework, especially for corridors with a high bus frequency and considerable stochasticity. The results demonstrate that incorporating stochasticity and the traffic conditions is essential if buses run with relatively short and variable headways. Moreover, we introduce an innovative application example for the evaluation of different bus
lane layouts, aiming to maximize the passenger throughput along a multi-modal urban corridor - contribution (2.5). The application example shows that the proposed framework can be used as an efficient modeling tool for practitioners. In particular, it can be used to identify a proper lane allocation strategy by computing the critical density of cars when a mixed lane should be switched to a dedicated bus lane or vice versa. It is important to note that such application would not have been possible without our proposed VT extensions, which account for both passenger dynamics and the impact of traffic conditions.

3. Designing multi-modal systems from the macroscopic perspective while taking into account the complex modal interactions.

In Chapter 5, we establish a general framework for the bus network design problem considering multiple trip length patterns, two types of user behavior, and the effects that the bus network structure might have on the traffic performance and passenger mode choice. For modeling different trip length patterns, we use the trip length distribution as an intermediate level of abstraction - contribution (3.1). To capture complex modal interactions and quantify the operating speeds, we apply the modeling framework developed in Chapter 4 - contribution (3.2). We use the operating speed for each mode to determine the mode choice at the trip length level. This way, we are able to solve the optimal bus network design problem under the free-flow/saturated traffic conditions in an analytical way, while considering more realistic settings including: a dynamic description of the peak hour demand; spatially non-uniform network topology, distribution of the passenger demand, and design parameters across cardinal directions; mixed traffic; and different mode choice decisions depending on trip lengths and walking preferences - contributions (3.3), (3.4), and (3.5).

Numerical analysis reveals that all the tested factors, including demand intensity, user behavior, and trip length patterns, have significant effects on the operator and user cost function - contribution (3.6). Results show that the probability of choosing any given mode follows certain distribution that varies across the trip length patterns, indicating the importance of modeling the mode choice at the trip length level. Furthermore, the results indicate that users can benefit if they are willing to adjust the number of transfers to minimize the walking distance at the origin and the destination. Moreover, we show
that the optimal bus network design determined for the uniform trip pattern underestimates the number of required buses, which leads to passenger congestion at stops during the peak period. This, however, does not happen when we take into account the actual trip length distribution for the bus network design. A comparison with a simplified approach that considers the bus system only, reveals the value of accounting for the complex modal interactions, especially for higher demand levels. Finally, we show that by allowing the design parameters to vary across cardinal directions we provide more flexibility for the bus system to serve the passenger demand while reducing the operator cost compared to the existing approaches.

4. **Optimizing the performance of urban systems from the macroscopic perspective while capturing multi-modal traffic dynamics.**

In Chapter 6, we study the subsequent phase of the transit network planning process, the frequency setting problem. We propose a novel *flexible bus dispatching system* in which a fleet of fully automated modular bus units, together with conventional buses, serves the passenger demand - contribution (4.1). These modular bus units can either operate individually or combined (forming larger modular buses with a higher passenger capacity). This provides enormous flexibility to manage the service frequencies and vehicle allocation, reducing thereby the operating cost and improving passenger mobility.

We develop an optimization model used to determine the optimal composition of modular bus units and the optimal service frequency at which the buses (both conventional and modular) should be dispatched across each bus line - contribution (4.2). We explicitly account for the dynamics of traffic congestion and complex interactions between the modes at the network level based on the 3D-MFD - contribution (4.3). To the best of our knowledge, this is the first application of the 3D-MFD and modular bus units for the frequency setting problem in the domain of bus operations.

Numerical results show the improvements in the total system cost made by adjusting the number of combined modular bus units and their dispatching frequencies to the evolution of both, the car and the bus passenger demand. A comparison with the commonly used approach that considers only the bus system (neglecting the complex modal interactions and congestion propagation) reveals the value of
the proposed modeling framework. Finally, a sensitivity analysis of the effect of the operating unit cost of modular bus units, the size of the bus network, and the size of the bus fleet sheds light on the robustness of the proposed optimization framework - contribution (4.4).

