A Dynamic Network Approach for Multimodal Urban Mobility: Modeling, Pricing and Control

THÈSE N^{O} 6456 (2014)

PRÉSENTÉE LE 12 DÉCEMBRE 2014
À LA FACULTÉ DE L'ENVIRONNEMENT NATUREL, ARCHITECTURAL ET CONSTRUIT
LABORATOIRE DE SYSTÈMES DE TRANSPORTS URBAINS
PROGRAMME DOCTORAL EN GÉNIE CIVIL ET ENVIRONNEMENT

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Nan ZHENG

acceptée sur proposition du jury:

Prof. A. J. Wüest, président du jury
Prof. N. Geroliminis, directeur de thèse
Prof. M. Bierlaire, rapporteur
Prof. H. Koutsopoulos, rapporteur
Prof. H. van Lint, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE
Suisse
2014

Acknowledgement

After I obtained Master degree of Transportation from Delft, I visited my bachelor supervisor Prof. Xiucheng Guo of Southeast University in Nanjing. It was right before I started my PhD in September 2010. I made Prof. Guo laugh after telling him that I was going for academia which he would never expect from me. Nevertheless I got his bless and faith as always, and came to Lausanne. At this moment, I finish my PhD on large-scale traffic modeling and control, a challenging and novel field that I had no idea about four years ago. I struggled to learn, published papers, and succeeded in being a better person I could be. I would not have had such life achievement without the help of the important people during my 4–year life, and I would like to take this opportunity to express my appreciation.

My heartful thanks first go to my supervisor Prof. Nikolas Geroliminis. Your creativity, vision, and enormous amount of knowledge have guided the shape of this thesis. You have been treating me with unlimited patience, and assisting me in improving academic thinking and the ways of delivering ideas and findings. I count myself extremely lucky to have you, not only that I have learned so much from the every single discussion with you, but also that I have been deeply influenced by your rigorous altitude towards science, passion towards life, and humor towards both science and life. I cannot say thank you enough.

Secondly, I would like to thank my PhD jury committee members, Prof. Johny Wüest, Prof. Michel Bierlaire, Prof. Haris Koutsopoulos, and Prof. Hans van Lint for dedicating their time in reading my thesis and giving valuable feedbacks. My special thanks go to Michel, for your many critical and helpful comments on my work, in private, during the joint internal seminars, and in Swiss Transportation Research conferences; to Hans, my former teacher from Delft and long-time admired musician, for coming a long way and being there for me; and to Haris, for joining the final defense far in Boston through video-conference.

I am very fortunate to work in LUTS, my beloved lab, with smart colleagues of diverse background. I wish to firstly thank the first generation of the lab, Burak, Jack, Kostas, Mehmet, Mohsen and Yuxuan, whom I spent most of my 4-year life with. It has been lovely with you guys. All the best and respect to the new members of the lab, Reza, Martin, Mikhail, Raphael, Tasos and Wei, I am sure you would excel further the research quality of LUTS. And thank you all for your tolerance towards my personality and my weird jokes and stories.

I would also like to mention my colleagues from the sister labs, Litep, Lavoc, and Transp-or for their kindly support and friendship, particularly to Alex, Antonin, Anastasia, Jingmin, Jianghang, Panos, Sofia and Xinjun. I really enjoy the collaborations with Guillaume from Civil Service, and Rashid, Prof. Axhausen from ETH Zurich. Furthermore, I have received encouragement from my old-time supervisors at TU Delft, Prof. Serge Hoogendoorn, Prof. Henk van Zuylen and Prof. Andreas Hegyi, and many colleagues working on transportation all over the world. I am grateful for all of you.

I need to thank my many friends outside of work. I cannot imagine my life in Lausanne without the friendship and assistance from the following three persons: Christine, the most kind and lovable secretary of all time, and Matthieu and Gary, the crazy sport lovers. I appreciate a million your time, effort and patience on solving the countless troubles that I frequently bring to you. Work-related, living issues, no matter what, you sit down with me and find solutions. You treat me as family, much bigger than being a friend.

For the many people that I encounter during sport and social activities, you have made my life colorful. It has been a pleasure for me to work with the enthusiastic members of the Chinese Student and Scholar Association Lausanne, and leading the organization of various unforgettable events. I am glad to be part of the Chinese badminton team Lausanne, as well as to be participants in the annual tennis tournaments of EPFL-CSP and Roger’s cup. I need to thank particular my tennis tutor Vincent who brought me into the competitive and mental world of tennis.

My long-time friends from China, Ning, Bai, Peng, Yu, Haixin, Mingmin, Zhendong, Meng, Yufei, Minwei, Huizhao, thank you for your greetings and advices from time to time.

Last but the most important, I wish to express my sincere gratefulness to my family. Your deep affection and patience have been moving and encouraging me for so long. With this PhD, I learn the way to perceive. With you, I learn the way to love.

I thank all the sages, families and friends, who have come along with me during my PhD journey, for smoothing the path and pointing the direction to success. My endless wishes to you all.

Abstract

The complexity in road infrastructure and human travel behavior makes the performance of urban transport system hard to predict, especially in real time. Recent advances in traffic flow theory at the network level, namely the Macroscopic Fundamental Diagram (MFD), reveals the existence of well-defined laws of congestion dynamics at aggregated levels. The MFD has been demonstrated as a reliable tool for monitoring and controlling the performance of large-scale urban road networks. The same knowledge for multimodal networks however is limited. As people travel through different modes of transport such as cars, buses, taxis that compete for scarce urban road infrastructure, it is critical to understand how this space can be allocated and managed for multimodality. The objective is to develop aggregated modeling and optimization approaches, which will contribute on the knowledge of congestion dynamics in cities of different structures and mode usages, and ultimately facilitate the design of efficient and equitable urban transport policies, such as dedicated bus lane allocation, parking limitation and congestion pricing.

Building on the knowledge of the single-mode MFD theory, a bi-modal MFD model considering the effect of mode conflict is proposed for mixed networks of buses and cars. A system-level model is developed where a city network is decomposed to multiple regions with different characteristics. Each region has its unique mode-specific MFD, which is assumed known in advance, given space allocation condition. The flow dynamics among regions are described by a regional level flow conservation law, where vehicles are categorized by families regarding their traveling directions and the movement throughput is represented by the bi-modal MFD. An approximate dynamic mode choice is also developed based on imperfect information of the time-dependent cost estimated with variables from the MFD. Combining the bi-modal MFD flow model, the mode choice model and the aggregated flow conservation model, a non-linear optimization framework is performed to optimize space allocation, minimizing the total passenger cost, given certain demand, city structure and road facility. Then, parking limitation is integrated in the proposed multi-modal system model, where vehicles cruising for parking are also integrated. The extra delay of cruising is captured by a geometric distribution related to the time-dependent parking availability and estimated at the aggregated level. The delay cost to other users is also estimated via the bi-modal MFD, and it shows the effect of cruising on all travelers of the system, even the ones moving towards their destination and do not require parking. Optimal parking pricing policies for on-street and garage parking are obtained through the optimization framework, as well.

To identify and quantify the patterns of multimodal congestion, the existence of a three-dimensional MFD (3D-MFD) for mixed bi-modal networks is investigated and analyzed via micro-traffic simulation studies. A 3D-MFD relates vehicular production of a network (flow, travel distance) to the density of cars and buses, where the impact of each mode on network performance can be directly observed. An exponential function is proposed for the analytical form of the 3D-MFD. To further compare the modal impact on performance, the Bus-Car Unit equivalent value is estimated by the function form of the 3D-MFD. Result indicates that this value is state- and mode-composition dependent rather than deterministic. In addition to the conventional vehicle-flow-based analysis, passenger dynamics expressed by a passenger 3D-MFD are derived from the vehicle 3D-MFD, which provides a different perspective of the flow characteristics in bi-modal networks. The properties of the passenger 3D-MFD are discussed, and a partition algorithm is applied to identify the impact of heterogeneity of mode composition and congestion distribution on the bi-modal modeling. Simulation study on 3D-MFD based perimeter-control shows promising performance in real-time control of congested multimodal city centers.

The final part of the thesis concerns the development of MFD-controlled cordon- and area-based pricing schemes for congestion management. Feedback-type control mechanisms are proposed to determine and adjust the time-dependent tolls, based on congestion level as expressed by the MFD. In the case of the area-based pricing, the pricing scheme also considers user’s adaptation to the toll cost, allowing a great flexibility in toll adjustment, and deals with the promotion of public transport usage by integrating incentive programs such as accessibility improvement of buses using the collected toll revenue. The performance of the pricing schemes is investigated in an existing agent-based model where the complex travel behavior in real-life is reasonably reproduced. Results demonstrate that both pricing schemes are effective in congestion reduction, while the area-based pricing scheme achieves higher efficiency as large welfare gain can be obtained. Smooth behavioral equilibrium in long-term operation is found under such pricing schemes. Furthermore, user heterogeneity with respect to value-of-time is introduced in the agent-based model. Significant differences are found in behavioral response and trip cost. By realizing and treating this heterogeneity, pricing strategies can achieve even higher efficiency and equitable benefit.

The findings of this thesis are of great importance because the multi-modal MFD can be utilized to: (i) monitor and predict traffic performance in urban networks, and (ii) develop simple policies and identify different management strategies maximizing vehicle or passenger throughput for cities with heterogeneous regions and users. Both (i) and (ii) can be achieved with feasible data inputs and implementation costs therefore are readily applicable in large-scale urban city networks. Policy-makers can rely on the MFD-based tools to adjust management strategies and operate a city at different mobility levels.

Key words: Macroscopic Fundamental Diagram (MFD), Multimodality, Aggregated system dynamics, Space allocation, Dynamic dedicated bus lane allocation, Optimization, Passenger hours travelled, Parking limitation, Cruising-for-parking, Parking pricing, Congestion pricing, Feedback control, Agent-based approach

Résumé

La complexité du réseau routier et du comportement des usagers rendent difficile la prévision des performances d’un système de transport, a fortiori en temps réel. Cependant, une avancée récente dans la théorie du trafic – le Diagramme Macroscopique Fondamental (MFD) – montre l’existence de lois bien définies régissant le comportement dynamique de la congestion à un niveau agrégé. Il a été prouvé que le MFD est un outil fiable pour la surveillance et le contrôle à large échelle de réseaux routiers urbains. Néanmoins, ces outils ont été développés principalement pour les flux automobiles et la gestion des réseaux multimodaux n’a été que très partiellement étudiée. Or, les différents modes de transports sont aujourd’hui en compétition non seulement en termes de demande mais aussi en termes d’espace urbain. Il est donc primordial de comprendre comment l’allocation de cet espace urbain entre les différents modes affecte les performances du système. Le premier objectif de ce travail est de développer une modélisation agrégée adaptée afin de mieux comprendre les dynamiques de la congestion en fonction de la structure urbaine et des modes de transports utilisés. Le second objectif est de créer des méthodes d’optimisation basées sur la modélisation précédente afin de faciliter l’élaboration de politiques de transports efficaces et équitables, telles que la création de voies réservées aux bus, la gestion du stationnement et la tarification liée aux encombrements.

En s’appuyant sur la théorie du MFD pour un seul mode de transport, un modèle de MFD bi-modal est proposé pour les réseaux mixtes bus/automobiles, qui prend en compte les conflits entre ces deux modes. Un modèle est développé pour l’ensemble du système, dans lequel le réseau urbain est constitué de différentes régions. Chaque région est régie par son propre MFD, qui est considéré comme connu, étant donnée l’allocation de l’espace. Les dynamiques du trafic entre les régions sont décrites à l’échelle régionale par une loi de conservation, où les véhicules sont répartis en différentes catégories selon leur destination et le débit est exprimé grâce à un MFD bi-modal. Un modèle d’approximation dynamique du choix modal est aussi développé, en considérant une connaissance imparfaite des coûts en fonction du temps, basée sur des variables du MFD. En associant le MFD bi-modal, le modèle de choix modal et le modèle agrégé de conservation du flux, une méthode d’optimisation non-linéaire est proposée afin de minimiser le coût des usagers en optimisant l’allocation de l’espace, étant données la demande et les infrastructures de transport. Ensuite la gestion du stationnement est intégrée dans ce système multi-modal, prenant en compte les véhicules en mouvement à la recherche

d'une place de stationnement. Le délai additionnel représenté par ce phénomène est modélisé de manière agrégée par une distribution géométrique dépendant de l'offre de stationnement en temps réel. Le coût pour les autres usagers est lui aussi estimé, via le MFD bi-modal, montrant l'impact de ce phénomène sur l'ensemble du réseau. Des politiques optimales pour le stationnement sont obtenues par la méthode d'optimisation proposée.

Afin d'identifier et de quantifier les phénomènes de congestion avec différents modes de transport, l'existence d'un MFD tridimensionnel (3D-MFD) pour les réseaux bimodaux est étudiée à l'aide de simulations microscopiques du trafic. Un MFD tridimensionnel associe la production d'un réseau (en termes de distance parcourue par l'ensemble des véhicules par unité de temps) aux densités de bus et de voitures. Une fonction exponentielle est proposée pour décrire analytiquement la forme du MFD tridimensionnel. Pour approfondir cette étude, l'impact des bus dans le réseau est approximé par un nombre de voitures équivalent. Cependant, les résultats montrent que cette valeur dépend de l'état du réseau et de la répartition modale. En outre, un MFD tridimensionnel alternatif est proposé, dans lequel la production du réseau est exprimée en termes de distance totale parcourue par l'ensemble des passagers (et non des véhicules) par unité de temps. Les propriétés de ce MFD tridimensionnel alternatif sont étudiées et un algorithme de partitionnement est utilisé afin d'identifier l'impact de l'hétérogénéité de la décomposition modale et de la congestion sur le MFD bimodal. Une stratégie de contrôle du périmètre basée sur le MFD tridimensionnel est étudiée via des simulations et met en évidence le potentiel important du contrôle en temps réel pour les réseaux multimodaux des centres-villes soumis à la congestion.

La dernière partie de cette thèse est dédiée au développement de stratégies de tarification de la congestion basées sur le contrôle d'une région soit à son périmètre soit sur l'intégralité de son territoire. Des mécanismes de commande par rétroaction sont proposés afin de déterminer les tarifs des péages en fonction du temps, en utilisant les estimations de la congestion obtenues avec le MFD. Dans le cas d'une région contrôlée sur l'intégralité de son territoire, la stratégie de tarification prend en compte l'adaptation du comportement des usagers en fonction du coût et le potentiel développement des transports en commun obtenu en réinvestissant les recettes des péages. Les performances de telles stratégies sont étudiées grâce à un modèle multi-agents déjà existant, reproduisant assez fidèlement le comportement complexe des usagers. Les résultats montrent que les deux stratégies sont efficaces pour réduire la congestion mais que la stratégie contrôlant l'intégralité du territoire de la région est plus efficace en termes de bien-être social. Il est à noter qu'un équilibre est atteint quand ces stratégies sont mises en œuvre

sur le long terme. Pour approfondir cette étude, une hétérogénéité de la valeur du temps pour les différents usagers est introduite dans le modèle multi-agent, conduisant à des différences significatives du comportement. En prenant en compte cette hétérogénéité, les stratégies de tarification permettent d'obtenir une efficacité encore supérieure et des bénéfices équitablement répartis.

Les résultats de cette thèse sont d'une importance majeure car le MFD tridimensionnel peut être utilisé pour : (i) surveiller et prédire la performance de réseaux urbains, (ii) développer des politiques simples et des stratégies de gestion maximisant la production du réseau, en termes de véhicules ou de passagers, pour des villes avec des régions et des usagers hétérogènes. Les résultats (i) et (ii) peuvent être obtenus avec les données généralement disponibles et pour des coûts de mise en œuvre raisonnables. Ainsi, ces mesures sont directement applicables aux réseaux urbains à grande échelle.

Mots clés : Diagramme Macroscopique Fondamental (MFD), Multimodalité, Dynamiques de systèmes agrégés, Allocation de l'espace, Allocation dynamique de voies réservées aux bus, Optimisation, Temps de parcours de l'ensemble des usagers, Gestion du stationnement, Véhicules en recherche de stationnement, Tarification du stationnement, Tarification liée aux encombrements, Contrôle par rétroaction, Modèle multi-agents.

Contents

Abstract I

Résumé V

List of Figures VI

List of Tables X

1 Introduction 1

- 1.1 Research Background and Motivation 2
  - 1.1.1 Space and Mobility 2
  - 1.1.2 Parking and Mobility 4
  - 1.1.3 Pricing and Mobility 5
  - 1.1.4 Congestion Dynamics in Urban Road Networks 7
- 1.2 Research Objectives 12
- 1.3 Research Contributions 14
- 1.4 Thesis Outline 16

2 A Multi-modal Multi-region Modeling Framework for Road Space Allocation 23

- 2.1 Bi-modal MFD Model and System Dynamics 26
- 2.2 Road Space Allocation Management 34
- 2.3 Case Study and Analysis 35
  - 2.3.1 Static space allocation 36
  - 2.3.2 Dynamic space allocation 38
  - 2.3.3 Dynamic space allocation with pricing 41
  - 2.3.4 Sensitivity analysis 43
- 2.4 Summary 46

3 Modeling and Optimizing Multimodal Urban Network with Limited Parking and Pricing. 49

- 3.1 System Dynamics with Parking Consideration 52
- 3.2 Parking Management 57
- 3.3 Case Study and Analysis 59
  - 3.3.1 System dynamics with parking pricing 60
  - 3.3.2 Pricing scheme evaluation 63
  - 3.3.3 Parking policy indication 64
- 3.4 Summary 66

4 A Three-dimensional Macroscopic Fundamental Diagram for Mixed Bi-modal Networks 69
4.1 A City-scale Three-dimensional MFD for Bi-modal Traffic 71
4.1.1 A three-dimensional MFD for vehicle flows (3D-*v*MFD) 71
4.1.2 A three-dimensional MFD for passenger flows (3D-*p*MFD) 73
4.1.3 A functional form of the 3D-MFDs 75
4.1.4 Bus-Car Unit (BCU) Equivalent 77
4.1.5 An MFD for buses operating in dedicated lanes 79
4.2 An Analytical Passenger Flows Three-dimensional MFD 81
4.3 Network Partitioning for Mixed Bi-modal Network 82
4.3.1 Homogeneity in mixed bi-modal networks 82
4.3.2 Vehicle 3D-*v*MFD after partitioning 85
4.3.3 Passenger 3D-*p*MFD after partitioning 86
4.4 Application of the 3D-MFD and Partition: Real-time Perimeter Flow Control 89
4.4.1 Single-region control 90
4.4.2 Two-region control 94
4.5 Summary 96
5 Congestion Pricing Schemes combining Macroscopic Fundamental Diagram and Agent-based Approach 99
5.1 The Multi-agent Based Traffic Simulator 100
5.2 The Existence of the MFD in the Agent-based Simulation 101
5.2.1 The MFD 101
5.2.2 An explanation of scatter in the MFD 102
5.2.3 The Fundamental Diagram (FD) and spill-back effect 103
5.2.4 The MFD-controlled pricing schemes 104
5.3 A Feedback-type Pricing scheme 105
5.3.1 The pricing controller 105
5.3.2 Case study setup 107
5.3.3 Performance of the pricing scheme 108
5.3.4 Behavioral changes 110
5.4 A Feedback-type Pricing Scheme with User Adaptation 112
5.4.1 The pricing controller 112
5.4.2 Case study setup 113
5.4.3 Performance of the pricing scheme 114

5.4.4 PT benefit and accessibility improvement 116

5.4.5 User adaptation.. 118

5.5 Impact of User Heterogeneity.. 119

5.5.1 Behavioral difference 119

5.5.2 Distributional effect under different pricing strategies 121

5.6 Summary.. 122

6 Conclusions.. 125

6.1 Thesis Summary 125

6.2 Research Contributions.. 127

6.3 Future Work.. 129

Reference... 133

Curriculum Vitae.. 143

List of Figures

Fig. 1.1Empirical evidence on the existence of the MFD relating network trip completion and the space mean density. 9
Fig. 1.2 Effect of spatial heterogeneity of congestion in the MFD. 10
Fig. 2.1 A multi-region, multimodal urban road system. 24
Fig. 2.2 System dynamics flow chart. 27
Fig. 2.3 The MFD of mixed car-bus network, the MFD of dedicated bus lanes and speed profile for mixed network. 32
Fig. 2.4 Illustration of the case study two-region network sturcture and properties. 36
Fig. 2.5 An illustration of the performance of the system with two cases of space allocation to buses 38
Fig. 2.6 System performance under different space allocation of bus lanes. 38
Fig. 2.7 10 solutions by the SQP optimization algorithm for different starting points. 39
Fig. 2.8 The optimal solution of a time-dependent bus lane space allocation strategy (pi: π). 40
Fig. 2.9 Comparison of static and dynamic space allocation 41
Fig. 2.10 Comparison of the dynamic allocation with and without pricing. 43
Fig. 2.11 Comparison of system performances under different pricing scenarios for different levels of demand increase. 45
Fig. 3.1 Dynamics of a bi-modal transport system with parking choice, with indication of principal variables. 52
Fig. 3.2 Flow movements in region network i with parking choices. 54
Fig. 3.3 System dynamics and traffic performance under optimal constant pricing. 61
Fig. 3.4 System dyanmics and traffic performance under the optimal Strategy P1 62
Fig. 3.5 System dynamics and traffic performance under the otpimal Strategy P2. 62
Fig. 3.6 An illustation of two types of competition between the parking facility operators... 66
Fig. 4.1 The 3D-*v*MFD for bi-modal traffic and its contour plot 73
Fig. 4.2 The 3D-*p*MFD for bi-modal passenger traffic, and the contour plot 74
Fig. 4.3 The exponential 3D surface plot of 3D-*v*MFD, 3D-*p*MFD. 77
Fig. 4.4 A 3D diagram relating accumulation of cars and buses with BCU. 78
Fig. 4.5 The MFD for buses the impact of dwell times on the shape of the MFD. 80
Fig. 4.6 Illustration of the partitioned network based on bi-modal heterogeneity. 85
Fig. 4.7 Comparing the time each partitioned region reaches congestion. 85