7.2 OUTLOOK

The potential of using the 3D-MFD for the purpose of monitoring and optimizing the performance of multi-modal systems, as well as for modeling the system dynamics and designing the optimal network structure is a relatively new and promising research area. Given the complexity of urban networks, it is foreseeable that research on the macroscopic analysis of multi-modal traffic will continue in the future. In the following, we list potential future research directions, summarized in four categories.

1. Improving estimation of traffic variables in the presence of limited data.

In Chapter 3, we assume that the traffic information is acquired either from LDD, FCD, or AVL data. Recent years have witnessed a significant progress in the vehicle and communication technology, largely driven by advances in vehicle automation. Connected and automated vehicles are becoming increasingly popular in the transportation sector, as they offer a great potential to mitigate congestion in cities through a more efficient utilization of the existing roads. Their deployment could be followed by the installation of advanced communication devices throughout a city, which can serve as an additional data source. Therefore, contributions (1.1) and (1.2) can be extended to investigate how the information provided by these advanced communication devices can be integrated into the proposed framework to yield a more accurate estimation of the macroscopic traffic variables.

2. Improving modeling of stochastic traffic processes.

The advantages of the proposed methodology for modeling stochastic bus operations in Chapter 4 are not restricted to the multi-modal case, but open possibilities to model other stochastic processes in the field of traffic. For example, contributions (2.1), (2.4), and (2.5) can be used to model adaptive traffic control systems, one of the
most commonly used traffic solutions in urban areas, i.e. contribution (2.1) allows to treat traffic signals as stochastic bottlenecks, similarly to bus stop shortcuts in this paper, whereas contribution (2.4) gives the opportunity to incorporate effects of traffic conditions, being the driving parameter for determining adaptive signal settings (36). Contribution (2.5) can be extended to propose the dynamic allocation of space across modes. Furthermore, contribution (2.3) allows to model different modes and therefore can, in combination with contribution (2.1), be used to investigate different types of infrastructures (e.g. managed lanes) following the rationale from contribution (2.5). It is worth mentioning that our stochastic VT approach allows to model different types of dependencies, such as those in bus operations. For example, while in some scenarios we can incorporate dependent bus arrivals regardless of the traffic conditions, for other scenarios we can relate the dependency of bus arrivals to the traffic conditions. Combining contributions (2.1), (2.2), (2.3), and (2.4) it is possible to define a convenient analytical model for a network with complex interactions between autonomous vehicles (AVs) and conventional vehicles. For such a case, contributions (2.1) and (2.2) can define the stochastic VT graph including the moving bottlenecks that conventional vehicles might potentially represent for AVs; contribution (2.3) can account for different cost functions of both vehicle types; and contribution (2.4) can model the influence of AV platoons on traffic conditions. Examples such as these underline the value of the proposed stochastic extension of VT and its potential across a wide range of applications. Future efforts can be devoted to further explore those opportunities.

3. **Improving design of urban networks.**

Formulations presented in Chapter 5 can be considered as the first building block towards the design of public transport networks for grid-like structures, which consider for the first time the interactions with the traffic system, mode choice behavior, and heterogeneous demand patterns and types of user behavior. Contributions (3.1), (3.2), (3.3), and (3.4) can be extended to model radial or ring street layouts. It is also worth mentioning that, although the real networks are more complicated, the formulated design problem for simplified networks gives a fast global assessment of the key design variables such as range of bus headways, overall fraction of dedicated bus lanes, and range of stop and line spacings per cardinal direction. These design
variables should be further analyzed at the operational level when applied to real networks. This can be also be a potential direction for future research.

4. **Improving operations and performance of multi-modal systems.**

Chapter 6 opens several research directions. First, it is interesting to study how contribution (4.1) can be integrated with other traffic management strategies, e.g. a multi-modal perimeter control (32, 151), to further improve the operations of multi-modal systems. Furthermore, contribution (4.2) can be extended to more complex scenarios (e.g. multiple-region networks), and a hierarchical control framework (31) can be developed that converts the macroscopic level bus dispatching decision to account for more detailed operational features (e.g. timetables, en-route combining and splitting of modular units along a bus line). Moreover, the en-route combining and splitting features can be used to further develop other bus-related strategies (e.g. stop skipping and dynamic routing). In addition, it would be interesting to incorporate the automated modular bus units into other promising paradigms, e.g. shared mobility. For example, we can design real-time demand responsive bus services with these units to further improve passenger mobility. We may also integrate the proposed system with mobility-on-demand systems, where the public transport operator maintains a larger fleet of automated modular units and some units provide last-mile on-demand services. Finally, it would be meaningful to evaluate the long-term benefits of the proposed strategy from a planning perspective by accounting for shift in mode choice, route choice, and departure time.