Fig. 4.8 The normalized 3D-MFD of the partitioned regions, the contour plots. 86
Fig. 4.9 Passenger MFD from the simulation, n_c vs. n_b, v_c vs. v_b, and the estimated passenger MFD for the two partitioned regions and the whole network. 87
Fig. 4.10 Comparison of n_c vs. n_b for the entire network, the measured P vs. the estimated P 88
Fig. 4.11 The composition rate δ for three representative scenarios, Ω =[0.02 0.15]............ 90
Fig. 4.12 Comparing overall traffic performances for 5 different scenarios........................... 91
Fig. 4.13 Comparing bus trajectories in several bus lines in the network during the heart of rush between pre-time control and perimeter control. 93
Fig. 4.14 Snapshot of Downtown San Francisco and clustering into two regions; red color indicates the city center; green color indicates the rest of the network, and illustration of the 3D-MFD 94
Fig. 4.15 Comparison of different performance indices for the two modes of traffic under single-region (1-region) and two-region controls (2-region). 95
Fig. 4.16 Cumulative distributions of TTKT for the two modes of traffic in the two regions under single-region and two-region controls. 96
Fig. 5.1 MFD of the center area network of Zurich 102
Fig. 5.2 Density distribution of the network at time 6h35, 9h15 and 11h40........................ 102
Fig. 5.3 Illustration of congestion propagation along consecutive links.............................. 103
Fig. 5.4 An illustration of the optimization algorithm of pricing.. 105
Fig. 5.5 The targeted cordon area for pricng the city of Zurich 108
Fig. 5.6 Comparison of traffic performance between no toll and the final toll..................... 109
Fig. 5.7 Comparison of link traffic in the no-toll scenario and the final toll scenario 109
Fig. 5.8 Time series of the flow of agents passing the cordon 111
Fig. 5.9 Time shift of activities by types, in the no toll and the final toll scenarios 112
Fig. 5.10 The MFD of the study site 114
Fig. 5.11 Density time series without and with optimal time-dependent pricing. 115
Fig. 5.12 The density series and toll rates of three simulations before reaching the optimal pricing.. 116
Fig. 5.13 Comparison of density series and resultant toll between “pricing only” and “pricing BA”... 118
Fig. 5.14 Equilibrium condition of the pricing schemes 119
Fig. 5.15 Comparison of mode choice, total toll paid and savings between two user groups. 120

Fig. 5.16 Comparison of mode shift, toll paid and savings for the two user groups among the three pricing strategies: pricing, pricing BA and pricing BAIp. 121

List of Tables

Table 2.1 Nomenclature of main variables and parameters in Chapter 2. 24

Table 2.2 Sensitivity analysis on β_1, β_2 and γ. 44

Table 3.1 Nomenclature of main variables and parameters used in Chapter 3. 50

Table 3.2 Comparison of the performances of different pricing schemes. 63

Table 4.1 Nomenclature of the main variables and parameters used in Chapter 4. 70

Table 5.1 Summary of the social of the proposed pricing 110

Table 5.2 Comparison of total and average travel time after pricing..................................... 116

Table 5.3 Welfare gain among the three pricing strategies.. 117

1 Introduction

With the rapid economic growth and social development, city centers are experiencing significant amount of daily commuting. Traffic congestion notoriously spreads in urban road networks. Management strategies ought to be developed and implemented, in order to maintain mobility levels and mitigate the negative impact generated by congestion on social welfare. These strategies should be efficient, equitable and sustainable. More importantly, they should be based on sound knowledge of congestion physics and practically applicable in the nowadays large-scale multimodal transport systems. This is a challenging task as the dynamics of urban transport systems are complex: (i) Detailed demand patterns (origin-destination table) with reasonable temporal-spatial resolution are almost impossible to obtain with a decent accuracy; (ii) disaggregated travel behavior (route and mode choices) involves unpredictable rationality and stochasticity, which is difficult to be integrated in a real-time traffic management; (iii) the capacity of individual road facilities varies by topology and signal settings, etc. Despite the presence of these complexities, recent advances in traffic flow theories encourage the transition from prediction-based approaches to monitoring-based ones. The latter approaches reveal the existence of well-defined laws of congestion dynamics at aggregated levels and they can be utilized with smart strategies to improve mobility in cities. The knowledge, however, is still limited for multimodal networks. As people travel through different modes of transport such as cars, buses, taxis that compete for scarce urban road infrastructure, it is critical to understand how this space can be allocated and managed. This thesis indeed attempts to make effort towards this direction. The objective is to develop aggregated modeling and optimization approaches, which will contribute on the knowledge of congestion dynamics in cities of different structures and mode usages, and ultimately facilitate the design of efficient traffic management policies, such as dedicated bus lane allocation, congestion pricing and parking limitation. In this first chapter of this thesis, we aim to review the existing literature on modeling and controlling multimodal urban transport system, and to identify the key issues that this thesis has to address.

1.1 Research Background and Motivation

Multimodality plays a critical role in mobility management. If not managed well, traffic congestion will be increasingly pervasive in urban areas. Constructing new infrastructure is an expensive solution, as extremely high cost is needed to keep pace with increase in demand and the induced demand (Small, 2004). Smart planning and allocation of the existing road space among different mode usage therefore is required. As people travel through various modes of transport compete for the limited urban road space, understanding mode conflict pattern, and the relation between space allocation and the resultant congestion, are important. By assigning space in priority to more efficient modes, the re-allocation of road space aims to change the mode share and decrease the car usage and congestion. In case the mode shift is insufficient to improve the mobility level, pricing strategies can facilitate this shift to the desired one. These pricings can be congestion-charge type, such as the cordon-based pricing in London and the area-based pricing in Singapore, but also through accessibility control such as parking limitation or bus service improvement which could possibly come from congestion toll collection. In the case of parking limitation, if the characteristics of the cruising-for-parking phenomenon can be captured, strategies such as on-street parking pricing (where availability significantly influences the extra travel delay) can be suitable policy in congestion alleviation. Sub-sections 1.1.1-1.1.3 will investigate multimodal mobility in literature from three dimensions: road space, parking limitation and congestion pricing. To develop sustainable multimodal networks, congestion dynamics should not be ignored. After all, the goal is to reduce congestion and the underlying approach of the management strategies should be consistent with the physics of congestion. To this end, a comprehensive review of the recent development on traffic flow theories is provided in sub-section 1.1.4.

1.1.1 Space and Mobility

Road space is an important factor in influencing mobility. Under a multimodal environment, space should be allocated taking into account spatiotemporal differences in the demand, the topology, the mode usage and the control characteristics. If these spatiotemporal decisions are made incorrectly, space could be wasted and mobility is being restricted as this wasted space could be used by more productive modes, e.g. modes with high passenger occupancy. The importance of multimodality by considering passenger throughput rather than vehicle throughput was earlier mentioned by Vuchic (1981) who suggested that evaluations of multimodal systems should integrate the occupancies of each mode. Empirical studies in

Californian freeways showed accordance with this argument as well. For example, Chen et al. (2005) questioned the effectiveness of high-occupancy vehicle (HOV) lanes and have shown that HOV lanes are underutilized and the passenger capacity of freeways has decreased, resulting in heavier congestion levels. Other studies on this discussion include Sparks and May (1971) who evaluated priority lanes for high occupancy vehicles on freeways by a mathematical model; Dahlgren (1998) and Daganzo and Cassidy (2008) studied how different modes use freeways, recognizing that different modes serve different numbers of passengers and should not be considered all equally. Most of these works are limited to small-scale systems as they looked at the effect on total passenger travel time if a lane on a specific road section was dedicated to multiple occupant vehicles. Nevertheless, the connection between road space for different occupancies vehicles and performance is highlighted as the passenger capacity of the modes is recognized.

A body of researches also looked at how street space (through HOVs or dedicated bus lanes) should be allocated to reduce vehicular congestion, such as Radwan and Benevelli (1983) and Black at al. (1992). These early research explorations towards space and mobility have limited applicability, however, since they assume steady-state traffic conditions and they ignore the dynamic fluctuations and spill-over effects that typically characterize urban traffic congestion. Currie et al. (2004) carried out impact analysis in planning studies of road space allocation, via a disaggregate micro-simulation which relies on intensive travel data inputs, such as time-dependent origin-destination tables and traffic supply curves (flow-density relationship) that are typically unavailable or unreliable for urban networks. In simple-structured and uncongested road networks (as of the time period those studies were done), these treatments may work sufficiently. However the impact of space on congestion is unfortunately more complicated nowadays. Other works including Pushkarev and Zupan (1977), Pickrell (1985), Kenworthy and Laube (2001) and Cameron et al. (2003) focused on the connection between performance of transport systems and land use for transportation as a whole, without further consideration on allocation by specific mode usage. Recent researches carried out qualitative analysis on space allocation, see in Gonzales et al (2010) and Gonzales and Daganzo (2012). The former work proposed dedicated bus lanes for the city of Nairobi, and tested the performance at intersection level via micro-simulation. The latter work recognized that the network capacity for cars is reduced if transit operations receive dedicated space. However neither quantified the impact of this dedicated space on the global performance.

Though the existing studies have made many effort in addressing space and multimodality, the limitations lie in application level and consideration of congestion dynamics. Furthermore, scenario analysis by micro-simulation and agent-based models require huge effort to be calibrated and cannot be utilized in a real-time framework. To fill in these research gaps, an approach should be developed which is suitable for large-scale road space allocation with proper treatment of traffic dynamics. Detailed research proposal will be provided in Sections 1.2 to 1.4.

1.1.2 Parking and Mobility

Parking limitation is another space-related issue in urban cities. The accessibility and cost of parking significantly influence people’s travel behavior, such as mode choice and facility choice (on-street or garage parking). Parking thus has direct influence on travel cost. Car-users may have to cruise for on-street parking space before reaching destinations and cause delays eventually to everyone, even users with destinations outside the limited parking areas.

Many studies have showed the impact of parking on mobility. One of the highly focused field addresses how parking limitation and parking policies influence people’s mode choice and the resultant behavioral equilibrium. This type of literature can be found for example in Arnott, and Rowse (1999), Anderson and de Palma (2004), Arnott (2006), Forsgerau and de Palma (2013), Qian et al. (2013) and elsewhere. By utilizing and extending the classical bottleneck-model of Vickrey (1969), behavioral adaptation given parking constraints is revealed. However static bottleneck capacity and redundant parking availability were assumed in these studies. In case of non-steady conditions, such as when congestion and the demand for parking are time-varying, the system dynamics become much more complicated; while in case of insufficient parking space, cruise-for-parking flows are different than the normal running flows and can lead to traffic congestion. Unfortunately the parking-economics focused approaches cannot capture those dynamics.

Studies taking the impact of cruising into account can be found in Shoup (2005), Arnott and Rorwse (2009), Arnott et al. (2013) for analyzing parking pricing policies;, in Martens et al. (2010) for examining spatial effect on searching-for-parking; and in Horni et al. (2013) where parking choice and searching-for-parking were incorporated in an agent-based model with simplistic traffic modeling. However, the inadequacy still lays in the treatment of traffic dynamics, particular the dynamics of the cruising flow. For example, trip length may change significantly as vehicles are cruising for parking. As more vehicles are cruising, the probability of finding a space decreases and eventually the trip length in the destination area increases. It is expected that the increase of trip length would lead to the increase of vehicle density, and thus a decrease in average speed and flow of the whole area. No work has been done quantitatively towards this direction with proper treatment of cruising and congestion.

Other studies attempting to study cruising can be found below. Gallo et al. (2011) incorporated parking cost in their traffic assignment model. Van Ommeren et al. (2012) carried out an empirical study on the diverse features of cruising time in the Netherlands. Guo and Gao (2012) estimated travel time including the delay from searching for on-street parking space. The proposed approaches either require data that are difficult to collect (e.g. detailed information of origin-destination and parking availability at destination) or they become computationally expensive in large-scale applications (e.g. microscopic traffic or agent-based models). Several works analyzed delay caused by on-street parking at intersections with specific layout, see for example in Yousif and Purnawan (2004) and Cao et al. (2014), which deal with single-lane downstream section of an intersection. Unfortunately their approach cannot be readily applied on a large-scale network.

To influence urban mobility via parking policies, the dynamics of cruising and its impact on congestion deserve further research attention. Similar modeling approach as proposed for road space and allocation can be applied with the integration of parking limitation.

1.1.3 Pricing and Mobility

Provided with well-designed space allocation between different models, urban networks may still be exposed to congestion and rarely sustain at the designed vehicle or passenger capacity. The reason is that travelers do not always make rational decisions rather than to maximize their own utility when making choices. Pricing therefore has been applied in many cities, with the goal of triggering rational travel behavioral changes. Charged by the external costs of facility usage, demand is redistributed in a way that the average user cost is optimized.

There exists an extensive body of studies on congestion pricing models. A comprehensive literature summary of these models can be found in Yang and Huang (2005). The theoretical background of pricing has relied on the fundamental concept of marginal cost, first introduced by Pigou (1920) and followed by Vickrey (1963) and other researchers. In the traffic assignment literature, tolls of this type belong to the first-best pricing category and have been proposed to drive a user equilibrium pattern (Wardrop, 1952) towards a system optimum. Despite their idealized theoretical basis, first-best pricing models have been impractical and difficult to implement. Merchand (1968) investigated second-best tolls using a general equilibrium model. According to the second-best pricing models, e.g. Arnott et al. (1990), Small and Yan (2000), Verhoef (2002), tolls are charged in a subset of selected links where the bottlenecks are located. Utilizing the classical bottleneck equilibrium approach (Vickrey, 1969) for pricing has received as much attention and effort as for the parking-related studies, such as Arnott (2007), Wu et al. (2011), Verhoef (2012) etc. The critical inadequacy in these works, which was pointed out by Geroliminis and Levinson (2009), is that the Pigouvian-type pricing assume a network supply curve (desired or input demand vs. average travel cost) which is not consistent with the physics of traffic. The total cost expressed in delay terms for a given desired demand over a period of time is sensitive, during congested conditions, to small variations of flow within the given period and depends on the initial state of the system and the level of congestion. In other words, such supply-demand relation is memory-less and does not consider the dynamics and evolution pattern of traffic congestion, leading to inefficient pricing implementations.

The second issue of pricing is that charging individual links cannot guarantee (improved) optimum travel condition and is also difficult to implement, while pricing at large aggregated scales (e.g. cordon- and area-based pricing) lacks a base that is theoretically sound and practically applicable. Nevertheless, pricing schemes of aggregated links and networks were developed and applied in different cities, such as cordon-based pricing in London and Stockholm, area-based pricing in Singapore, and elsewhere. Many studies discussed the resultant improvements in travel condition and the political, equity and other relevant concerns of these real-life implementations. However, it is unfortunate that few studies explain or aware the underlying mechanisms for pricing determination, except the general principle of Singapore’s pricing was briefly explained in Liu et al. (2013), where the time-dependent cordon-pricing scheme was adjusted based on regular survey on traveling speed. On the theoretical side, Anderson and Mohring (1997) examined congestion on the Twin Cities road network by charging drivers marginal costs and using a user equilibrium assignment for a single period. Yang and Huang (2005) examined the principle of marginal-cost pricing in a road network. Maruyama and Sumalee (2007) compared the performance of cordon- and area-road pricing schemes regarding their efficiency and equity via simulation. Liu et al. (2013) proposed a speed-based toll similar to the one applied in Singapore. The ambiguity in most of these models, once more, falls on that traffic conditions are considered either stationary or using the traditional demand-supply curve for congestion modeling. The estimated congestion toll based on idealized or memory-less traffic models may not be optimal and the system may be either still congested if underpriced or very uncongested if overpriced.

To develop efficient pricing schemes, especially for large-scale urban networks, an approach must capture the aggregated dynamics of congestion towards pricing. Furthermore it is crucial to consider heterogeneity (e.g. distribution of VOTs among users) and users' behavioral adaptation (e.g. change of travel behavior in response to pricing). Few studies provide quantitative approach to evaluate the effect of pricing on multimodality. Questions such as how to adjust pricing rate based on user's behavioral changes or what are the impacts of an incentive program of using public transport (PT) on the performance of pricing should be addressed. It has to be highlighted that despite the vast literature in pricing, field tests are quite limited and this is among other reasons due to user acceptability. Providing incentives to other modes of transport (e.g. return a fraction of the tolls paid to users that switched to public transport mode) can make such policies more attractive for real cases.

1.1.4 Congestion Dynamics in Urban Road Networks

The main challenge in mobility management of large-scale multimodal urban networks is to develop a multimodal dynamic modeling framework at the network level. This framework should capture congestion dynamics of the multimodal system and links the performance of road networks to road space allocation, parking limitation, pricing and other policies. In this section we provide an overview of the literature on traffic flow models, shedding light on the research directions of this thesis.

There is a strong understanding and vast literature of congestion dynamics and spreading in car-to-car-scale or road-scale traffic systems with a single mode of traffic, e.g. a highway or urban street section with cars. Traffic scientists, mathematicians and physicists have contributed to the field of traffic flow theories from many aspects. Because a detailed review of numerous publications is beyond the scope of this thesis, readers are advised to refer to Helbing (2001) for an overview. Briefly speaking, the main modeling approaches can be classified as follows: (i) Car-following models deal with the non-linear interactions and dynamics of individual vehicles (e.g. Gipps, 1981). To address computational burden, (ii) cellular automata describe the dynamics of vehicles in a coarse-grained way by discretizing space and time, (e.g. Nagel and Schreckenberg, 1992; Daganzo, 1994). (iii) Gas-kinetic models formulate a partial differential equation for the spatio-temporal evolution of the vehicle density and the velocity distribution (e.g. Helbing and Treiber, 1998). (iv) First-order flow models are based on a partial differential equation for the density and a fundamental diagram relation (e.g. Lebacque, 1996; Leclercq et al, 2007). (v) Second-order models contain an additional equation for non-steady state conditions. In continuum models, a network is approximated as a continuum in which users choose their routes in a two-dimensional space (e.g. Jiang et al., 2011). It has been broadly shown through simulation and field experiments, e.g., Munoz and Daganzo (2003), Geroliminis and Daganzo (2008), Helbing et al. (2009), that the classical linkage between pertinent variables flow, speed and density on a spatially disaggregated level (e.g. one link) is very scattered and does not follow a well-defined curve. One of the reasons is that traffic systems are not in steady-state conditions at a link level, particularly in urban networks where traffic signals, turning movements and spill-overs are widely present and intervene with traffic. Management strategies based on these models cannot guarantee efficiency and are also difficult to be implemented in real time due to computational complexity.

On the other hand, models based on aggregated traffic dynamics appear to reflect traffic performance and congestion evolution at network level and could overcome the computational burden if they can highlight the main important characteristics of dynamic congestion. The idea of an aggregated traffic flow model for car-only urban networks was initially proposed in Godfrey (1969) and was re-initiated later in Mahmassani et al. (1987) and Daganzo (2007). The demonstration of the existence and dynamic features of a Macroscopic Fundamental Diagram, the MFD, was only recent in Geroliminis and Daganzo (2008), with empirical data from Yokohama, Japan. According to the definition, an MFD relates the production (total distance travelled by all vehicles per unit time or trip completion rate) or space mean speed (as defined by Edie (1963)) to vehicle accumulation (total amount of vehicles) of a network. The same work by Geroliminis and Daganzo (2008) showed that the MFD is a property of the network itself (infrastructure and control) if the network is homogeneously congested and not very sensitive to demand, i.e. the space-mean flow is maximum for the same value of critical density of vehicles, for many origin-destination tables. Fig. 1.1(a) displays the MFD observed from field data of the city center network in Yokohama. Each point in the MFD plot corresponds to a 5-min aggregation (space-mean) of density and network circulating flow data collected in different days and at different times. Remarkably, the scatters follow closely a well-defined curve, although the same relationship does not hold for individual links, as shown in Fig. 1.1(b) the flow-density relationship on two links in the same network.