7.3 **Concluding Remarks**

With a rapid increase in urbanization rates across the world, traffic congestion has become a substantial burden for urban development in cities worldwide. The need to provide new mobility solutions that ensure a more efficient, sustainable, and resilient transportation system is therefore apparent and urgent. Recent advances in vehicle technology and communication devices have offered a great potential to mitigate congestion in urban areas through a more efficient utilization of the existing road infrastructure. In this dissertation, we explored such potential by proposing novel macroscopic tools for monitoring, modeling, design, and optimiza-
tion of multi-modal transportation systems considering new technologies and data sources. The proposed tools can be seen as guidelines for policy makers and practitioners on how to better design, monitor, model, and operate urban transportation systems that will propel cities into a richer and more sustainable future. Ultimately, this dissertation aims at the promotion of multi-modality and sustainable mobility in urban transportation systems.
**Result 1.** The expected walking time at the origin and the destination for Type 2 user behavior in Chapter 5 is given by Eqs. 5.27 and 5.30:

**Case 1** \((l_x \geq l_y)\).

\[ A_{2,b}(D) = \frac{(6l_xl_y + 6l_xs_x - 2l_y^2 - 3l_y s_x + 3l_y s_y)}{(12l_x w)}. \]

**Case 2** \((l_x < l_y)\).

\[ A_{2,b}(D) = \frac{(6l_xl_y + 6l_y s_y - 2l_x^2 - 3l_x s_y + 3l_x s_x)}{(12l_y w)}. \]

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

**Case 1** \((l_x < l_y)\). Regions (1) and (3) consist of two subregions: a rectangle with sides \(l_x - l_y\) and \(l_y/2\); and two right triangles with legs \(l_y/2\) and \(l_y/2\). The first subregion has the area of \(P_{(1)}^{(1)} = (l_x - l_y)l_y/2\) and the associated average walking distance of \((1/2)(l_y/2) = l_y/4\) along N-S and \((1/2)(s_x/2) = s_x/4\) along E-W directions, i.e. the average walking distance is \(A_{(1)}^{(1)} = (l_y/4 + s_x/4)/w\). The area of the second subregion (combining the two triangles) is \(P_{(1)}^{(1)} = l_y^2/4\), with the associated average walking distance of \((1/3)(l_y/2) = l_y/6\) along N-S and \((1/2)(s_x/2) = s_x/4\) along
E-W directions, i.e. $A^{(1)}_{\triangle} = (l_y/6 + s_x/4)/w$. Similarly, regions (2) and (4) have each the area of $P^{(2)}_{\triangle} = l_y^2/4$, with the associated average walking distance of $(1/3)(l_y/2) = l_y/6$ along E-W and $(1/2)(s_y/2) = s_y/4$ along N-S directions, i.e. $A^{(2)}_{\triangle} = (l_y/6 + s_y/4)/w$. Then, assuming that the expected access and egress distances traveled are the same (similarly to 59, 60), the expected walking time can be computed by considering all possible combinations of the subregions in which the origin/destination might be located:

$$A_{2,b}(D) = 2 \left( A^{(1)}_{\square} \cdot \frac{P^{(1)}_{\square}}{P^{(1)}_{\square} + P^{(1)}_{\triangle} + P^{(2)}_{\triangle}} + A^{(1)}_{\triangle} \cdot \frac{P^{(1)}_{\triangle}}{P^{(1)}_{\square} + P^{(1)}_{\triangle} + P^{(2)}_{\triangle}} + A^{(2)}_{\triangle} \cdot \frac{P^{(2)}_{\triangle}}{P^{(1)}_{\square} + P^{(1)}_{\triangle} + P^{(2)}_{\triangle}} \right)$$

$$= \left( 6l_xl_y + 6l xs_x - 2l_y^2 - 3l_y s_x + 3l_y s_y \right) / (12l_xw).$$

**Figure A.1:** Regions defining potential locations of an origin/destination within a rectangle defined by sides $l_x$ and $l_y$ in case: (a) $l_x \geq l_y$; and (b) $l_x < l_y$.