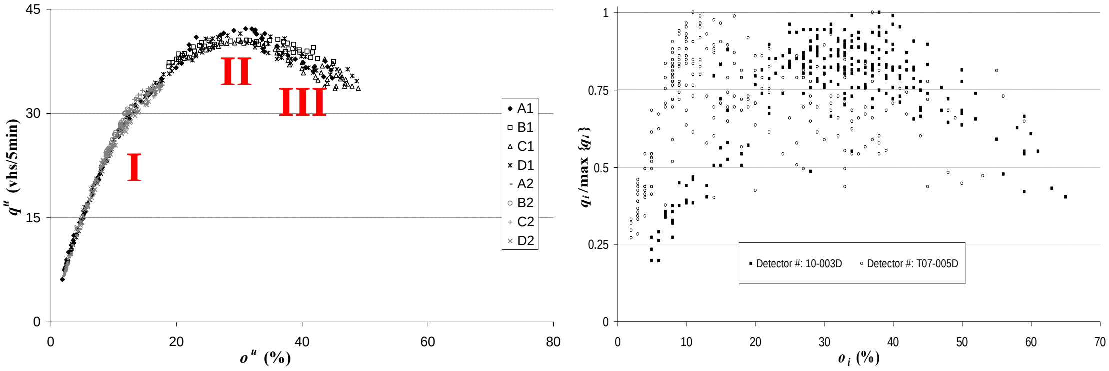

Fig. 1.1(a) Empirical evidence on the existence of the MFD relating network circulating flow and the space mean density, Yokohama city center network in Japan, and (b) fundamental diagram of flow-density relation on two individual links in the same network. Data source: Geroliminis and Daganzo (2008).

An MFD facilitates performance monitoring at network-level. A graphical explanation is added in Fig. 1.1(a). In the figure, three traffic regimes can be observed in the MFD scatter plot. Regime I corresponds to under-saturated states where most of the links in the network sufficiently serve vehicles with an actual throughput smaller than the maximum possible. Regime II represents saturated states. In this regime, most of the links are filled with permanent queues and many operate at capacity. This capacity is approximately constant, but never larger than the quantity $\sum_i L_i g_i s_i$, where L_i is the link length, g_i is the duration of green phase and s_i is the saturation flow of link i. Furthermore, the critical density K_{cr} that maximizes flow is achieved in Regime II. While in Regime III, production decreases with more vehicles joining the network (oversaturated states) and long queues or spillbacks are observed in many links. It can be seen that an MFD provides explicitly the optimal traffic state of the networks, which is to operate closely to the critical density in Regime II. The goal of an MFD-based management strategy would be to avoid states in Regime III instead to maintain in Regime II or not exceeding K_{cr}, as much as possible. These properties can be of great importance to unveil management policies in such a way that maximize the network capacity of vehicles or even the passenger capacity in multimodal networks, if similar patterns can be observed.

Recent works highlighted that the spatial distribution of congestion can affect the shape of the MFD with higher flows observed for less spatial heterogeneity, e.g. Mazoumian et al. (2010), Geroliminis and Sun (2011), Daganzo et al. (2011), Mahmassani et al. (2013), and Knoop et al. (2013) explored the properties of the MFD via simulation analysis. Analytical theories have been derived for estimating the MFD as a function of network and intersection parameters, using variational theory for homogeneous and heterogeneous network topologies, respectively (Daganzo and Geroliminis (2008), Geroliminis and Boyaci (2012), Leclercq and Geroliminis (2013)). The conditions of a well-defined MFD, stability and scatter phenomenon are analyzed through many other simulation studies and experimental tests, such as in Buisson and Ladier (2009), Ji et al. (2010), Gayah and Daganzo (2011), Saberi and Mahmassani (2012), Ji and Geroliminis (2012), and Mahmassani et al. (2013).

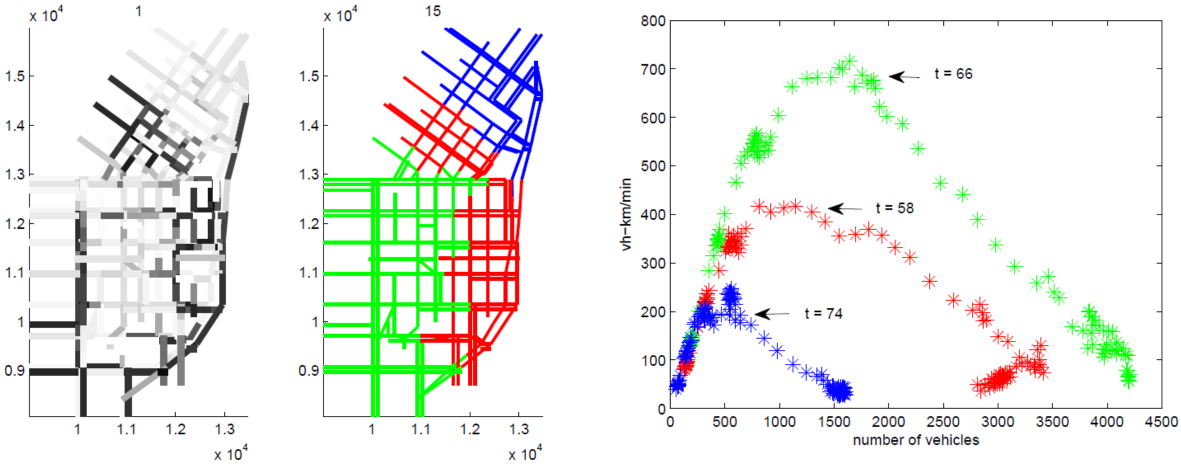

Fig. 1.2 Effect of spatial heterogeneity of congestion in the MFD: (a-left) Snapshot of the San Francisco network during peak hour (colors indicate density level, the darker the more congested), (b-middle) the network partitioned in 3 regions, and (c-right) MFDs of the 3 partitioned regions, with the congestion starting times indicated. Data source: Ji and Geroliminis (2012).

It should be further emphasized that the MFD of freeway road networks likely exhibit hysteresis and capacity drop phenomenon. Geroliminis and Sun (2011) concluded that this phenomenon mainly is caused by heterogeneity in congestion distribution and they observed different congestion centers exist in the studied network. This property should not be neglected (both for urban and freeway networks) as each congestion center may reach the congested regime at different times. Control strategies based on a global MFD for the whole city thus may not be effective as they treat regions with different level of congestion in a similar manner. A recent study by Ji and Geroliminis (2012) presented a partitioning algorithm to identify regions with different congestion levels. Fig. 1.2 displays an application of partitioning for a simulation of the San Francisco network. Fig. 1.2(b) shows the result of the partitioning algorithm, given congestion distribution over the network illustrated in Fig. 1.2(a) (the darker the more congested), while Fig. 1.2(c) plots the MFDs of the three partitioned regions. It can be observed that there is a clear distinction between congested and uncongested regimes for all three regions. The time each of the regions reaches the congested regime is very different.

With successful partitioning, network-level monitoring and optimization would be a feasible approach, as congestion is homogeneous inside each partitioned region and all the regions have a well-defined MFD with low scatter.

Once the MFD is obtained and given, real-time traffic performance can be monitored with input data that can be easily collected via current communication technologies, such as GPS, Bluetooth, loop detectors etc. Management strategies then can be developed, examples including MFD-based perimeter flow control which can be found in Keyvan-Ekbatani et al. (2012), Geroliminis et al. (2013), Aboudolas and Geroliminis (2013), and Haddad et al. (2013).

As multiple modes compete for limited urban space, conflicts and interactions are developed resulting in congestion. Congestion dynamics and mode conflict patterns of multimodal urban networks need to be analyzed. Existing literature on multimodal traffic mainly focuses on design and operation of special lanes (Daganzo and Cassidy (2008), Li et al. (2010), Tirachini and Hensher (2011)), optimization of signal control (Mesbah and Currie (2011), Christofa et al. (2013)). However, few of these works were dedicated on the modeling of traffic dynamics and even fewer on the influence of each mode on the network performance. Furthermore, most of the aforementioned works fall short either in the scale of application or the treatment of congestion dynamics (small scale and/or static models). For example, Li et al. (2010) and Mesbah et al. (2011) are based on a BPR ((Bureau of Public Roads, 1964) type of model, which works only for static conditions. Daganzo and Cassidy (2010) and Tuerprasert and Aswakul (2010) utilized the link-scale Fundamental Diagram, which can experience high scatter and therefore cannot provide accurate estimates of speed and travel time at the network level. These somehow detailed models can also be computationally complex when applied at the network scale.

Mobility of multimodal urban networks needs to be analyzed under various network structures, as well. A multimodal network can be treated as a system of multiple regions, where each region exhibits similar congestion patterns. This treatment is plausible thanks to the development of the MFD theory, as the dynamics in each region can be represented by an MFD. However, capturing such multimodal system dynamics is still challenging. This is because the MFD of a multimodal network, if exists, is expected to be influenced by space allocation and mode usage. Traditional models based on detailed OD information and traffic assignment models, or small-scale traffic flow models, can be computationally costly when applied for large-scale networks and require data that are difficult to obtain. With the respect to this issue, an MFD-type based system model is suitable and can serve as a strong alternative for multimodal network modeling and control. Recent works towards this direction can be found in Gonzales and Daganzo (2012) (2013), who examined system optimum solutions using MFD for a transport system with cars and public transit share space for the peak-hour commune problems.

Furthermore, the impact of operational and service characteristics and interaction between different modes when sharing the same space should not be ignored. While literature has focused on congestion modeling for vehicular traffic, the effect of service stops (like when a bus or taxi stops to board and alight passengers) in the overall performance of a large-scale urban system requires further research effort. It is intuitive that the effect of these stops in the network capacity is almost negligible during light demand conditions, but nowadays city centers are experiencing high level of congestion and the frequency in time and space of these stops is significantly high. One bus creates more congestion than a moving car. On the other hand as mentioned in Section 1.1, the additional delay per passenger carrying is smaller for large bus occupancies, making buses a more sustainable mode of transport compared to cars. A semi-analytical approach building on variational theory was developed to estimate an MFD for networks with cars and buses (Boyaci and Geroliminis, 2011). Buses service stops were treated as periodic red phases of fictitious traffic signals that decrease the capacity of the road at the bus stop locations. While such a framework can provide important insights for network capacity for different operational characteristics of buses, it cannot directly integrate the heterogeneity in the spatial distribution of buses and cars in the network.

The promising tool to advance the understanding of multimodal traffic dynamics is the extension of an MFD for car-only urban traffic to a multimodal framework. If an MFD-type model exists for multimodal networks and it unveils similar pattern as observed in the MFD, system model can be developed to produce the aggregated behavior and be utilized to design and test management strategies.

1.2 Research Objectives

To address the aforementioned research gaps, this thesis aims to develop a macroscopic modeling and optimization approach for influencing mobility and multimodality of urban networks. The approach shall contribute on the knowledge of congestion dynamics in cities of different structures and mode usages, and ultimately facilitate the design of effective and efficient urban transport policies, such as dedicated bus lane allocation, congestion pricing and parking limitation. To this end, four objectives are identified.

Objective I: To develop a network-level dynamic traffic flow model for optimizing road space allocation in multimodal urban networks. In order to gain insightful understanding of the complex multimodal dynamics, bi-modal dynamics of cars and buses should be studied as a first step. The developed bi-modal traffic model needs to capture congestion dynamics, e.g. the effect of bus operation (such as dwell time of service stops) in the global performance. The model should also link traffic performances to space allocation between cars and buses. Based on the bi-modal traffic model, a system model needs to be developed to reproduce the system dynamics under given demand, city structure and road infrastructure. With such system model, an optimization framework can be built up for designing space allocation strategies.

Objective II: To develop a system dynamic model for designing parking polies in multimodal urban networks with limited parking. Given the achievement of Objective I, the focus moves on the integration of cruising-for-parking in the bi-modal traffic model, such that the change of system dynamics due to parking limitation and cruising can be reflected, and the cost of cruising can be estimated e.g. the increase of average cruising delay or the reduction in average speed. Incorporating this parsimonious model to the system model developed in Objective I, the ultimate objective is to design optimal parking pricing strategies through optimization.

Objective III: To develop a new type of network-level traffic flow model, which captures the congestion dynamics by modes, the modal conflict patterns, and passenger flow dynamics. Completing objectives I and II, one obtains a modeling approach for mixed-traffic networks where the joint effect of all vehicles in congestion can be evaluated. The key objective now is to propose a model that can monitor and quantify the impact of each individual mode on the global traffic performance in multimodal networks, for example the proposed model should be able to analyze the network vehicle and passenger flow throughput at different mode composition rates. This modeling tool is also expected to identify mobility levels with respect to passenger flow under different presence of buses, so that passenger-dynamics-based analysis can be carried out.

Objective IV: To develop network-level area-based and cordon-based congestion pricing schemes for multimodal networks based on the concept of the MFD. The proposed pricing schemes aim at controlling congestion for an entire network, thus the underlying pricing control mechanism must be consistent with congestion physics at aggregated scales. To ensure the efficiency of the pricing schemes in congestion management, traveler’s behavioral adaptation ought to be taken into account for toll rates adjustments. Furthermore, the promotion of public transport service through redistribution of the toll revenue should be treated, as the goal is to increase multimodality. The distributional effect of benefit and loss over heterogeneous users under such pricing schemes should also be investigated.

1.3 Research Contributions

The contributions of this thesis can be summarized in four categories, which correspond to the four objectives defined in Section 1.2. By achieving these objectives, this thesis succeeds in developing a macroscopic tool for understanding the congestion dynamics in multimodal networks and developing strategies for mobility and multimodality management.

A Multi-modal Multi-region Modeling Framework for Road Space Allocation (Chapter 2)

- A bi-modal Macroscopic Fundamental Diagram model (bi-modal MFD) is developed for capturing congestion dynamics of a mixed-traffic network of cars and buses, relating the network production to the network space mean accumulation of the mixed-traffic. Space allocation affects the shape of the bi-modal MFD.
- An aggregated system dynamic model is proposed for reproducing the aggregated flow movements and (time-dependent) travel cost in a bi-modal multi-region transport system for given city structure and road infrastructure.
- As traffic performance of the system is linked to space allocation, an optimization framework is established to determine space allocation and mode usage strategies that minimize the total cost of all the travelers. Two allocation strategies are obtained: a static and a dynamic dedicated bus lane allocations.
- The developed modeling and optimization framework can be utilized for real-life application, as the required data that are feasible to collect. It is also demonstrated that this framework is not very sensitive to the changes on the values of the main parameters.

Modeling and Optimizing Multimodal Urban Network with Limited Parking and Pricing (Chapter 0)

- The system model developed in Chapter 2 is extended with parking limitation. The dynamics of cruising are integrated in the system dynamic model.
- A parsimonious model is proposed to estimate cruising delay and a dynamic average trip length, assuming the average probability of finding a parking space follows a geometric distribution and depends on the time-dependent parking space availability.
- The cruising effect is integrated in the bi-modal MFD model as well, where delay to all car-users including those who do not participate in cruising for parking and those travel outside of the congestion region, can be estimated.
- Close-to-optimum pricing schemes can be obtained through the optimization framework, such as a real-time parking pricing control scheme, where the price of parking depends on congestion level and cruising delay. Furthermore, the model can be applied to investigate competition behavior of the parking market, if the operation authority of different types of parking facilities belongs to different parties.

A Three-dimensional Macroscopic Fundamental Diagram for Mixed Bi-modal Traffic (Chapter 4)

- The existence of a three-dimensional Macroscopic Fundamental Diagram (3D-MFD) for bi-modal mixed urban networks is demonstrated through micro-simulation, relating the total network production to the accumulation of cars and buses if a region is (roughly) uniformly congested. With the 3D-MFD, the individual modal impact on the global traffic performance can be directly identified and quantified.
- A passenger 3D-MFD is proposed and derived where the passenger production over the accumulation of cars and buses can be observed and it provides a different perspective of traffic dynamics of bi-modal networks.
- Network partition is applied to obtain the 3D-MFD for networks with heterogeneity in mode composition and congestion level. Recognizing this heterogeneity improves the accuracy of the 3D-MFD modeling.
- An exponential model is proposed for the functional form of the 3D-MFD. With the function form, the Bus-Car Unit equivalent value is derived and found to be state- and mode composition-dependent instead of a constant value as is usually assumed in literature.
- The 3D-MFD can be significantly useful for monitoring and managing congestion in bi-modal networks, e.g. a 3D-MFD based perimeter flow control can achieve high

efficiency in congestion reduction and equitable benefit between users of different modes and in different parts of the network.

Congestion Pricing Schemes combining Macroscopic Fundamental Diagram and Agent-based Approach (Chapter 5)

- An approach combining the MFD and agent-based model is proposed to design city-level congestion pricing schemes.
- A cordon-based pricing is developed via a Proportional feedback-type control mechanism, where the pricings are determined and adjusted based on aggregated congestion level, as expressed by the MFD.
- An area-based pricing is developed via a Proportional-Integral feedback-type control mechanism, where the pricings are determined and adjusted based on aggregated congestion level and behavioral adaptation, expressed by the MFD as well.
- Pricing is found more efficient: when behavioral adaptation is considered; when incentive programs are integrated in the pricing schemes through redistributing the toll revenue to promote bus usages; and when heterogeneity of travelers such as difference in value-of-time is recognized and treated separately.
- MFD-based pricing strategies are practical for implementation, given their reasonable data input requirements and implementation costs.

1.4 Thesis Outline

This thesis consists of six chapters. The main four chapters, Chapters 2 to 5 intend to address the four stated objectives respectively. The outline of each chapter is given below.

Chapter 2 extends the single-mode MFD model to a multi-modal one with focus on bi-modal case, and constructs a dynamic system model. A city network is decomposed to multiple regions. Each region has its unique mode specific MFD or bi-modal MFD which is assumed known in advance, given space allocation condition (the shape of this MFD is investigated later in Chapter 4). The flow dynamics among regions are described by a regional level flow conservation law, where vehicles are categorized by families regarding their traveling directions and the movement throughput and speed is represented by the bi-modal MFD. An approximate dynamic mode choice (that can be integrated in an optimization framework) is also given to passengers to choose modes based on imperfect information of the time-dependent cost, where the cost is identified from the MFD. Combining the bi-modal MFD, the mode choice model and the aggregated flow conservation model, a non-linear optimization framework is proposed to allocate road space between cars and buses, such as a time-dependent dedicated bus lane allocation strategy in the congested center, given the total available road space of a city. A case study is carried in a hypothetical two-region network to prove the rationale of the aggregated system model. Two space allocation strategies are introduced and optimized with the objective of minimizing total cost of all travelers. Furthermore, a sensitivity analysis on the key parameters of the model is carried out. Performance under high demand scenarios is also investigated. The preliminary results of this chapter are presented in:

- Optimizing space allocation for multimodal transport system: A macroscopic approach. Zheng, N. and Geroliminis, N., 2012. Paper presented at the 12th Swiss Transportation Research Conference (STRC). Ascona, Switzerland.
- Re-distribution of urban road space for multimodal congested networks with MFD representation. Zheng, N. and Geroliminis, N., 2012. Paper presented at the 92nd Annual Meeting of Transportation Research Board. Washington D.C., USA.

The most important findings of chapter 2 are published in:

- On the distribution of road space for urban multimodal congested networks. Zheng, N. and Geroliminis, N., 2013. Transportation Research Part B 57: 326-341. The paper was also presented at the 20th International Symposium on Transportation and Traffic Theory (ISTTT), Noordwijk, The Netherlands, 2013.

Chapter 3 integrates parking limitations into the dynamic model and the optimization framework developed in Chapter 2. The system model is enhanced by incorporating a cruising-for-parking vehicle family, where vehicles before reaching their destination should cruise due to the limitation of parking space. The cost of cruising is estimated by a parsimonious model where the probability of finding a parking space is assumed to follow a geometric distribution and depend on the time-dependent parking availability. While the total system delay due to cruising estimated by the bi-modal MFD for the whole network. It also shows the effect of cruising on users who even do not require parking search. The impact of parking on the performance of the system is analyzed utilizing the same idealized network. Two parking facilities are available: on-street parking with limited space requiring cruising, and garage parking with infinite capacity requiring higher parking fee. Two parking pricing strategies are designed, optimized and tested. Furthermore, the developed system model is utilized to investigate parking market competition between on-street and garage parking. A bi-objective and a bi-level optimization framework are formulated for cooperative and non-cooperative price competition, respectively. The preliminary results of this chapter are presented in the following conferences, while a journal article is under preparation:

- A dynamic approach for influencing multimodal mobility of buses and cars with limited parking. Zheng, N. and Geroliminis, N., 2014. Paper presented at the 14th Swiss Transportation Research Conference (STRC). Ascona, Switzerland.
- A dynamic approach for influencing multimodal mobility with limited parking. Zheng, N. and Geroliminis, N., 2014. Paper presented at the 3rd Symposium of the European Association for Research in Transportation (hEART). Leeds, UK.
- Modeling and optimization of multimodal urban network with limited parking and dynamic pricing. Zheng, N. and Geroliminis, N., 2015. Paper submitted to the 94th Annual Meeting of Transportation Research Board. Washington D.C., USA.

Chapter 4 has two main objectives: (i) It investigates the existence of aggregated relationships that describes the performance of urban bi-modal networks with buses and cars sharing the same road infrastructure and (ii) identifies how this performance is influenced by the interactions between modes and effect of bus stops. To this end, a three-dimensional vehicle MFD is developed, relating the accumulation of cars and buses, and the total circulating vehicle flow in the network. This relationship experiences low scatter and can be approximated by an exponential function. An analytical model is then proposed to estimate a three-dimensional passenger MFD, which provides a different perspective of the flow characteristics in bi-modal networks, by considering that buses carrying more passengers. With the function form of the 3D-MFD, it is also derived and shown that a constant Bus-Car Unit equivalent value cannot describe the influence of buses in the systems as congestion develops. A partitioning algorithm is performed to cluster the network into a small number of regions with similar mode composition and level of congestion. Interactions between buses and cars in the partitioned regions are analyzed and compared. The preliminary results of this chapter are presented in:

- A city-scale Three-dimension Macroscopic Fundamental Diagram: Simulation findings. Zheng, N., Aboudolas, K. and Geroliminis, N., 2013. Paper presented at the 13th Swiss Transportation Research Conference (STRC). Ascona, Switzerland.

- Investigation of a city-scale three-dimensional Macroscopic Fundamental Diagram for bi-modal urban traffic, Zheng, N., Aboudolas, K. and Geroliminis, N., 2013. Paper presented at the 16th International IEEE Annual Conference on Intelligent Transportation Systems (ITSC). The Hague, the Netherland.
- A city-scale macroscopic fundamental diagram for mixed bi-modal urban traffic, Geroliminis, N., Zheng N. and Aboudolas, K., 2014. Paper presented at the 93rd Annual Meeting of Transportation Research Board. Washington D.C., USA.

The last section of Chapter 4 demonstrates the potential use of the 3D-MFD to real-time traffic management, via perimeter flow control (note that the author of this thesis mainly contributes in the application of the developed controller and numerical analysis). The presented simulation results and analyses are included in the following papers where the author of this thesis serves as the second author:

- Perimeter flow control in bi-modal urban road networks: A robust feedback control approach. Aboudolas, K., Zheng, N. and Geroliminis, N., 2014. In the proceedings of the 2014 European Control Conference: 2569-2574.
- Robust control of bi-modal multi-region urban networks: An LMI optimization approach. Aboudolas, K., Zheng, N. and Geroliminis, N., 2014. In the proceedings of the 17th International IEEE Annual Conference on Intelligent Transportation Systems (ITSC). Qingdao, China (a journal version is under preparation considering the above two articles).

Chapter 4 as a whole addresses research objective III, given research background in Sub-section 1.1.4. The complete version of the modeling part is published in:

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

Chapter 5 concerns the development of network-level congestion pricing schemes for multimodal urban networks with the concept of the MFD and implementation in an agent-based approach, with the consideration of user adaptation and user heterogeneity. Two time-dependent schemes are designed through a Proportional (P) and a Proportional-integral (PI) type feedback controllers respectively, where prices are adjusted based on the level of congestion and behavior adaptation of the travelers that are observed offline from the MFD. Within the PI pricing scheme, the toll revenue is redistributed to promote the usage of buses.

Furthermore, two groups of users are differentiated with respect to their value-of-time in order to investigate equity. To ensure the MFD-based analysis is suitable, the output of the agent-based model is firstly investigated for consistency with the aggregated congestion physics as expressed by the MFD. Two case studies are then carried out implementing the two pricing schemes in the Zurich center network and the Sioux-Fall network respectively. Comparing the results of the two pricing schemes, the PI-based scheme is found more efficient as it is flexible in toll adjustment. Efficiency can be further improved when the promotion of bus usage by toll redistribution is incorporated in the pricing scheme. Considering user heterogeneity also results higher efficiency and more equitable benefit to users. The preliminary results of this chapter are presented in:

- A dynamic cordon pricing model – A robust and effective road pricing model, Zheng, N. and Geroliminis, N., 2012. Research presented at the European Conference on Transport Research (Transport Research Arena). Athens, Greece, 2012. (This work obtained The Year 2012 Golden Award of the Doctoral Transportation Research Competition at the conference).
- A dynamic cordon pricing scheme combining a macroscopic and an agent-based model. Zheng, N., Waraich, R., Axhausen, K. and Geroliminis, N., 2012. Paper presented at the 91th Transportation Research Board 2012 Annual Meeting., Washington, D.C., USA.
- Area-based pricing scheme for multimodal systems and heterogeneous users in an agent-based environment. Zheng, N. Rérat, G. and Geroliminis, N., 2015. Paper submitted to the 94th Transportation Research Board Annual Meeting. Washington, D.C., USA.

Chapter 5 as a whole addresses research objective IV, built in the research background of Sub-sections 1.1.3 and 1.1.4. Two complete versions are published respectively in:

- A dynamic cordon pricing scheme combining the Macroscopic Fundamental Diagram and an agent-based traffic model. Zheng, N., Waraich, R., Axhausen, K. and Geroliminis, N., 2012. Transportation Research Part A 46(8): 1291–1303.
- A time-dependent area-based pricing scheme for multimodal urban networks with user adaptation: An agent-based approach. Zheng, N. and Geroliminis, N., 2015. In the proceedings of the 17th International IEEE Annual Conference on Intelligent Transportation Systems (ITSC). Qingdao, China. (A journal version is under

preparation combining the 2015 TRB paper and IEEE ITSC 2014 conference papers listed above)

Chapter 6 summarizes the main findings and highlights the contribution of this thesis to the existing literature.

Please note that Chapters 2 to 5 are self-standing articles. For the sake of readability, a nomenclature is provided in the beginning of each chapter for the main variables and parameters utilized in that chapter. Same notations may be used with different meanings in different chapters.

2 A Multi-modal Multi-region Modeling Framework for Road Space Allocation

In this chapter, a quantitative approach is proposed, linking traffic performance and road space allocation for multimodal urban transport systems. Such system can be treated as an interconnected network of regions (sub-networks) with one or more modes moving. In this extension, different parts of a city can be subject to different management strategies (see for example Fig. 2.1). Perhaps bus-only streets are allocated only in the central region while other parts of the city allow vehicles to operate in mixed traffic. It is important to understand how the space should be allocated and their impacts on traffic performance. With the recent findings in the macroscopic modeling and dynamics of traffic in cities which have provided knowledge of single-mode/ single-region cities and single-mode/multi-region cities, understanding the dynamic of multi-mode, multi-region cities is promising. Operational characteristics of different modes should be considered. Despite the different features of modes in terms of occupancy (number of passengers), driving behavior (speeds, acceleration and deceleration profiles, length), duration of travel, etc., all of these vehicles when moving to an urban environment make stops related to traffic congestion (e.g. red phases at traffic signals) and other stops, which also cause delays to the transportation system as a whole, e.g. buses stop at bus stops to board/alight passengers, taxis stop frequently and randomly when they search/pick up/deliver passengers, cars may stop/maneuver when search/find a parking spot. While there is a good understanding and vast literature of the dynamics and the modeling of congestion for congestion-related stops, the effect of service or general purpose stops in the overall performance of a transportation system still remains a challenge.

The proposed approach eventually should answer how road space should be allocated to each mode to minimize the travel costs of users traveling with all modes, providing quantitatively how changes in infrastructure, demand, or traffic management in one region have impact on the multimodal traffic performance. We will show that (i) the proposed approach captures the operational characteristics of different modes, (ii) the resulting system dynamics are consistent with the physics of traffic given different road space strategies, e.g. with or without dedicated bus lanes, (iii) allocation of road space can be readily optimized in a static and dynamic way, and (iv) pricing strategies can further improve the efficiency of the system with less space dedicated for buses.

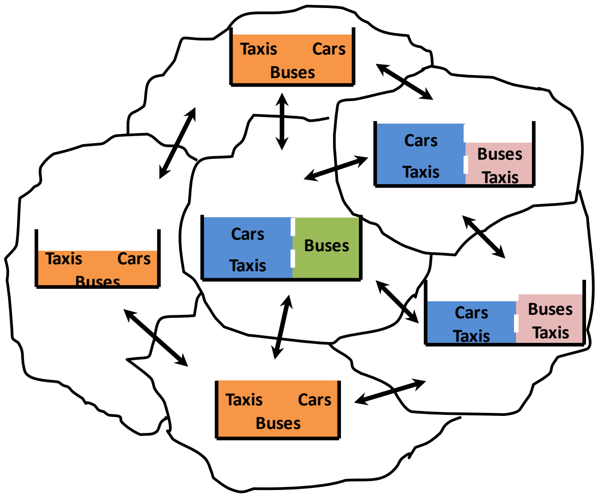

Fig. 2.1 A multi-region, multimodal urban road system.

In Sections 2.1and 2.2, the methodological frameworks for multimodal system dynamics and road space optimization are developed, respectively. Section 2.3 presents the results of a case study with a two-region city and two modes, buses and cars. A summary of this chapter is provided in Section 2.4.

Table 2.1 provides a nomenclature of main variables and parameters for Chapter 2.

Table 2.1 Nomenclature of main variables and parameters in Chapter 2.

Variables	Description
$Q_i^k(t)$	demand generated at time t in region i with next destination region k
$Q_i^{km}(t)$	demand generated at time t in region i with next destination region k, choosing mode m
$p_i^{kb}(t)$	Bus mode share for demand $Q_i^k(t)$
$O_i^m(n_i^b(t), n_i^c(t))$	flow exiting region i of mode m at time t as a function of accumulations
$r_i^{km}(t)$	fraction of O_i^m from region i with final destination k at time t
$P_i^m(t)$	total distance traveled by all vehicles of mode m in region i at time t
$O_{i\to j}^{km}(t)$	transfer flow of mode m from region i to j with final destination k at time t

$\pi_i(t)$	space share allocation plan of region i at time t
$U_i^{km}(t)$	utility of using mode m at time t in region i with next destination region k
$n_i^{km}(t)$	accumulation of mode m in region i with next destination region k at time t
$n_i^{m}(t)$	accumulation of mode m currently in region i at time t
ob_i^c	average number of passengers per car in region i
$ob_i^{kb}(t)$	average number of on-board passengers per bus in region i with destination k at time t
$\delta_{i\rightarrow j}^k(t)$	binary variable showing the sequence of the trip, with value equal to 1 if a trip from region i to k passes through j immediately after leaving i, and 0 otherwise
$OB_i^{kb}(t)$	total number of bus on-board passengers in region i with destination k at time t
$\alpha_i^b(t)$	parameter for the effect of dwell times on bus speed in region i at time t
$TT_i^b(t)$, $V_i^b(t)$	average travel time and speed by using mode m in region i at time t
$TT_d^b(t)$	extra travel time spent due to bus dwell times in region i at time t
$\overline{L}'_{ib}$	average trip length of bus passengers in region i
$\overline{L}_{im}$	average trip length of mode m in region i
C_{bus}	storage capacity of buses (persons per vehicle)
$D_i^b(t)$	bus discomfort (crowdedness) in region i at time t
θ_i	percentage of on-board passengers reaching destination i
γ, β_1, β_2	parameters for mode choice and bus crowdness

2.1 Bi-modal MFD Model and System Dynamics

Consider a city as a multi-region system, where the road network of the city is divided into N regions as in Fig. 2.1, denoted by $i = 1, \ldots \ldots, N$. Criteria for partitioning a region (approximate size a few hundred links each) are: homogeneous distribution of congestion within each region to obtain a low scatter MFD (see in details in Ji and Geroliminis (2012)), similar topological characteristics and similar type of mode usage. In this thesis, we assume both the partition result and the MFDs are given, where the MFDs can be estimated analytically (Boyaci and Geroliminis (2012)) or approximated given data collected from sensors (Keyvan-Ekbatani et al. (2012), Ortigosa et al. (2013), Leclercq and Geroliminis (2013)).

Any region i is partitioned into sub-regions, each one containing a specific type of mode usage, e.g. it can be dedicated bus lanes, mixed traffic lanes, car-only lanes or any other special usage lanes. The strategy of allocating fraction of space to each sub-region in region i at time t is denoted by vector $\pi_i(t)$ (which can be static e.g. 20% of space for bus-only lanes all the time, or dynamic e.g. 10% during off-peak while 20% during peak). Without losing the aggregated treatment of π, detailed distribution of the space is not considered here e.g. which roads have bus lanes (nevertheless an average cost of accessing those lanes is included in the model).

Demand defined as regional origin-destination is also considered known (for instance, demand from region i to k at time t). We will later relax this assumption and investigate how uncertainty and errors in the demand influence our approach.

The goal of road space optimization for such multimodal multi-region transport systems is to minimize the total passenger hours travelled (PHT) over time for all modes of transport m that serve the total demand, by redistributing the road space in areas with different usages of the city. Mathematically, the optimization problem is

$$\min_{\boldsymbol{\pi}(t)} Z = \sum_{t,i,m} PHT_{t,i,m}\,(\boldsymbol{\pi}(t)) \tag{2-1}$$

Problem (2-1) is subject to the dynamics of the transport system. The dynamic interactions are illustrated in Fig. 2.2, (change the index of variable vectors to bold) where $\boldsymbol{n}(\boldsymbol{t})$ is the vector of accumulation of vehicle in different regions, $\boldsymbol{O}(\boldsymbol{t})$ the outflow, $\boldsymbol{Q}(\boldsymbol{t})$ the demand and $\boldsymbol{C}(\boldsymbol{t})$ the cost of travel for each of the regions. Given the initial state of the system at time t_0

and a space allocation strategy $\boldsymbol{\pi}(\boldsymbol{t})$, the system modeled by three parts that interact at every time step: The average traveling speed and travel cost of each mode at each region, the flow dynamics within and between regions and mode choice of the generated demand at each region. We develop the methodological framework in the following sections.

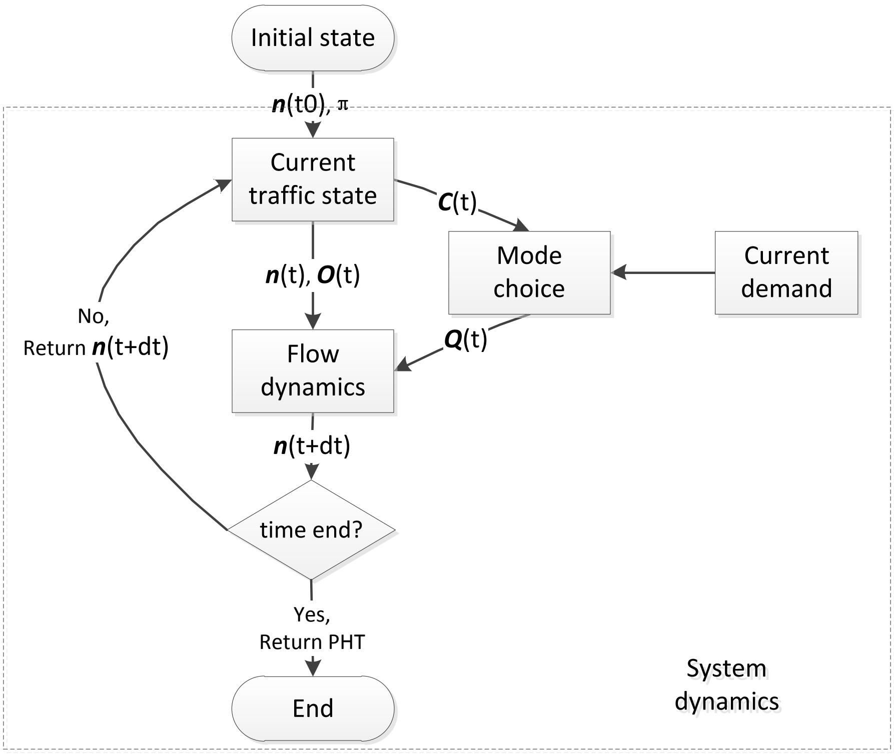

Fig. 2.2 System dynamics flow chart.

Traffic flow dynamics

Let us now describe in detail the dynamics, by looking at individual region i. Let $Q_i^{km}(t)$ be the demand generated in region i with final destination k choosing to travel with mode m at time t. The total demand generated during this interval from i to k, $Q_i^k(t) = \sum_m Q_i^{km}(t)$. Route choice for each demand pair from i to k is defined exogenously with binary variables $\delta_{i \to j}^k(t)$, that show the sequence of the trip, with value equal to 1 if a trip from region i to k passes through j immediately after leaving i, and 0 otherwise. It is assumed that all trips generated from i with destination k follow the same sequence of regions in their route. In our model it is assumed that the sequence of regions is not altered by traffic conditions (the sequence of passing regions, e.g. $i \to j \to k$, is exogenously given). Let us now consider two modes bus and car, with indices c and b respectively $m = \{b, c\}$. Let $n_i^{km}(t)$ be the accumulation of mode m in region i with final destination k, $O_{i \to j}^{km}(t)$ be the transfer flow of mode m from region i to j with final destination k. Note that $n_i^m(t) = \sum_{k=1}^N n_i^{km}(t)$. Note also that for $i = k$, $O_{i\to j}^{km}$ is zero $\forall j \neq i$. $O_{i\to i}^{im}$ is the trip ending flow with final destination region i.

As long as flow $O_{i\to j}^{km}(t)$ enters j, it becomes part of accumulation $n_j^{km}(t)$. The dynamic equations between the state variables n_i^{km}, flow variables $O_{i\to j}^{km}$, and the demand Q_i^{km} of the system with N regions for the two modes are

$$n_i^{kc}(t+1) \ = n_i^{kc}(t) + \frac{Q_i^{kc}(t+1)}{ob_i^c} - \sum_{j=1}^{N} O_{i\to j}^{kc}(t) + \sum_{l=1}^{N} O_{l\to i}^{kc}(t), \tag{2-2}$$

$$n_i^{kb}(t+1) \ = n_i^{kb}(t) - \sum_{j=1}^{N} O_{i\to j}^{kb}(t) + \sum_{l=1}^{N} O_{l\to i}^{kb}(t), \tag{2-3}$$

where, ob_i^c is the occupancy of cars in region i, which is a constant (say 1-1.3 passenger per car) and will remain the same until the end of the trip and $\frac{Q_i^{kc}(t+1)}{ob_i^c}$ is the number of generated car trips. Both equations are discrete versions of a mass conservation differential equation, where t represents an interval. Equation (2-2) states that the change of car accumulation n_i^{kc} over time equals to:

(i) The internally generated demand from region i to k (2nd item, RHS of (2-2)), minus

(ii) The outflow from region i with final destination k (3rd item, RHS of (2-2)), plus

(iii) The inflow from all regions adjacent to i with final destination k, which is also the outflow for these regions (last item, RHS of (2-2)).

A graphical illustration of Equation (2-2) with indications of the flow items can be found in Figure 2.3.

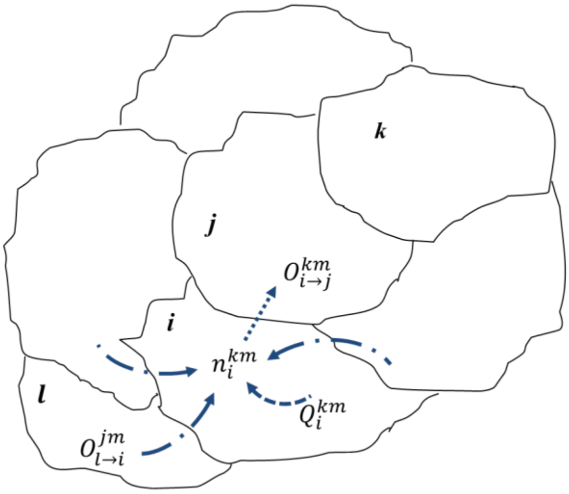

Figure 2.3 An illustration of the flow conservation as described by Equation (2-2), where index m corresponds to c, cars

Equation (2-3) is similar to (2-2), but without new generation of bus trips (Q^{kb}_i) to the system. It is assumed that the amount of buses in service is usually constant during typical commuting hours. Note that the number of buses in each region at different times depends on the relative speeds and MFDs of the regions.

Flow $O^{km}_{i\to j}(t)$ is the minimum of two terms: the sending flow from region i, which depends on the accumulations of region i and the boundary capacity of region j, BC_j, which is a function of the receiving region's accumulations. Nevertheless, this constraint can be ignored during the optimization process. The physical reasoning besides this assumption is that (i) boundary capacity decreases for accumulations much larger than the critical accumulation that maximizes flow (see Geroliminis and Daganzo (2007)), and (ii) optimized space allocation will not allow the system to get close to gridlock. Note also that $O^{km}_{i\to j}(t) = 0, if\ \delta^{k}_{i\to j}(t) = 0$.