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

\[
A_{2,b}(D) = 2 \left( A^{(1)}_\triangle \cdot \frac{p^{(1)}_\triangle}{p^{(1)}_\triangle + p^{(2)}_\square + p^{(2)}_\triangle} + A^{(2)}_\square \cdot \frac{p^{(2)}_\triangle}{p^{(1)}_\triangle + p^{(2)}_\square + p^{(2)}_\triangle} \right) + A^{(2)}_\triangle \cdot \frac{p^{(2)}_\triangle}{p^{(1)}_\triangle + p^{(2)}_\square + p^{(2)}_\triangle} \right) 
= (6l_xl_y + 6ls_y - 2l_x^2 - 3ls_x + 3ls_s) / (12lw).
\]

Result 2. The expected number of transfers per trip for Type 2 user behavior in Chapter 5 is given by Eqs. 5.28 and 5.31:

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

\[
\varphi_2(D) = \Pr(\varphi > 0) \left( 4l_x^2l_y^2 - 2l_xl_y^3 + l_y^4 \right) / (2l_x^2l_y^2).
\]

Case 2 \((l_x < l_y)\).

\[
\varphi_2(D) = \Pr(\varphi > 0) \left( 4l_x^2l_y^2 - 2l_xl_y^3 + l_x^4 \right) / (2l_x^2l_y^2).
\]

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

**Case 1** \((l_x \geq l_y)\).

\[
\varphi_2(D) = 0 \cdot \Pr(\varphi = 0) + \Pr(\varphi > 0)(1 \cdot \Pr(\varphi = 1) + 2 \cdot \Pr(\varphi = 2)) \\
= \Pr(\varphi > 0)\left(1 \cdot \frac{2(P^{(1)} + P^{(1)})P^{(2)}(P^{(1)} + P^{(1)})}{(P^{(1)} + P^{(1)} + P^{(2)})^2} + 2 \cdot \frac{(P^{(1)} + P^{(1)})^2 + (P^{(2)})^2}{(P^{(1)} + P^{(1)} + P^{(2)})^2}\right) \\
= \Pr(\varphi > 0)(4l_x^2l_y^2 - 2l_xl_y^3 + l_y^4)/(2l_x^2l_y^2).
\]

**Case 2** \((l_x < l_y)\).

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

**Result 3.** The expected waiting time per user including that at the origin and all transfer stops for Type 2 user behavior in Chapter 5 is given by Eqs. 5.29 and 5.32:

**Case 1** \((l_x \geq l_y)\).

\[
W_{2,p}(D) = \Pr(\varphi = 0)\bar{H}/2 + \Pr(\varphi > 0)(H_x(8l_x^2l_y^2 - 4l_xl_y^3 + l_y^4) \\
+ H_y(4l_x^2l_y^2 + l_y^4))/(8l_x^2l_y^2).
\]
Case 2 \((l_x < l_y)\).

\[
W_{2,b}(D) = \Pr(\varphi = 0)\bar{H}/2 + \Pr(\varphi > 0)(H_x(4l_x^2l_y^2 + l_x^4) + H_y(8l_x^2l_y^2 - 4l_x^3l_y + l_x^4)) / (8l_x^2l_y^2).
\]