The sending flow of $O^{km}_{i\to j}(t)$ is estimated by the macroscopic fundamental diagram O^m_i. The MFD of mode m of region i is denoted as $O^m_i(n^b_i(t), n^c_i(t))$. For our approach, the given condition is that the city is partitioned in homogeneously congested regions and that each region has a well-defined MFD. MFD O^m_i can be analytically estimated as a function of accumulation for different mode usages and space allocations using variational theory. Then,

$$O^{km}_{i\to j}(t) \ = min\left\{O^m_i(n^b_i(t), n^c_i(t)) \cdot r^{km}_i(t), BC_j(n^b_i(t), n^c_i(t))\right\} \qquad (2\text{-}4)$$

where $r_i^{km}(t)$ is the proportion of the outgoing flow O_i^m from region i to j with final destination k. If the average distance traveled by travelers within region i is assumed to be independent of their destinations, then the proportion $r_i^{km}(t)$ can be estimated as a ratio of the related accumulations, according to Little's formula (Little, 1961).

$$r_i^{km}(t) = \frac{n_i^{km}(t)}{n_i^m(t)} \tag{2-5}$$

We should also estimate the dynamics of passengers for each mode and every region. With respect to the car, we just need to multiply both sides of Equation (2-2) by ob_i^c. For simplicity, we assume $ob_i^c = 1$. With respect to the bus, we need to estimate how the occupancy of bus (in passengers) evolves over time according to demand and finished trips. The dynamics of bus passengers are very different than bus dynamics. A simple on-board passenger mass conservation equation is

$$\begin{aligned} OB_i^{kb}(t+1) = OB_i^{kb}(t) + Q_i^{kb}(t+1) - \sum_{j \neq i}^{N} O_{i\to j}^{kb}(t) \cdot ob_i^{kb}(t) + \\ \sum_{l=1}^{N} O_{l\to i}^{kb}(t) \cdot ob_l^{kb}(t) - b_i^k \cdot OB_i^{kb}(t) \cdot (1 - (1-\theta_i)^z) \end{aligned} \tag{2-6}$$

where

$$ob_i^{kb}(t) = \frac{OB_i^{kb}(t)}{n_i^{kb}(t)}, \tag{2-7}$$

$$\theta_i = \begin{cases} \left(\dfrac{\bar{L}'_{ib}}{s_i}\right)^{-1}, & if\ i = k \\ 0, & otherwise \end{cases} \tag{2-8}$$

In Equation (2-6), OB_i^{kb} is the number of bus on-board passengers currently in region i with final destination k. (2-7) estimates ob_i^{kb}, the average number of on-board passengers per bus from region i to k. The RHS of (2-6) consists of the following terms: (i) the passenger inflow at time $t+1$ for the specific OD pair, (ii) the passengers that move outside region i while in the bus, (iii) the passengers that move inside region i while in the bus and (iv) the passengers that finish their trip inside region i. The last term is non-zero only for $i = k$ (binary $b_i^k = 1$ for $i = k$ and zero otherwise). This term approximates the passenger trip endings in region i as the mean of a Bernoulli trial which is repeated z times for each passenger with probability of success θ_i. Variable z is the number of stops that a bus travels during interval t, which is $V_i^b(t)T/s_i$ (T is the duration of the interval and $V_i^b(t)$ is the space-mean speed of buses in region i, which will be described later). At each stop in region i, every on-board passenger from the previous time step has probability θ_i of reaching its destination. $\overline{L'}_{ib}$ is the average trip length of bus passengers and s_i is spacing between bus stops for region i, i.e. passengers travel $1/\theta_i$ stops on average. Note that the trip length of a passenger, $\bar{L}'_{ib}$, is different than the trip length of a bus in the region, $\bar{L}_{ib}$. If more detailed data exists from real measurements, the generic model can be adjusted accordingly with a non-homogeneous Bernoulli process (or a Brownian motion). How the demand Q_i^k is divided between the two modes, Q_i^{kc} and Q_i^{kb} is discussed later.

Multimodal travel time estimation

The multimodal traffic flow dynamics of the previous section considered that the demand is known for each mode. This will be an input of a mode choice model, which is presented in the next section. To obtain this goal, it is important to identify the speed and travel time of each mode of transport. The estimation of space-mean speed $V_i^m(t)$ for each mode depends on the type of usage, i.e. mixed traffic of cars and buses or dedicated bus lanes. Keep in mind that variational theory estimation for shared-use of car and bus estimates the network flow and of cars for given car density given the conflicts and the operational characteristics of buses (frequency, dwell times etc.). It does not estimate the speed of buses and extra effort is needed to account for the additional delay due to dwell times. This extra step is not required in separated bus lanes as there is only one mode of transport.

The space-mean speed in a time interval is by definition the ratio of the total distance travelled and the total hours traveled. Let $P_i^m(t)$ be the total distance travelled by vehicles of mode m (bus or car) in region i at time t and $\bar{L}_m$ the average trip length of mode m in region i. Then, for steady state queuing systems $P_i^m(t) = O_i^m(t) \cdot \bar{L}_{im}$ (Little, 1961) and $V_i^m(t) \stackrel{\text{def}}{=} P_i^m(t) \ / n_i^m(t)$. Geroliminis and Daganzo (2008) have shown that trip length of cars in downtown Yokohama was a time invariant variable. The average trip length of buses can be easily estimated as buses have fixed routes. By using the above equations and variational theory (Daganzo and Geroliminis, 2008; Geroliminis and Boyaci, 2012), we can estimate (i) O_i^c and V_i^c for car-only regions and mixed traffic car-bus regions and (ii) O_i^b for bus-only regions. For shared-use car-bus regions, the speed of buses further decreases because of dwell times for picking up and dropping off passengers by a parameter $0 < \alpha_i^b(t) \leq 1$ and $V_i^b(t) =$

$V_i^c(t)\cdot\alpha_i^b(t)$. If we assume that buses have the same speed with cars when they are not stopped at bus stops then $\alpha^b(t)$ can be approximated as

$$\alpha_i^b(t) = \frac{TT_i^b(t)}{TT_i^b(t)\ + TT_d^b(t)} \tag{2-9}$$

where $TT_d^b(t)$ is the total time spent at bus stops (total dwell time) in interval t, and $TT_i^b(t)$ is the average travel time of buses in region i excluding TT_d^b, which is given by $\bar{L}_{ib}/V_i^c(t)$. Average travel time of cars in region i, $TT_i^c(t)$, can be estimated as $\bar{L}_{ic}/V_i^c(t)$.

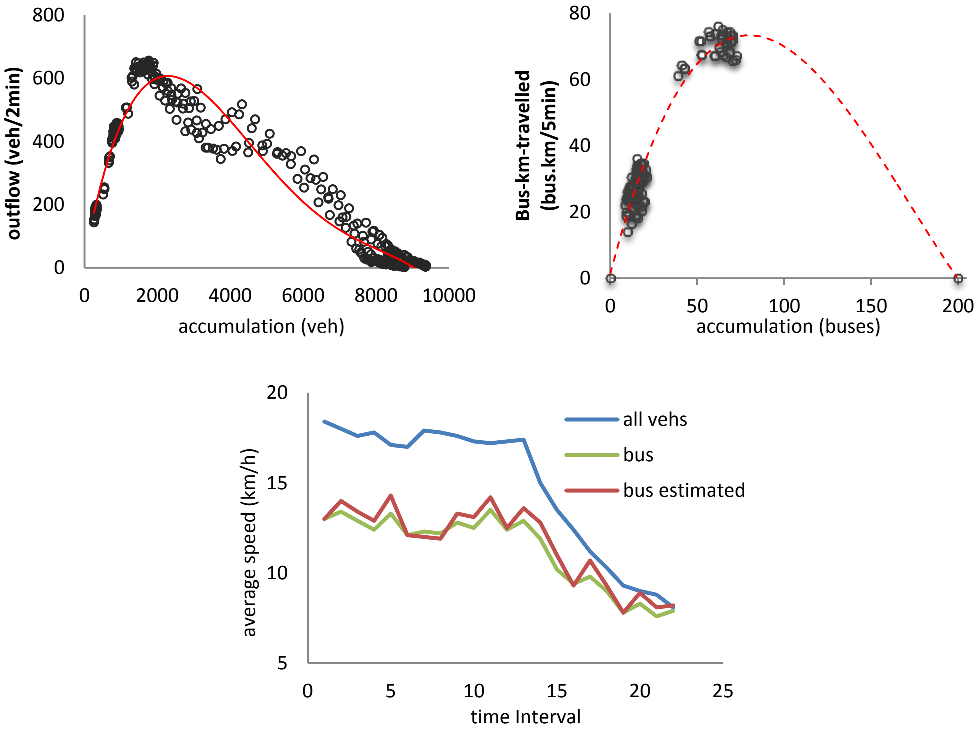

Fig. 2.4 (a-top left) The MFD of mixed car-bus network, (b-top right) the MFD of dedicated bus lanes and (c-down) speed profile for mixed network.

To validate all of the above assumptions, we performed a micro-simulation of San Francisco network taken from (Geroliminis and Daganzo, 2007), which provides a low scatter MFD and can be estimated with variational theory with small error. While in the previous version this was a car-only simulation, we performed additional simulations for shared-use bus-car lanes and for dedicated bus lanes with different bus frequencies and dwell times. Fig. 2.4 summarizes the MFDs in some of the scenarios for mixed traffic (Fig. 2.4(a)) and the dedicated bus lanes (Fig. 2.4(b)), while Fig. 2.4(c) shows the average speed of cars, the speed of buses and estimated speed of buses according to the aforementioned methodology (time interval is 5min). A detailed analysis of these results will be reported later in Chapter 4.

Aggregated mode choice

Suppose that the selection of the preferred mode is made by the travelers only once when they enter the network and start their trip. The total demand generated at each time interval $Q_i^k(t) = \sum_m Q_i^{km}(t)$ is assigned to the preferred modes through an endogenous approach. Mode choice of a traveler only happens at this moment and remains the same until her trip is finished. The choice of a mode is based on the utility of the mode, which is expressed as the travel cost of using that mode at the beginning of the trip. We consider that the travel cost of the whole route consists of travel time for cars and buses $TT_i^c(t)$ and $TT_i^b(t)$, on-board spatial discomfort $D_i^b(t)$ for buses and extra costs $C_i^c(t)$ for cars vs. buses (e.g. parking, tolls, fare) for each region i and interval t. Calculation of utility for car and bus for a specific origin destination pair (from i to k) is given by (2-10), where the summation term considers the set of all regions that a trip passes on its route from i to k, $\{S_i^k\}$, that can be estimated using binaries $\delta_{i \to j}^k(t)$.

$$\begin{aligned} U_i^{kc}(t) &= -\sum_{j \in \{S_i^k\}} \left(TT_j^c(t) + C_j^c(t)\right) \\ U_i^{kb}(t) &= -\sum_{j \in \{S_i^k\}} \left(TT_j^b(t) + D_j^b(t)\right) \end{aligned} \tag{2-10}$$

where

$$D_j^b(t) = \gamma \cdot \left(\frac{ob_j^b(t)}{C_{bus}}\right)^2 \tag{2-11}$$

The rationale of the spatial discomfort D_j^b is that (i) the utility of using bus decreases if number of on-board passengers (crowdedness) increases and (ii) it prevents the buses from reaching overcrowded conditions if buses are much faster than cars. Spatial discomfort of passengers D_j^b is a function of the average occupancy of a bus at time t, which is the ratio between the average on-board passengers per bus in region j, $ob_j^b = \sum_{k=1}^N ob_j^{kb}$ and the storage passenger capacity of a bus, C_{bus}. The spatial discomfort term is transformed to time unit by parameter γ, which should be calibrated with real data. Nevertheless, during the optimization process, which is described in details in the next section, we noticed that by changing the chosen value of γ by ±50%, objective function varies only by 2-3%. A different formulation with the same logic can be proposed for D_j^b, which should always be a monotonically increasing function of bus average occupancy.

Given the described utilities, we propose to model the mode choice of buses over time $p_i^{kb}(t)$ (for cars is $1 - p_i^{kb}(t)$) and $Q_i^{kb}(t)$ with a sequential approach, where the difference of utilities between the modes for a given origin destination pair is $\Delta U_i^k(t) = U_i^{kc}(t) - U_i^{kb}(t)$.

$$\begin{gathered} p_i^{kb}(t+1) = p_i^{kb}(t) + \beta_1 \cdot \Delta U_i^k(t) + \beta_2 \cdot \left(\Delta U_i^k(t) - \Delta U_i^k(t-1)\right) \\ Q_i^{kb}(t+1) = Q_i^k(t+1) \cdot p_i^{kb}(t+1) \end{gathered} \tag{2-12}$$

Equation (2-12) states that (i) if traffic conditions do not change, mode choice percentage remains the same as of the previous period and (ii) if traffic conditions change, mode choice percentage changes in proportion to the difference in utilities as expressed by ΔU_i^k, and to the evolution of ΔU_i^k. In such a way, we assume that all travelers make smart choices based on current and historic information. Note that travelers choose the mode of travel only once in the beginning of their trip based on the traffic conditions of this time. Parameters β_1 and β_2 should be calibrated using real data. Alternatively, one can use the traditional logit-form model. Since we are trying to model a dynamic choice process and we consider that a dynamic estimation of a logit model might be a challenging task, we utilize the proposed approach while a more complicated mode choice model with diverse utility considerations can be developed and estimated given real data. Furthermore, it is worth mentioning that Equation (2-12) tries to succeed a set point, which is an equal utility of modes in our case (equilibrium in the choices). This type of controllers is a generic widely used control loop feedback mechanism. The equation calculates an error value as the difference between a measured process variable and a desired set-point. It attempts to minimize the error ΔU_i^k by adjusting the process inputs, p_i^{kb}. The adjustment depends on the present error (term β_1) and the accumulation of past errors (term β_2). Sensitivity analysis of parameters γ, β_1 and β_2 is performed later.

2.2 Road Space Allocation Management

Road space can be deliberately allocated between competing modes. Although the allocation of this space is a policy-oriented decision, it should be informed by the correct congestion physics and system dynamics. After the introduction of the system modeling in the previous section, we can now build up an optimization framework for space allocation. Given the space allocation vector $\boldsymbol{\pi}$ (which include temporal and spatial variables), the total passenger hours travelled (PHT) can be obtained as (where T is the duration of the interval t)

$$PHT(\boldsymbol{\pi}) = \sum_t \sum_i \sum_k \left(n_i^{kc}(t) \cdot ob_j^{kc} + OB_i^{kb}(t) \right) \cdot T \tag{2-13}$$

Optimization of (2-13) is highly non-linear. We solve this problem by a non-linear programming method, the sequential quadratic programming (SQP). SQP method solves a sequence of optimization sub-problems, each of which optimizes a quadratic model of the objective subject to a linearization of the constraints. Consider a non-linear programming problem of the form $min_{\pi}\, PHT(\pi)$ subject to constraints $g_M(\pi) \geq 0$ and $h_M(\pi) = 0$. In our case, $g_M(\pi)$ will state the constraints of implementation of space allocation in practice, e.g. space should not be frequently reallocated. $h_M(\pi)$ will be system dynamics equations introduced in the previous sections. The general Lagrangian for this problem is defined as (λ and σ are Lagrange multipliers)

$$\mathcal{L}(\pi, \lambda, \sigma) = PHT(\pi) + \sum_M \lambda_M \cdot g_M(\pi) + \sum_M \sigma_M \cdot h_M(\pi) \tag{2-14}$$

At an iteration π_K, an approximation is made of the Hessian of the Lagrangian function, using a quasi-Newton updating method. This is then used to generate a quadratic programming sub-problem, and it can be solved using any QP algorithm. The solution is used to form a search direction d_k for searching a new iterate $\pi_{K+1} = \pi_K + \alpha_K d_K$, where the step length α_K is determined by an appropriate line search procedure. The iteration will continue until stop criterion is achieved: the change of objective function is below a certain threshold. For detailed description of the SQP algorithm, please refer to Nocedal and Wright (2006). We apply this algorithm for multiple initial values (around 1000) to avoid convergence to local minima, which might be the case for a non-smooth objective function.

2.3 Case Study and Analysis

An application of the developed methodological framework is performed with a case study of a two-region city. Consider an urban network with two concentric regions, as shown in Fig. 2.5. Mixed traffic of buses and cars occurs in the outside region (periphery), while a fraction of road space in the center region is dedicated to buses. The radius of the center region is R_1 =1.6km and of the periphery is $R_1 + R_2 = 3.2$km. The road networks of the two regions are well connected at the border. We simulate an urban road traffic system for 4-hours (80 time units), a typical morning or evening period. Demand has a symmetric trapezoidal shape with time and the length of peak period is equal to 1hr. A 70% fraction of the demand generated in the periphery will travel to the city center and 30% fraction of the demand generated in the center will travel to the periphery of the city. The dynamics of the system follow the equations defined in Sections 2.1 and 2.2, while MFDs for different sub-areas are estimated with variational theory. Two modes of transport are considered available in the system, car and bus. A fraction 10% of the users are captive and do not have access to cars. The network topology, signal settings and traffic parameters are as follows: (i) two-lane one-way roads with block lengths $l_{b1} = 154$m and $l_{b2} = 308$m, (ii) signal settings for all intersections, green, $g = 40sec$; cycle, $c = 90sec$ and offset, $o = 0sec$, (iii) a triangular fundamental diagram for each link with free flow speed for buses and cars, $u_f = 15m/sec$; jam density, $k_{jam}^{c} = 0.15veh/m$ for cars and $k_{jam}^{b} = 0.05veh/m$ and congested wave speed for buses and cars, $w = 5m/sec$.

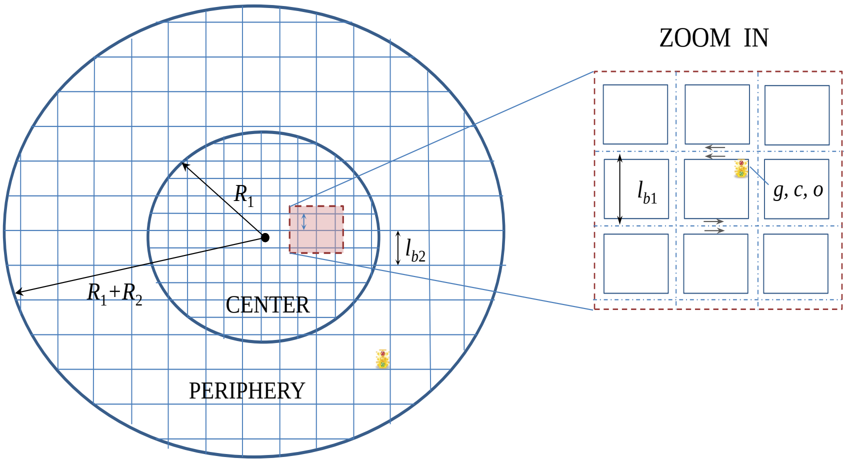

Fig. 2.5 Illustration of the case study two-region network sturcture and properties.

Static and dynamic space allocations are investigated. We also investigate how an area-based pricing strategy can further facilitate demand shift from cars to public transport. Sensitivity analysis for the non-physical parameters of the model is performed, while the effect of demand increase and uncertainty is also integrated in the analysis.

2.3.1 Static space allocation

We first investigate the performance of the traffic system, where a constant amount of space is allocated to dedicated bus lanes in the center region during all times. The results of the basic scenario where buses and cars share the whole network system without bus lanes, and of a scenario where 10% and 15% of the space in the center region is dedicated to bus lanes, are shown in Fig. 2.6. Note that the system performance of the latter scenarios significantly improves. The highly congested part disappears, the center of the city operates close to the maximum outflow of the car MFD, and total person hours traveled (PHT) during peak hours reduces by more than half.

Since we apply a static space allocation which has only one space variable to be optimized, we can simply enumerate all the possible values of space allocation and compare the results. Result of such an experiment is shown in Fig. 2.7(a), where PHTs are plotted over percentage of space for bus lanes. The small “kink” for $\pi = 3\%$ is because of the behavior of captive users that choose bus even if car is a faster mode. The optimal space allocation can be found around 10%. The SQP optimization approach of the previous section provides identical results. From the figure, an interesting observation is that the resulting PHTs are similar for $\pi = 10\%$ to 15%. This means that small variations in the demand or errors in the parameters of the model will still produce close-to-optimality results and PHTs will remain low.

Let us check more carefully the efficiency of the allocated space, which is defined as the ratio between the passenger trips completed during the total simulation time and the total lane kilometers of the space for buses. Space allocations of 10% and 15% cases are compared, where efficiency of bus lanes is shown in Fig. 2.7(b) and efficiency of space of the whole center region in Fig. 2.7(c). If we look at the space efficiency of the bus lanes, it is clear that with the increase of allocated space for buses, the trip completion rate per ln-km actually decreases. Note that for the whole region (Fig. 2.7(c)), the two cases are almost identical. The reason is that: More space for buses allows to transport more persons on bus lanes, but at the same time it constraints the capacity of car network and car users travel at a slower rate. Also, higher number of bus passengers increases the dwell times and the benefit of more space is smoothed off. Given the fact that the cost of operating dedicated bus lanes is high, the optimal solution should still hold at the lower value of 10%.