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

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

\[
W_{2,b}(D) = H(i) \cdot \Pr(\varphi = 0) + \Pr(\varphi > 0)(H(ii) \cdot \Pr(ii)
+ H(iii) \cdot \Pr(iii) + H(iv) \cdot \Pr(iv))
\]

\[
= H(i) \cdot \Pr(\varphi = 0) + \Pr(\varphi > 0)\left(H(ii) \cdot \frac{(P^{(1)} + P^{(1)})^2}{(P^{(1)} + P^{(1)} + P^{(2)})^2}
+ H(iii) \cdot \frac{(P^{(2)})^2}{(P^{(1)} + P^{(1)} + P^{(2)})^2} + H(iv) \cdot \frac{2(P^{(1)} + P^{(1)})(P^{(2)})}{(P^{(1)} + P^{(1)} + P^{(2)})^2}\right)
\]

\[
= \Pr(\varphi = 0)\bar{H}/2 + \Pr(\varphi > 0)(H_x(4l_x^2l_y^2 - 4l_x^3l_y + l_x^4)
+ H_y(4l_x^2l_y^2 + l_x^4)) / (8l_x^2l_y^2).
\]
Case 2 ($l_x < l_y$).

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

\[\blacksquare\]

Result 4. When the trip origins and destinations are uniformly and independently distributed across the network, Eq. 6.27 becomes equivalent to that of (59) for grid-like structures:

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

Proof. It suffices to show that the maximum number of on-board passengers along E-W/N-S directions is across the vertical/horizontal cordon that bisects the city, i.e. $\beta = 1/2$ (see Fig. 5.2). For brevity, we only show this for the $eb$ direction. Given the symmetry of the network, derivations for other cardinal directions are done in an entirely analogous manner using the corresponding subscripts.

To be able to obtain the same result as (59), we need to use the same assumption that the abscissa and the ordinate of both, the trip origins $(x_1, y_1)$ and the destinations $(x_2, y_2)$, are distributed uniformly and independently on $[0, \phi_x]$ and $[-\phi_y, \phi_y]$, respectively. As abscissae $x_1$ and $x_2$ are independent uniforms, their difference $x_1 - x_2$ has a triangular distribution, hence the user trip length in the $eb$ direction $\{\ell_x = x_1 - x_2 : \ell_x \geq 0\}$ has the following probability density function: $\{f(\ell_x) = (\phi_x - \ell_x)/\phi_x^2 : \ell_x \in [0, \phi_x]\}$. Furthermore, due to the independence, the joint probability density function becomes $f(\ell_x, \ell_y) = f(\ell_x)f(\ell_y)$. Finally, since (59) does not consider
the mode choice, we also assume that \( \Pr(b | \ell_x, \ell_y) = 1, \forall \ell_x, \ell_y \). It follows that the maximum number of on-board passengers in the \( eb \) direction is:

\[
Q_{eb} = \max_{\beta \in [0,1]} \left\{ \int_0^{\phi_x} \int_{-\phi_y}^{\phi_y} \Pr(J_{\beta}(\ell_x)) \Lambda \Pr(b | \ell_x, \ell_y) f(\ell_x, \ell_y) d\ell_y d\ell_x \right\}
\]

\[
= \max_{\beta \in [0,1]} \left\{ \frac{\Lambda}{\phi_x^2} \int_0^{\phi_x} \Pr(J_{\beta}(\ell_x))(\phi_x - \ell_x) d\ell_x \right\}
\]

\[
= \max_{\beta \in [0,1]} \left\{ \frac{\Lambda}{\phi_x^2} \left( \int_0^{\min\{\beta \phi_x, (1-\beta) \phi_x\}} \frac{\ell_x}{\phi_x - \ell_x} (\phi_x - \ell_x) d\ell_x + \int_{\min\{\beta \phi_x, (1-\beta) \phi_x\}}^{\max\{\beta \phi_x, (1-\beta) \phi_x\}} \frac{\min\{\beta \phi_x, (1-\beta) \phi_x\}}{\phi_x - \ell_x} (\phi_x - \ell_x) d\ell_x + \int_{\max\{\beta \phi_x, (1-\beta) \phi_x\}}^{\phi_x} (\phi_x - \ell_x) d\ell_x \right) \right\}
\]

\[
= \max_{\beta \in [0,1]} \left\{ \frac{\Lambda}{\phi_x^2} \left( \frac{1}{2} (\max\{\beta \phi_x, (1-\beta) \phi_x\})^2 + \min\{\beta \phi_x, (1-\beta) \phi_x\} (\max\{\beta \phi_x, (1-\beta) \phi_x\} - \min\{\beta \phi_x, (1-\beta) \phi_x\}) + \phi_x (\phi_x - \max\{\beta \phi_x, (1-\beta) \phi_x\}) - \frac{1}{2} (\phi_x^2 - (\max\{\beta \phi_x, (1-\beta) \phi_x\})^2) \right) \right\}
\]