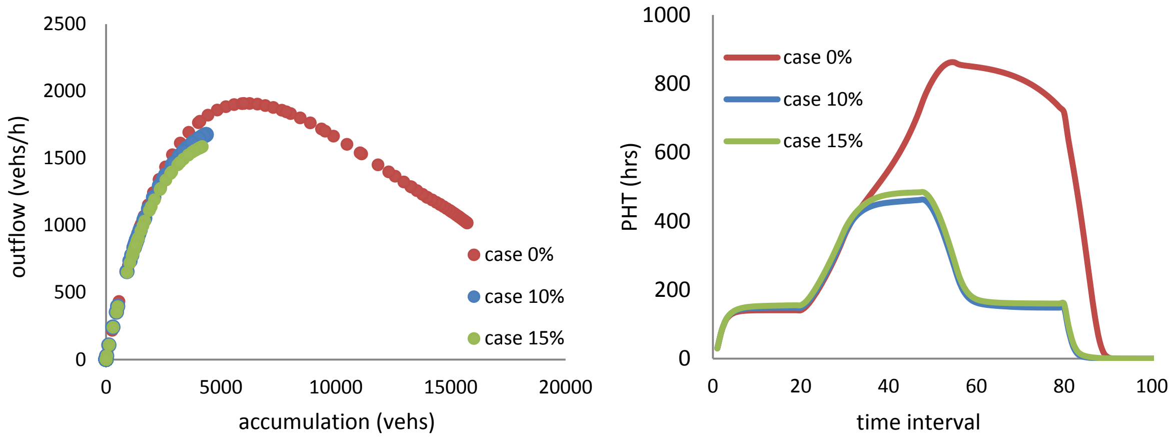

Fig. 2.6 An illustration of the performance of the system with two cases of space allocation to buses: (a) The car MFD in center region (cases 10%, 15% cars only, case 0% cars and buses mixed), and (b) PHT for both modes over time.

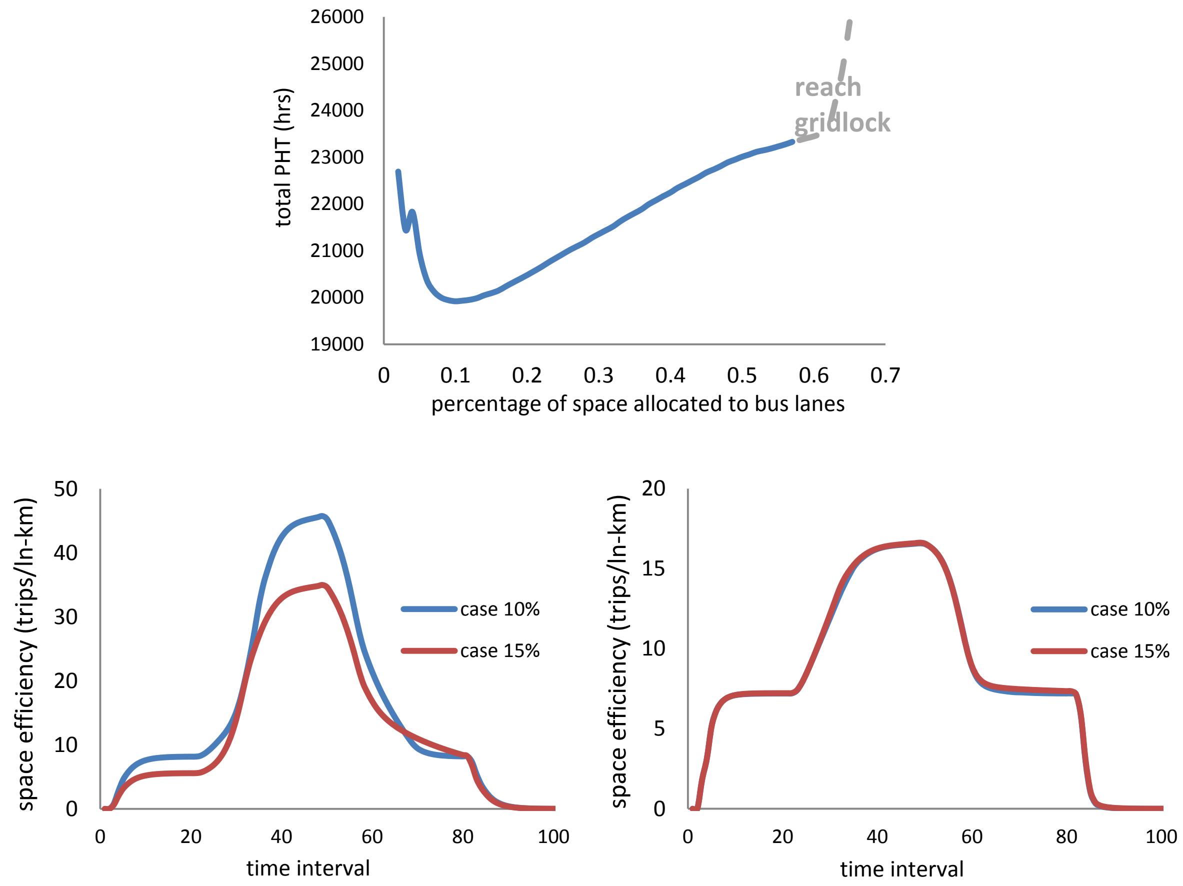

Fig. 2.7 (a-top) Total PHT, (b-down left) efficiency of bus lanes given difference space allocation, and (c-down right) efficiency of the whole space with different space allocation of bus lanes.

2.3.2 Dynamic space allocation

It is clear in Fig. 2.7(b) that the utilization of bus lanes is higher during peak hour, while it is much less during the onset and the offset of congestion. A dynamic space allocation strategy, where space is allocated in a time-dependent way is expected to increase spatial efficiency and further minimize the total travel cost. Nevertheless, given the associated infrastructure, the change of space should not happen many times during the day as implementation will be difficult. In this section, we investigate the optimal solution for a four-variable, three-period dynamic space allocation strategy. The three periods are the peak hour and the off-peak before and after the peak hour. The four optimized variables are: (i) the starting time t_1 and (ii) the ending time t_2 of the high space allocation, (iii) space allocated to bus lane during the off-peak periods π_1 and (iv) during the peak period π_2.

We apply the algorithm described in Section 2.2. The iteration and the solution of 10 executions with different random initial solutions out of the 1000 are shown in Fig. 2.8. The dotted line in both figures indicates the best solution among these 10 examples. We see that by applying a multiple random initial search, we are indeed able to avoid local minima, for example note a solution that gives 30% of the space to buses during the peak period.

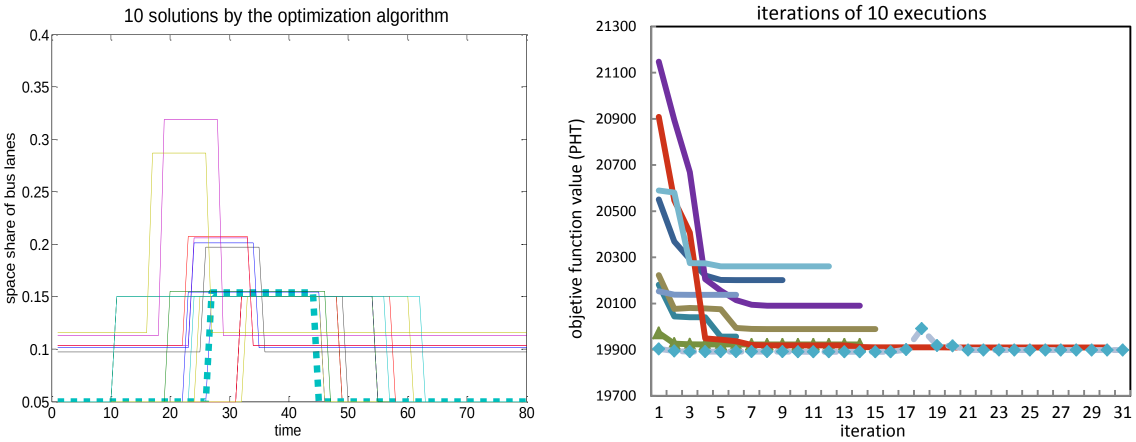

Fig. 2.8 (a) 10 solutions by the SQP optimization algorithm for different starting points, and (b) the corresponding optimal solutions after a number of iterations for each execution. The dotted line corresponds to the best solution.

Fig. 2.9 shows the optimal solution for this dynamic space allocation strategy. The line “pi” is the resulting space share of bus lanes over time, while “demand” is the demand profile. Note that the optimal strategy allocates less space when demand is low, but three times more during the “identified” peak period. The starting time and the ending time of the identified peak period makes physical sense, as it starts earlier than demand rate reaches its maximum and finishes later than demand rate starts decreasing. In such a way, the strategy “forces” travelers to choose to travel by bus proactively, before the beginning of peak hour, therefore all travelers avoid experiencing high travel cost. It also “tolerates” travelers to choose cars as the main mode of travel after the end of the peak hours. The optimal PHT is 19715 hours, better than the optimal PHT by a static space allocation, which is 20216 hours (10% further improvement from the $\pi = 0$ solution). Note that this improvement is succeeded with 15% less total space allocated for buses, as expressed by the total lane-km-hrs of individual bus lanes. Furthermore, a dynamic allocation gives flexibility in space management under demand fluctuation which is not possible in the static case.

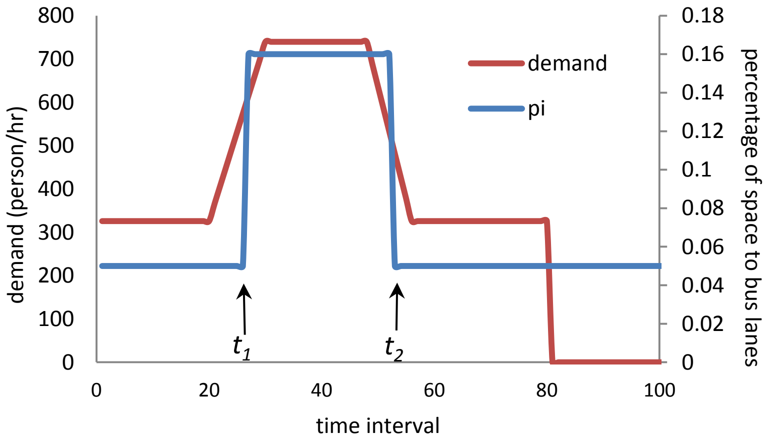

Fig. 2.9 The optimal solution of a time-dependent bus lane space allocation strategy (pi: π).

Now let us compare the efficiency of allocated bus lanes and the bus occupancy during the peak hour between the static strategy and the dynamic strategy. The results are shown in Fig. 2.10. During the off-peak period, the dynamic strategy allocates less space to buses, dropping from 10% to 5%, which increases the efficiency of bus lanes (as expressed by passengers served per lane-kilometer of bus space). While during the peak period, the dynamic strategy has lower utilization rate when the allocated space increases from the constant 10% to 16%, nevertheless the dynamic strategy actually serves more demand for buses during the period (shown by the right figure). Note that two sharp changes for the dynamic case are observed during the space changes, at times t_1 and t_2 (off-peak to peak, and peak to off-peak). Nevertheless, the number of served passengers is smooth vs. time (see Fig. 2.10(b)). These changes represent the transition of operation between the two space allocation scenarios.

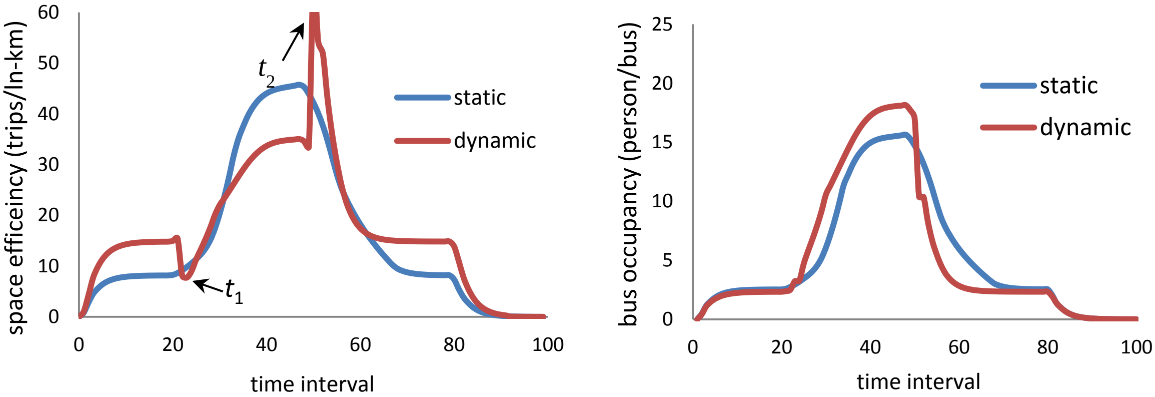

Fig. 2.10 Comparison of static and dynamic allocation: Time series of (a) space efficiency and (b) bus occupancy.

As a high utilization rate indicates that a large amount of demand chooses to travel by bus, one might question if this strategy will create cases of buses full of passengers (at capacity C_{bus}), which is neither comfortable nor realistic. This will not happen in the developed model, because Equation (2-10) ensures that mode choice of buses will reduce significantly when buses are crowded. If we check the occupancy of buses, as shown in Fig. 2.10(b), we see for both the static strategy and the dynamic strategy that buses are occupied less than half of their capacity (40 persons/bus). In fact, if more demand shifts from cars to buses, while buses are given enough space to travel with high speed, PHT could be further reduced. This is investigated in the next section.

2.3.3 Dynamic space allocation with pricing

Following the discussion, we will investigate measures to improve the occupancy of buses. To trigger mode shift from cars to buses, we apply an area-based pricing in the peak-hour, where car users have to pay when they drive in the center of the city. The cost of pricing is added into (2-10) and influences the mode choice. Since toll is in money units, we transform it to time by multiplying with a value-of-time (VOT). In our application we utilize a value of 16 Swiss Franc (CHF) per hour, which is estimated based on an analysis by Axhausen et al. (2005) for Swiss travelers. The start and end of pricing is assumed to be the same as of the peak-hour space allocation, between t_1 and t_2. The amount of toll charged c_1 is an additional variable to be optimized (5 in total). In our analysis we assume that toll will only affect the mode choice and not the departure time of the trip. While an analysis of the morning commute for multiple modes is analyzed for a simpler system in Gonzales and Daganzo (2012), departure time is not included in this current work as it will make the modeling part very tedious and intractable.

We execute this five-variable optimization with the SQP algorithm with multiple initial searches. The total PHT is 18215hrs, which is about 12% smaller than the PHT without pricing and dynamic space allocation. The space allocation during the peak period decreases from $\pi =$16% without pricing to 14%, while values for t_1 and t_2 are identical for the two cases. The savings in PHT are about the same with the total toll paid divided by VOT. While this is the case in the single bottleneck model (Vickrey, 1969), in systems governed by a variable capacity (like the MFD model), savings can be much larger, because congested states are avoided (Geroliminis and Levinson, 2009). But, the no toll case operates close to capacity, because of an efficient space allocation, while the toll paid is similar to delay savings. Nevertheless, as we will show later, an increase in the demand of travel in a city can create additional congestion and given that a new design of space allocation is not an easy solution (due to high infrastructure cost), variations in pricing can significantly improve the state of the system and avoid congestion.

The occupancy of buses over time is compared in Fig. 2.11(a) where “dynamic with pricing” is the resulting bus occupancy of the pricing strategy. As expected, the buses are more occupied after pricing. Besides, it can be observed that the occupancy goes to zero earlier when approaching the end of the simulation (interval 81), indicating that passengers are able to finish their trips earlier. This shows that a faster travel speed is achieved for all users and it corresponds to a reduction in PHT. Fig. 2.11(b) shows the MFD states for cars in the center of the city. The two different maximum values for the dynamic case represents the two different space allocations. With pricing the system operates at a more reliable state as this is less than the critical accumulation that maximizes outflow (n=4000veh). As we will show later, in this case the system can absorb increase in demand or small stochastic fluctuations without entering the congested regime, while in the case without pricing, a demand increase will also be associated with higher congestion.

The pricing strategy also succeeds higher demand shift and utilization rates of the buses during peak hours (Fig. 2.11(c)). Fig. 2.11(d) shows how the utility of bus and car change over time for two different O-D pairs (center to center and periphery to center). Note that when bus share is minimum, car utility (during the off-peak) is significantly higher, while during the other times the two utilities are about the same. Thus, model (2-12) succeeds in identifying a dynamic equilibrium of mode choice.

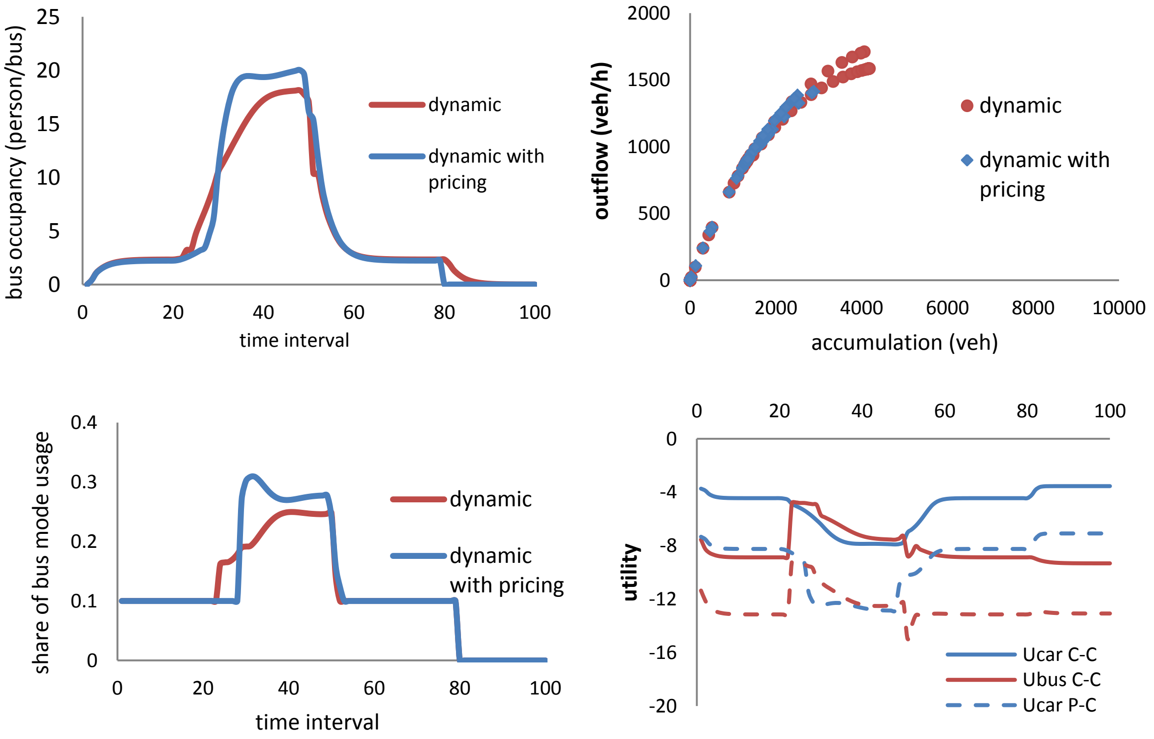

Fig. 2.11 Comparison of the dynamic allocation with and without pricing: (a-top left) Occupancy of buses vs. Time; (b-top right) MFDs of the center region for cars; (c-down left) Mode share of bus vs. time and (d-down right) Utilities of car and bus vs. Time for trips from center to center (C-C) and periphery to center (P-C) for the “dynamic with pricing” case.

2.3.4 Sensitivity analysis

In this section we discuss how sensitive is the resulting PHT, if parameters of mode choice β_1 and β_2 in Equation (2-12) and discomfort parameter γ are given different values. We also investigate how an increase in the demand will create additional congestion and if this can be resolved without a new re-distribution of the urban space. Table 2.2 shows the results of a sensitivity test. The basic scenario is the one where the solution of optimal static allocation is applied. For each of the three parameters, values vary by percentages as shown in the first row of the table. When one parameter is changing, the other two remain constant and the corresponding change of PHT is calculated. We see that the changes of β_1 and β_2 have small impact on the resulting PHT, while the change of γ has relatively greater impact. This result shows a good property of β_1 and β_2. While for γ, a higher value of γ prevents users from using buses therefore worse off the whole system significantly, while a low γ encourages users and consequently reduces PHT. But when γ is too low, buses attract too much demand, the dwell times increase and reduce the speed of buses therefore PHT starts to increase again.

Table 2.2 Sensitivity analysis on β_1, β_2 and γ.

	change of parameters	-75%	-50%	-25%	Base	+25%	+50%	+75%	+100%	+200%
change of PHT (%)	β_1	0.7	-0.03	0.1	0	0.02	0.01	0.2	0.3	0.6
	β_2	0.2	0.08	0.04	0	-0.09	0.1	-0.2	-0.2	-0.4
	γ	-1.4	-1.5	-0.8	0	0.7	1.4	2	3	5

Note that for β_1 and β_2, an extreme high value will result to an "all-or-nothing" mode choice: all users choose to travel with the better mode but then the travel cost of this mode becomes huge suddenly and result a complete choice of the other mode in the next time period, creating unrealistic instability and oscillations. For γ, a high value will prevent travelers from choosing buses, which not only underutilizes the bus space but also results in heavier congestion for cars. In our model, we tune the parameters via trial-and-error, with the objective of having a coherent and stable mode choice process. As mentioned already, these parameters need to be calibrated when the model is applied in a real city.

During the optimization process, perfect information about the demand profile was assumed. In reality demand can experience stochastic fluctuations or erroneous measurements, as it might not be straightforward to accurately measure, even at the aggregate level. To investigate how the model reacts under demand uncertainty, the effect of unbiased and biased errors are introduced for the optimal solution, obtained by the trapezoidal demand. With respect to the unbiased demand, we consider fluctuations from the deterministic trapezoidal demand as a standard normal distribution $\mathcal{N}(0,1)$, multiplied by a degree of error λ, i.e. $\hat{Q}_i^k(t) = Q_i^k(t) \cdot \left(1 + \lambda \cdot \mathcal{N}(0,1)\right)$. We perform multiple runs for values of $\lambda \in [0, 0.25]$ and the predetermined space allocation and tolls. Even for large degrees of error, the total passenger hours travelled (PHT) do not change more than 5%, which highlights that the approach is robust and not very sensitive to random demand fluctuations.

In the case of biased error, for example because of a demand increase, the analysis is more complicated as PHT can significantly increase. The reason is that the base case represents a high demand case, which without efficient space allocation creates significant congestion, while the optimized solution operates at the network capacity in the center region for the roads devoted to cars. Thus, this demand increase will create states in the congested part of the MFD and increase the PHT. Nevertheless, a new redistribution of the urban space might not be a feasible solution due to high infrastructure costs and it should not be performed frequently. Instead, a change in the price of toll is easy to be implemented and for some range of demand increase can still lead the system to an efficient state and avoid states in the congested part of the MFD, which results in network capacity loss. To quantitatively analyze the above, we estimate the total PHT of the two-region model as demand increases for the optimal space allocation and tolls of the base scenario. Afterwards, we keep the space allocation identical and optimize the system with respect to the value of the toll in the peak hour for different demand levels.

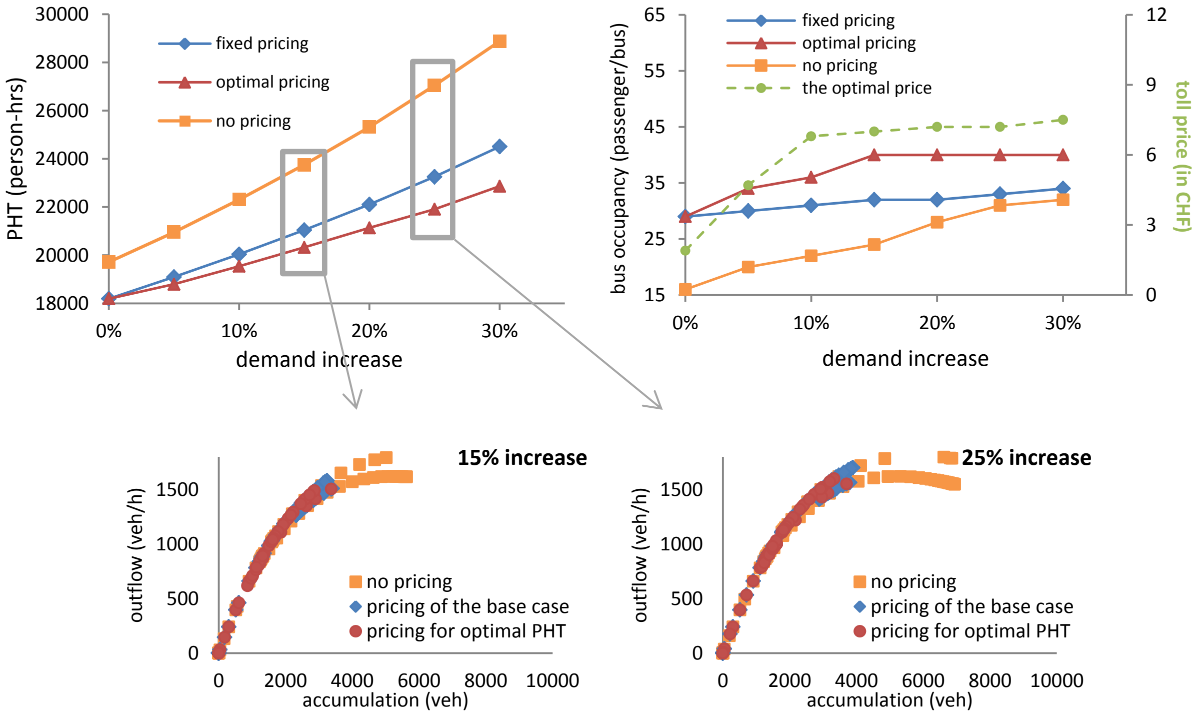

Fig. 2.12 Comparison of system performances under different pricing scenarios for different levels of demand increase: (a-top left) PHT and (b-top right) average bus occupancy over different demand increase percentages. (c-down left) and (d-down right) show the resultant MFDs under the three pricing when there are 15% increase and 25% increase of demand.

The results are summarized in Fig. 2.12, where PHT for different demand increase is estimated for (i) the dynamic allocation without toll, (ii) the dynamic allocation with fixed toll of the base scenario and (iii) the dynamic allocation with toll re-optimization for each demand level. We also show the average bus occupancy (in passengers/bus) during the peak hour for the three scenarios and the price of the optimal toll. Note that with price re-optimization, bus utilization is much higher while congested states are avoided for demand increase up to 25%

(see the associated MFDs for cars in the city center). For higher increase (>25%) buses become full and congestion is unavoidable. The price of toll stabilizes for higher demand as buses are full and cannot attract more passengers. A solution in this case is to introduce more frequent buses in the system, which will allow for lower bus occupancy, but it might decrease the efficiency of the bus network. Nevertheless, pricing appears to be an effective measure for dealing with uncertainty in demand.

2.4 Summary

In Chapter 2, we presented a macroscopic approach for allocating road space among modes of transport with the objective of minimizing the total hours traveled by travelers (PHT). We extended the single-mode MFD to a bi-modal one, where the effect of bi-modal operation in global performance was considered. For example, the effect of dwell time and the share of dedicated bus lanes (of the total space) on network space-mean speed were quantified by the bi-modal MFD. A system model then was constructed for a multi-region network, with a network-level flow conservation model and an aggregated dynamic mode choice model. Given this system model, the performance under different road space strategies can be predicted, provided with data input that can be readily collected such as regional origin-destination tables and road space allocation plans. We tested the rationale of this modeling approach with a two-region bi-modal city case study, and investigated the performance of two space allocation strategies for the center region of the city where demand was high and heavy congestion existed. A static and a dynamic (time-dependent) allocation strategies were obtained through non-linear optimizations respectively. We found that the dynamic allocation strategy managed to minimize PHT in a more efficient way as it utilized the bus lane space during the off-peak period and served a higher amount of passengers during peak period. By implementing pricing during the peak period, further reduction of PHT was achieved, as pricing led a higher demand shifted and increased the utilization of buses which was the faster traveling mode. We also observed the existence of user equilibrium in our macroscopic model, e.g. the utilities of traveling by modes became equivalent during the peak-period. Furthermore, we carried out sensitivity analysis and showed the robustness of our approach towards the fluctuations in demand inputs and model parameters.

3 Modeling and Optimizing Multimodal Urban Network with Limited Parking and Pricing

Chapter 2 presented a macroscopic approach to model the dynamics of a bi-modal transport system with infinite parking capacity. Unlimited parking is rarely realistic in urban city centers. Parking limitation has significant effect in mobility. The objective now is to integrate the treatment of parking limitation into the developed macroscopic modeling framework. The extended approach shall capture the flow characteristics of cruising for parking, estimate the direct cost of users with different mode choice, and enable the development and optimization of parking polies for improving multimodal mobility. Similar aggregated models treating parking limitation can be found in Geroliminis (2014), where a parking model built into an MFD modeling framework was proposed and perimeter flow control strategy was developed to reduce both parking cost and total cost for travelers, for single-mode systems; and in Arnott and Inci (2010), where equilibrium conditions under parking constraint (space, duration) are derived for a single-region single-mode and single-parking-facility system under steady-states.

Consider a system where travelling by car and searching for an on-street parking is the only available mode of transport (see for example Arnott and Rorwse (2009), Arnott and Inci (2010), Geroliminis (2014)). In this case under high demand the system has a stable equilibrium, which is close to gridlock. Parking will always be full and whenever there is a free spot, this will be occupied immediately by the cruising vehicles as very well stated in Arnott and Inci (2010) and also showed in Geroliminis (2014). If passengers have efficient alternative choices (e.g. change of departure time, utilizing a parking garage, switch to public transport) and some of these choices might be properly priced, then the system might end up in non-gridlock states. This chapter investigates the traffic dynamics of a system with limited on-street parking, unlimited but priced garage parking and a public transport alternative. It develops a feedback-based pricing scheme that does not require prediction of the state of the system and shows that hyper-congestion and cruising can be avoided. It also investigates how close such a traffic management strategy can get to the system optimum pricing with perfect information of the future conditions of the system. While detailed micro- or agent-based simulations could provide a scenario analysis of such a system, we follow a “dynamic aggregated approach” consistent with the physics of traffic congestion that can contribute to develop some strong physical intuition for such a challenging problem. Time of departure is not considered, as it will make the solution approach too hard for analytical derivations.

The chapter is organized as follows. Section 3.1 presents the system model under parking limitation consisting of four modules: the traffic flow model, system model, cruising-for-parking model and choice model, respectively. Section 3.2 develops two parking pricing strategies for congestion and cruising management. Utilizing the system model, case studies are carried out on the performance of the pricing strategies and the results are shown in Section 3.3. Furthermore, system performances under parking pricing competition (between parking facility operators) are estimated and preliminary results are discussed.

A standing-alone nomenclature of this chapter is provided in Table 3.1 for the main variables and parameters.

Table 3.1 Nomenclature of main variables and parameters used in Chapter 0.

Variables	Description
$P_i^m(t)$	Total distance traveled (production) in region i by mode m at time interval t
$N_i^m(t)$	The accumulation of mode m currently in region i at time t
$G_i^m(N_i^m(t))$	The production MFD for mode m in region i , in function of the accumulation N_i^m at time t
$Q_{i\to j}^{km}(t)$	Demand generated using mode m in region i approaching to the neighbor region j at time t with final destination regions k
$O_{i\to j}^{km}(t)$	Trip completion/transfer flow of mode m from region i approaching to the bounded neighbor regions j with final destination regions k at time t
$O_i^m(t)$	Vehicle outflow of mode m exiting region i at time t
$I_i^m(t)$	Incoming flow of mode m from external regions to region i at time t
$NP_i^m(t)$	Passenger accumulation on mode m currently in region i at time t
$OP_i^m(t)$	Passenger outflow on mode m exiting region i at time t
$N_{x,i}(t)$	Accumulation of car family x in region i at time t ($x = r, s, o, g, os$)
$O_{x,i}(t)$	Outflow flow of car family x in region i at time t
$\omega_{os,i}^c(t)\|\omega_{g,i}^c(t)$	Fraction of trip-finishing cars for on-street (or garage parking) in

	region i at time t
$\omega_i^m(t)$	Fraction of passengers travelling with mode m in region i at time t
$O_{r\rightarrow s,i}(t)\|O_{r\rightarrow g,i}(t)$	The transfer flow from the running family to searching family or to garage parking in region i at time
$l_{x,i}$	The average trip distance of car family x travelled in region i
$occ_i^m(t)$	The average occupancy of mode m in region i at time t
$N_{os,i}(t)$	The occupied on-street parking spaces at time t of region i
A_i	The total amount of on-street parking spaces of region i
$\varphi_i(t)$	The probably of finding an available on-street parking space at t in i
$L_i(t)$	The average cruising distance before finding an available on-street parking space at time t in region i
$T_{cru,i}(t)$	The average cruising delay at time t in a region i
$p_{os}(t)\|\,p_g(t)$	Pricing rates for on-street parking (or garage parking) at time t

3.1 System Dynamics with Parking Consideration

In this section, we present the macroscopic approach for modeling multimodal traffic dynamics with limited parking. The embedded traffic, parking, system and choice models will be described in details. An overview of such system is illustrated in Figure 3.1 below and will be explained in the later text.

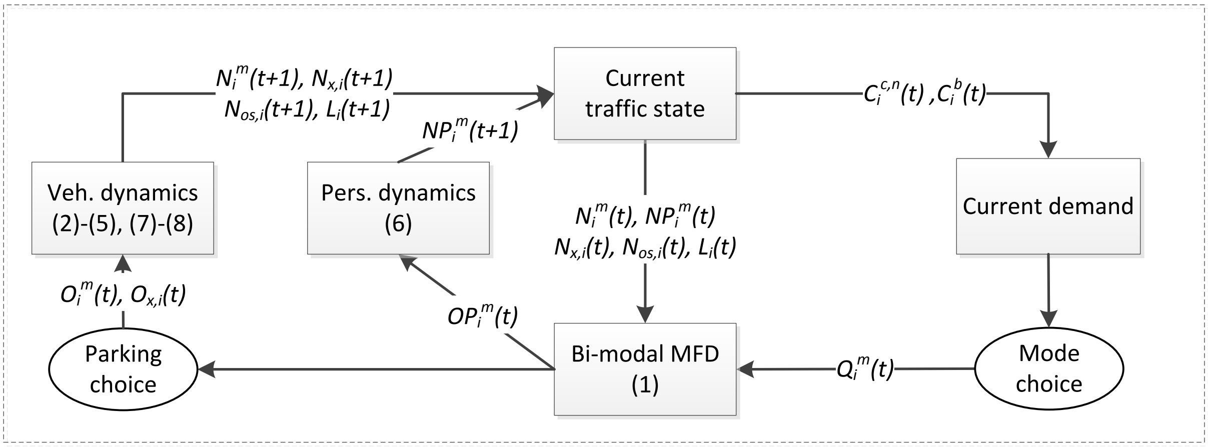
$N_i^m(t+1)$, $N_{x,i}(t+1)$
$N_{os,i}(t+1)$, $L_i(t+1)$
Current traffic state
$C_i^{c,n}(t)$,$C_i^b(t)$
$NP_i^m(t+1)$
Veh. dynamics (2)-(5), (7)-(8)
Pers. dynamics (6)
$N_i^m(t)$, $NP_i^m(t)$
$N_{x,i}(t)$, $N_{os,i}(t)$, $L_i(t)$
Current demand
$O_i^m(t)$, $O_{x,i}(t)$
$OP_i^m(t)$
Parking choice
Bi-modal MFD (1)
$Q_i^m(t)$
Mode choice

Fig. 3.1 Dynamics of a bi-modal transport system with parking choice, with indication of principal variables.

Traffic flow model

We utilize in this work both the single-mode and the multimodal MFD (cars and buses) as the traffic model, relating the total distance travelled in region i of mode m (m can be any vehicular mode that either utilizing dedicated road space or sharing space with other modes) at time t, the production $P_i^m(t)$, to vehicle accumulation $N_i^m(t)$. Mathematically, it is written as follows, for single-mode traffic (e.g. dedicated lanes separating cars and buses) and mixed traffic:

$$P_i^m(t) = G_i^m\left(\overrightarrow{N_i^m}(t)\right) \tag{3-1}$$

where $\overrightarrow{N_i^m}$ represents accumulation of all modes utilizing the specific road infrastructure; G_i^m is MFD functions which can be obtained via analytical approximations e.g. as in Geroliminis and Boyaci (2012) and Leclercq and Geroliminis (2013). Under steady state condition, we can approximate the trip completion rate, the outflow of a network $O_i^m(t)$ from the production by: $O_i^m(t) = P_i^m(t)/l_i^m$ where l_i^m is the average trip distance. We can also obtain the average speed by its definition $v_i^m(t) = P_i^m(t)/N_i^m(t)$.

System model

A city network shall be clustered into regions, as illustrated in Fig. 2.1 in Section 2.1 of Chapter 2, applying the same criteria for clustering: (i) homogeneous distribution of congestion within each region to obtain a low-scatter MFD (see Ji and Geroliminis (2012) for more details) and (ii) similar type of mode usage. Given the regional accumulation $N_i^m(t)$, the regional MFD G_i^m estimates production $P_i^m(t)$ and outflow $O_i^m(t)$ of vehicles, and of persons $OP_i^m(t) = O_i^m(t) \cdot occ_i^m(t)$ provided the average passenger occupancy of mode m, $occ_i^m(t)$. Given the traffic demand of mode choice m, $Q_i^m(t)$, the dynamics of each partitioned region i of this bi-modal system can be described by the change of $N_i^m(t)$ for vehicle flow, and the change of $NP_i^m(t)$ for person flow. A discrete version of the dynamics can be found as follows:

$$N_i^m(t+1) = N_i^m(t) + \frac{Q_i^m(t)}{occ_i^m(t)} + I_i^m(t) - O_i^m(t) \tag{3-2}$$

$$N_i^m(t+1) = NP_i^m(t) + Q_i^m(t) + IP_i^m(t) - OP_i^m(t) \tag{3-3}$$

In the equations, demand $Q_i^m(t)$=$\sum_{k \neq i} \sum_j Q_{i \to j}^{km}(t)$, where $Q_{i \to j}^{km}$ is the demand generated in region i approaching to the neighbor regions j with final destinations k. The regional outflow $O_i^m(t) = \sum_{k \neq i} \sum_j O_{i \to j}^{km}(t)$, is constrained by the receiving capacity of the approaching regions j and the boundary capacity between i and j. Variable I_i^m denotes the total incoming vehicle flow from the neighbor regions, $I_i^m(t) = \sum_k \sum_{j \neq i} O_{j \to i}^{km}(t)$, while $IP_i^m(t)$ is the incoming flow in person units. Note that (i) for route choice between an origin-destination pair, a regional route choice model can be applied to determine the sequence of the passing regions (Yildirimoglu and Geroliminis, 2014), and (ii) for details on the distributions of flows over the different regional ODs, for example $N_{i \to j}^{km}$ over N_i^m, readers can refer to Equation (2-5) of Chapter 2.

To estimate the cruising time for cars, it is indispensable to decompose $N_i^c(t)$ in Equation (3-2) into the three movement families that were introduced in the previous subsection: $N_i^c(t) = N_{r,i}(t) + N_{s,i}(t) + N_{o,i}(t)$. Buses are assumed to have one family of vehicles with external destinations and no generated "demand", as buses operate usually circular routes across the regions with small time-varying service frequencies. Equations (3-2) and (3-3) without the demand terms Q are sufficient for describing the dynamics of buses. Fig. 3.2 displays flow movements of buses and cars with parking choices in region i, with all state variables included.

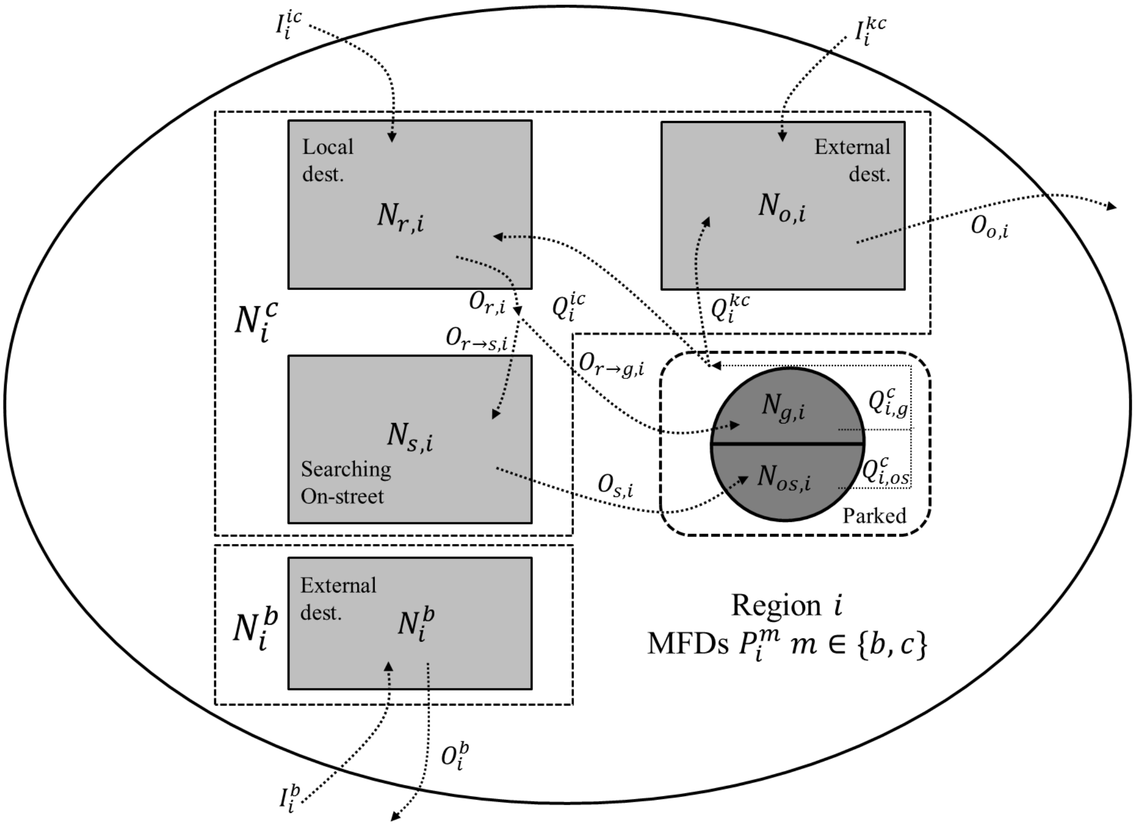
I_i^{ic}
I_i^{kc}
Local dest.
$N_{r,i}$
External dest.
$N_{o,i}$
$O_{o,i}$
N_i^c
$O_{r,i}$
Q_i^{ic}
Q_i^{kc}
$O_{r\to s,i}$
$O_{r\to g,i}$
$N_{g,i}$
$Q_{i,g}^c$
$N_{s,i}$
Searching On-street
$O_{s,i}$
$N_{os,i}$
$Q_{i,os}^c$
Parked
External dest.
N_i^b
N_i^b
Region i
MFDs P_i^m $m \in \{b, c\}$
O_i^b
I_i^b

Fig. 3.2 Flow movements in region network i with parking choices.