\[
= \max_{\beta \in [0,1]} \left\{ \frac{\Lambda}{\phi_x^2} \left( \min\{\beta \phi_x, (1-\beta) \phi_x\} \max\{\beta \phi_x, (1-\beta) \phi_x\} - \frac{1}{2} (\min\{\beta \phi_x, (1-\beta) \phi_x\})^2 + \frac{1}{2} \phi_x^2 - \phi_x \max\{\beta \phi_x, (1-\beta) \phi_x\} + \frac{1}{2} (\max\{\beta \phi_x, (1-\beta) \phi_x\})^2 \right) \right\}.
\]

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

**Case 1** \( (\beta \phi_x \leq (1-\beta) \phi_x) \).

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

\[
= \Lambda (\beta - \beta^2).
\]

**Case 2** \( (\beta \phi_x > (1-\beta) \phi_x) \).

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

\[
= \Lambda (\beta - \beta^2).
\]
Notice that both cases yield the same result. Therefore, we can compute the value of $\beta$ for which the maximum value of $Q_{eb}(\beta)$ is reached by solving the following equation:

$$\frac{dQ_{eb}(\beta)}{d\beta} = 0 \leftrightarrow \Lambda(1 - 2\beta) = 0 \leftrightarrow \beta = \frac{1}{2}. $$

It then follows that the maximum number of on-board passengers in the $eb$ direction is $Q_{eb} = Q_{eb}(\beta = 1/2) = \Lambda/4$, which is the same result as that of (59). \hfill \blacksquare


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


CURRICULUM VITAE

PERSONAL DATA

Name Igor Dakic
Date of Birth 13. Dec. 1990
Place of Birth Belgrade, Serbia
Citizen of Serbia

EDUCATION

01/2015 – 08/2016 Florida Atlantic University, United States
Master of Science in Civil Engineering

10/2013 – 12/2014 University of Belgrade, Serbia
Master of Science in Transportation Engineering

10/2009 – 09/2013 University of Belgrade, Serbia
Bachelor of Science in Transportation Engineering

EMPLOYMENT

01/2022 – present Consultant
AFRY, Switzerland

09/2021 – 11/2021 Lead Transport Systems Engineer
EuroTube Foundation, Switzerland

09/2016 – 08/2021 Research Assistant
ETH Zurich, Switzerland

01/2020 – 06/2020 Visiting Research Assistant
Georgia Institute of Technology, United States

03/2019 – 08/2019 Visiting Research Assistant
University of Lyon, France

01/2015 – 08/2016 Research Assistant
Florida Atlantic University, United States
AWARDS & SCHOLARSHIPS

2019  
*Doc.Mobility fellowship*  
Swiss National Science Foundation

2014 – 2015  
*Dositeja scholarship*  
Serbian Foundation for Young Talents

2013 – 2014  
*Studenica Foundation scholarship*  
Serbian Unity Congress

2013  
*The best student of generation*  
University of Belgrade, Serbia

2012 – 2013  
*City Government scholarship*  
City of Belgrade, Serbia

2012  
*Michael Pupin Foundation* (top 20 students in Serbia)  
Serbian National Defense Council

2012  
*The best Third Year Student*  
University of Belgrade, Serbia

2011 – 2012  
*Nikola Oka Foundation scholarship*  
University of Belgrade, Serbia

2011  
*The best Second Year Student*  
University of Belgrade, Serbia

2010 – 2011  
*Ministry of Science and Education scholarship*  
Government of Serbia

2010  
*The best First Year Student*  
University of Belgrade, Serbia


I. Dakic, M. Menendez. On the use of Lagrangian observations from public transport and probe vehicles to estimate car space-mean speeds in bi-


**PEER-REVIEWED CONFERENCE PROCEEDINGS**


**INTERNATIONAL CONFERENCES**