Assume $occ_i^c(t) = 1$ and constant, the flow conservation of the families I, II, III, and the two families of the parked cars can be written as follows and illustrated by Fig. 3.2:

$$N_{r,i}(t+1) = N_{r,i}(t) + Q_i^{ic}(t) + I_i^{ic}(t) - O_{r,i}(t) \tag{3-4a}$$

$$N_{s,i}(t+1) = N_{s,i}(t) + O_{r\to s,i}(t) - O_{s,i}(t) \tag{3-4b}$$

$$N_{o,i}(t+1) = N_{o,i}(t) + Q_i^{kc}(t) + I_i^{kc}(t) - O_{o,i}(t) \tag{3-4c}$$

$$N_{os,i}(t+1) = N_{os,i}(t) - Q_{i,os}^c(t) + O_{s,i}(t) \tag{3-4d}$$

$$N_{g,i}(t+1) = N_{g,i}(t) - Q_{i,g}^c(t) + O_{r\to g,i}(t) \tag{3-4e}$$

In Equation(3-4a), $Q_i^{ic}(t)$ is the internal demand from region i to i; I_i^{ic} is the total incoming flow from external regions to region i and ending their trips in region i, where $I_i^{ic} = \sum_{j\neq i} O_{j\to i}^{ic}(t)$. In Equation (3-4b), $O_{r\to s,i}(t)$ is the transfer flow from running (family I) to searching-for-parking (family II), $O_{r\to s,i}(t) = O_{r,i}(t) \cdot \omega_{os,i}^c(t)$ and $\omega_{os,i}^c(t)$ denotes the percentage of trip-finishing cars that pursing on-street parking. The term $Q_i^{kc}(t) = \sum_{k\neq i}\sum_{j\neq i} Q_{i\to j}^{km}(t)$ in Equation (3-4c) denotes the total demand generated in region i with external destinations k; $I_i^{kc}(t) = \sum_{k\neq i}\sum_{j\neq i} O_{j\to i}^{km}(t)$ is the total through transfer flow; while the last term $O_{o,i}(t) = \sum_{k\neq i}\sum_{j\neq i} O_{i\to j}^{km}(t)$ is the total outflow exiting region i. Equations (3-4d) and (3-4e) describe the dynamics of the parking flows, where $Q_{i,os}^{c}(t)$ and $Q_{i,g}^{c}(t)$ are the demand generated from on-street parking and garage parking, $Q_{i,os}^{c}(t) + Q_{i,g}^{c}(t) = Q_i^{ic}(t) + Q_i^{kc}(t)$.

The family-specific outflow $O_{x,i}(t)$, $x \in \{m, g, s, o\}$ can be obtained by Little's formula:

$$O_{x,i}(t) = \frac{N_{x,i}(t)}{N_i^c(t)} \cdot \frac{P_i^c(t)}{l_{x,i}} \tag{3-5}$$

where $N_i^c(t) = \sum_x N_{x,i}(t)$, and $P_i^c(t) = P_i^c(N_i^c(t))$ is estimated by (3-1). $l_{x,i}$ is the average trip distance of family x travelled in region i, where $l_{m,i} = l_{g,i} = l_{o,i} = l_c$ and l_c is network specific and given. $l_{s,i} = L_i(t)$ is estimated in the parking model introduced later .

Mode and parking facility choices

To determine the mode split of the newly-generated demand (between car and bus) a Nested-Logit model is applied based on the latest known trip disutility (costs) $C_i^{c,n}(t)$ for travelling with cars with parking facility choice $n = \{on\ street\ vs. garage\}$, and $C_i^b(t)$ for traveling with buses. For cars, trip disutility includes travel time, cruising delay and the costs of parking. While for buses, trip cost consists of travel time, accessing time and the discomfort of on-board overcrowding (see detail in Zheng and Geroliminis (2014)). In our case, the nest is required by the parking facility choice n. Denote $\omega_{g,i}^c(t)$, $\omega_{os,i}^c(t)$,the percentage of mode choice of cars with garage or on-street parking, and $\omega_i^c(t)$ and $\omega_i^b(t)$, the percentage of mode choice of cars and buses. We assume the travelers choose their mode of transport, either cars or buses, in the beginning of their trips. The estimation of bus share $\omega_i^b(t)$ is given by $\omega_i^b(t) = \frac{exp\left(\tau_b \cdot C_i^b(t)\right)}{exp\left(\tau_c \cdot C_i^b(t)\right) + exp\left(\tau_c \cdot C_i^c(t)\right)}$, where $C_i^c(t) = \frac{1}{\beta} \cdot ln \sum_n exp\left(\beta \cdot C_i^{c,n}(t)\right)$. Scale parameters $\beta, \tau_b, \tau_c$are calibrated to avoid oscillation in mode shift. By definition $\omega_i^c(t) = 1 - \omega_i^b(t)$. Given the total demand $Q_i(t)$, $Q_i^m(t)$ can be estimated by $\omega_i^c(t)$ and $\omega_i^b(t)$. Given the car demand generation $Q_i^c(t)$, $Q_{i,os}^c(t)$ and $Q_{i,g}^c(t)$ in Equation (3-4) can be obtained, as we can assume a fixed ratio that demand generated from the parking facilities (meanwhile releasing the parking space) is given.

Note that the car-users have limited knowledge on the possible availability of the on-street parking by the time they start their trip. Then if they chose to travel by car, when they get in the proximity of the destination, they re-evaluate the costs of on-street and garage parking and they make a new decision. If they decide to travel bus, they do not have such an option. Therefore the final choice on parking facility is determined on-site. While the a priori and the final decision are the same in the majority of the case, we consider such a framework more realistic. The distribution of the arriving flow $O_{r,i}(t)$ between the two parking facilities can be estimated trough a Logit model by $\omega_{g,i}^{c}(t)$ and $\omega_{os,i}^{c}(t)$, where $\omega_{os,i}^{c}(t) = exp\left(\beta \cdot C_{i}^{c,os}(t)\right) / \left(exp\left(\beta \cdot C_{i}^{c,os}(t)\right) + exp\left(\beta \cdot C_{i}^{c,g}(t)\right)\right)$. The production of $O_{r,i}(t) \cdot \omega_{os,i}^{c}(t)$ thus is the input to the cruising family $N_{s,i}(t)$.

Parking model

Consider a city network that has two typical parking facilities: on-street parking and garage parking. The on-street parking has total parking space A_i and the occupied parking space at time t is $N_{os,i}(t)$. The available parking space thus is $A_i - N_{os,i}(t) \geq 0$, while the garage parking is assumed to have infinite capacity without cruising delay. To reflect the impact of parking limitation on traffic, a parking model must be introduced. A bi-modal MFD model with unlimited parking such as the one in Zheng and Geroliminis (2013) considers two families of car movements: (I) cars moving towards internal destination, and (II) towards external destinations. We now extend the treatment to three families: (I) cars running towards internal destination but not yet search for parking $N_{r,i}(t)$, (II) cars reaching destination and searching for on-street parking space and (III) cars moving towards external destination regions $N_{o,i}(t)$. Cars reaching destination and successfully parked are denoted by $N_{os,i}(t)$ and $N_{g,i}(t)$ for on-street parking and garage parking, respectively. The interactions among the different families will be illustrated later. For simplicity, mode index c (for cars) is skipped in these family notations.

For family II, the cruising-for-parking process is considered to be repetitive Bernoulli trials, until an available parking spot is obtained, which is expressed by a geometric distribution (number of trials until the first success). The probability $\varphi_i(t)$ that a parking spot is available when being reached, is in direction proportion to the ratio between the available space $A_i - N_{os,i}(t)$ and the total parking space A_i (in this work we consider only one region with parking difficulties, but this can be easily extended to multiple regions)

$$\varphi_i(t) = \frac{A_i - N_{os,i}(t)}{A_i} \tag{3-6}$$

We assume the parking spaces evenly distributed around the destinations, with a spacing s_i. Alternatively one may apply a two-dimension cruising model, provided with detailed data on spatial distribution of parking spaces and the cruising behavior of cars.

Given the property of Bernoulli trial, it is apparent that the average number of parking spaces passed by cars in family II before finding an available parking space (number of trials before success) is $1/\varphi_i(t)$. The average cruising distance $L_i(t)$ thus can be obtained by Equation (3-7), while a similar model with exists in Geroliminis (2014):

$$L_i(t) = s_i \cdot \frac{1}{\varphi_i(t)} \tag{3-7}$$

Given the speed $v_i^c(t)$ from the MFD, the cruising delay $T_{cru,i}(t)$ can be obtained as:

$$T_{cru,i}(t) = \frac{L_i(t)}{v_i^c(t)} = \frac{s_i}{\varphi_i(t) \cdot v_i^c(t)} \tag{3-8}$$

Regarding the pricing of using the facilities, denote $p_{os}(t)$ and $p_g(t)$ for on-street parking rate and garage parking rate at time t. Cruising inside garages is neglected in this study.

3.2 Parking Management

Studies have demonstrated that the MFD modeling can contribute to develop traffic management strategies, examples including space allocation bus lanes of Chapter 2 and dynamic traffic signal perimeter control (Haddad et al. (2013), Aboudolas and Geroliminis (2013)). We propose two dynamic pricing schemes (strategy P1 and P2) to determine on-street parking pricing $p_{os}(t)$ and garage pricing $p_g(t)$ such that the congestion of traffic and parking is reduced. Pricing is only applied in the center region which experiences more congestion, i.e. a region index is omitted from the above pricing variables. Note that for the strategies under discussion, we assume to have full authority of both pricings whereas in reality they belong to parties with different operating objectives (and competition might occur). We will address this subject in the final part of the paper.

Strategy P1 develops a generic widely used control loop feedback mechanism. It tries to succeed two set points, related to (i) the maximum production of the network (in terms of vehicle-kilometers travelled per time interval) and (ii) small cruising time for on-street parking. The strategy calculates an error value as the difference between a measured process variable and the desired set point. It attempts to minimize the error by adjusting the prices $p_{os}(t)$ and $p_g(t)$.

Strategy P2 considers that the dynamic evolution of the system is known and solves a single full horizon optimization problem to minimize the total costs experienced by all passengers. While such an approach can be considered to provide close to system optimum conditions, it is considered an ideal scenario, which is also difficult to implement. Nevertheless, it provides an upper limit for comparison purposes with more feasible strategies like P1 or time-independent pricing.

A congestion- and cruising-responsive feedback parking pricing scheme (Strategy P1)

This pricing scheme should aim to reduce traffic congestion caused by cars and by cruising. Car users will pay a parking fee based on the magnitude of congestion they cause at the moment they enter the network. A feedback-type controller is employed to update the time-dependent prices $p_{os}(t)$ and $p_g(t)$, where the control variables are the total accumulation $N^c(t)$ and the accumulation of cruising family $N_{s,i}(t)$. The concept is that garage users have to pay for the hyper congestion due to large accumulations of cars in the network, while on-street users have to pay for the cost of cruising as well. In this way, both types of congestion can be eliminated. Equations (3-9) and (3-10) describe mathematically the two pricing control mechanisms respectively, where c_1 and c_2 are control gain parameters. Equation (3-9) states that the garage price for the next time interval $t+1$, $p_g(t+1)$, increases if the accumulation of car users $N^c(t)$ exceeds the critical accumulation of the network N^{cr} (where the maximum network production is reached and network production decreases if $N^c(t)$>N^{cr}). N^{cr} is derived from the MFD. Equation (3-10) indicates that the on-street price $p_{os}(t+1)$ charges the same amount for the reduction of network production, and an additional amount for causing cruising delay which is proportional to the difference between $N_s(t)$ and a pre-defined threshold N^{sT}. Having c_1 parameter in both (3-10) and (3-11) simply means that all car users have to pay for congestion independently of the parking choice. Parameter N^{sT} can be interpreted as the tolerated amount of cruising vehicles, a policy factor influencing the service level of on-street parking. We will show in a later graph how N^{sT} is chosen. Strategy P1 does not require any prediction and is based only on quantities that can be estimated with existing sensing techniques.

$$p_g(t+1) = p_g(t) + c_1(N^c(t) - N^{cr}) \tag{3-9}$$

$$p_{os}(t+1) = p_{os}(t) + c_1(N^c(t) - N^{cr}) \ + c_2(Ns(t) - N^{sT}) \tag{3-10}$$

A system optimum parking pricing scheme (Strategy P2)

The second parking pricing scheme aims at achieving a system optimum. The goal of this pricing scheme is to minimize the TPC that serves the total demand by optimizing the parking pricings in the center region controlled by a central operator with a full knowledge of the system, while such a strategy is difficult to be implemented due to prediction limitations and day to day variations of demand. Mathematically it can be described as the following:

$$\min_{p_{os}(t),p_g(t)} Z = \sum_{t,i,m} TPC_i^m(t) = \sum_{t,i,m} NP_i^m(t) \cdot T \tag{3-11}$$

where T is the duration of the time interval t. Problem (3-11) is subject to the system dynamics introduced in the previous section. We also add a constraint that $p_{os}(t) < \ p_g(t)$. The reason is the following. Since on-street parking requires cruising time, if $p_g(t)$ was not larger than $p_{os}(t)$, the cost of garage parking would be less and the optimization will consequently rule out the option of on-street parking. Pricing without this constraint will be evaluated as a future work. The optimization problem (3-11) is highly non-linear. We solve this problem by sequential quadratic programming. We apply this algorithm for multiple initial values (around 1000) to avoid convergence to local minima, which might be the case for a non-smooth objective function.

3.3 Case Study and Analysis

The proposed approach is tested in an idealized two-region bi-modal network. Mixed traffic of buses and cars occurs in the outside region (periphery), while an optimum fraction of road space in the center region (center) is pre-determined and dedicated to bus usage only (e.g. solution from Chapter 2). The radius of the center region is 1.6km and of the periphery is 3.2km. The road networks of the two regions are well connected at the border. We simulate an urban road traffic system for 4-hours (80 time units), a typical morning period. Demand has a symmetric trapezoidal shape with time and the length of peak period is equal to 1hr. A 70% fraction of the demand generated in the periphery will travel to the city center and 30% fraction of the demand generated in the center will travel to the periphery of the city. A mixed bi-modal MFD for the periphery and two single-mode MFDs for the center are utilized and given. Two modes of transport are considered available in the system, car and bus. A fraction 10% of the users are captive bus users and do not have access to cars. The network topology, signal settings and traffic parameters are as follows: two-lane one-way roads with block lengths l_{b1} =154m and l_{b2} =308m; signal settings for all intersections, green, $g = 40sec$; cycle, $c = 90sec$ and offset, $o = 0sec$. The average distance between parking spaces $s = 20m$. The duration of on-street parking is 1hr. A sensitivity analysis on parking duration will be reported as future work.

In the first sub-section, we show the resultant dynamics of the two-region bi-modal system under parking limitation and pricing, illustrating the mechanism of the modeling approach. Then we give performance comparison of four parking pricing policies: a base scenario with free on-street parking, a flat parking-rate scenario for both garage and on-street, Strategy P1 and Strategy P2. A discussion is followed in the final sub-section to address indications on parking policy from the results. The base scenario applies a time-constant parking garage fee p_g which creates crowding for the cruising traffic. The flat parking rate scenario estimates constant values of p_{os} and p_g by minimizing objective function (3-11).

3.3.1 System dynamics with parking pricing

Fig. 3.3 displays the system dynamics for the center region (the region index is skipped in the text), under the optimal constant pricing scheme where $p_{os} = 0.4\$/hr$ and $p_g = 1.6\$/hr$. In Fig. 3.3(a) and (d), it can be observed that mode share of buses increases during peak-hour as traveling with cars experience higher travel costs than buses. From the outflow-accumulation MFD in Fig. 3.3(b) (with the critical accumulation $N^{cr} = 5200$veh and a decreasing branch up to $N^c(t) = 9000$veh for car network), we can clearly confirm the car network experiences congestion. Judging from the time series of cruising delay in Fig. 3.3(c), limitation of parking contributes to the high travel cost of cars as well. Nevertheless, even if demand is high, on-street parking is not fully occupied because travelers have alternatives of lower total cost (e.g. public transport or garage with fee). Note that such a pricing scheme cannot fully avoid neither congestion nor cruising. The travel cost of buses remains nearly the same since buses are operated on the dedicated lanes with scheduled frequencies, although a slight increase can be found in Fig. 3.3(d) during the peak-hour, which is a reflection of speed reduction due to the longer dwelling time for boarding more passengers. Congested states can be observed in the MFD in Fig. 3.3(b), where different values of outflow occur for the same accumulation. The reduction of outflow can be explained by the reduction of on-street parking availabilities. As cars have to cruise longer distance to compete for a parking space, the outflow drops accordingly (outflow is production over trip length).

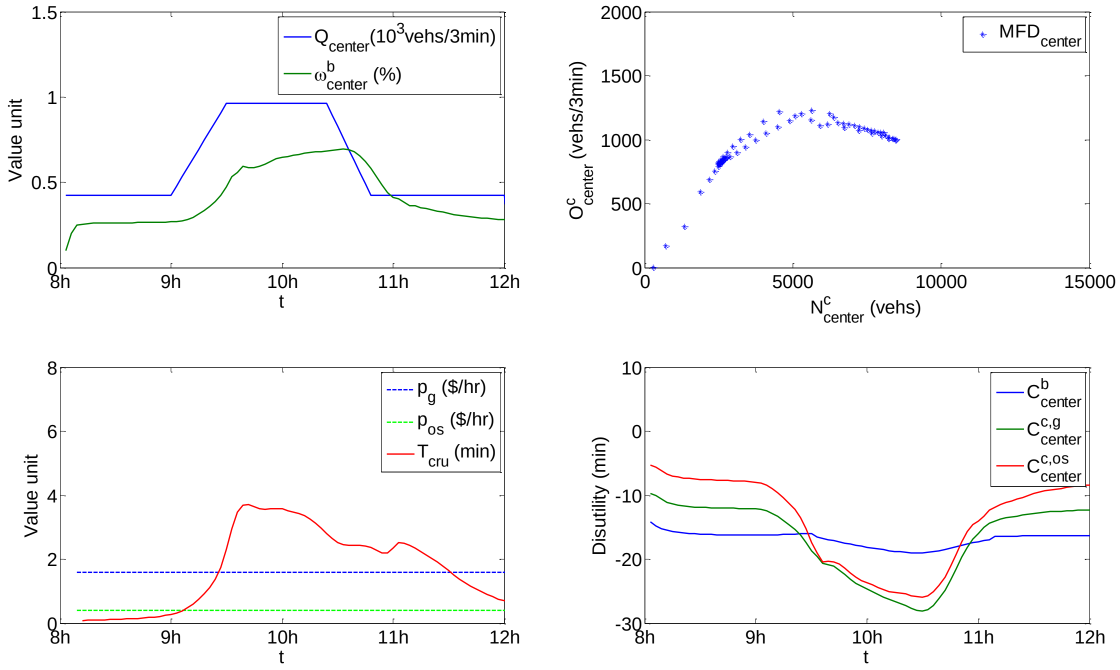

Fig. 3.3 System dynamics and traffic performance under optimal constant parking pricing scheme for the center region: (a-top left) mode share of bus $\omega^b{}_{center}$ and total travel demand over time, (b-top right) the MFD of the center region car network with congestion observed, (c-down left) the prices of p_{os} and p_g, and cruising-for-parking time, and (d-down right) the cost of traveling with each mode over time.

Fig. 3.4 illustrates the resultant system performance under the feedback-type time-dependent pricing Strategy P1. Prices are updated every 15min (5 interval units). In Fig. 3.4(c), the time-dependent pricing rates $p_{os}(t)$ and $p_g(t)$ are plotted where higher pricing rates are found for the peak hour. Fig. 3.4(a) plots the cruising condition of the basic scenario. For the applied Strategy P1, the objective is to control the cruising delay under 3min therefore we choose a value of 900veh for N^{sT}. Then shown in Fig. 3.4(d), the accumulations $N^c(t)$ and $N_s(t)$ fluctuate closely around the critical values N^{cr} and N^{sT}, though there are a few cases where congestion exceeds the desired states. This shows consistency with the expected system dynamics by feedback controllers.

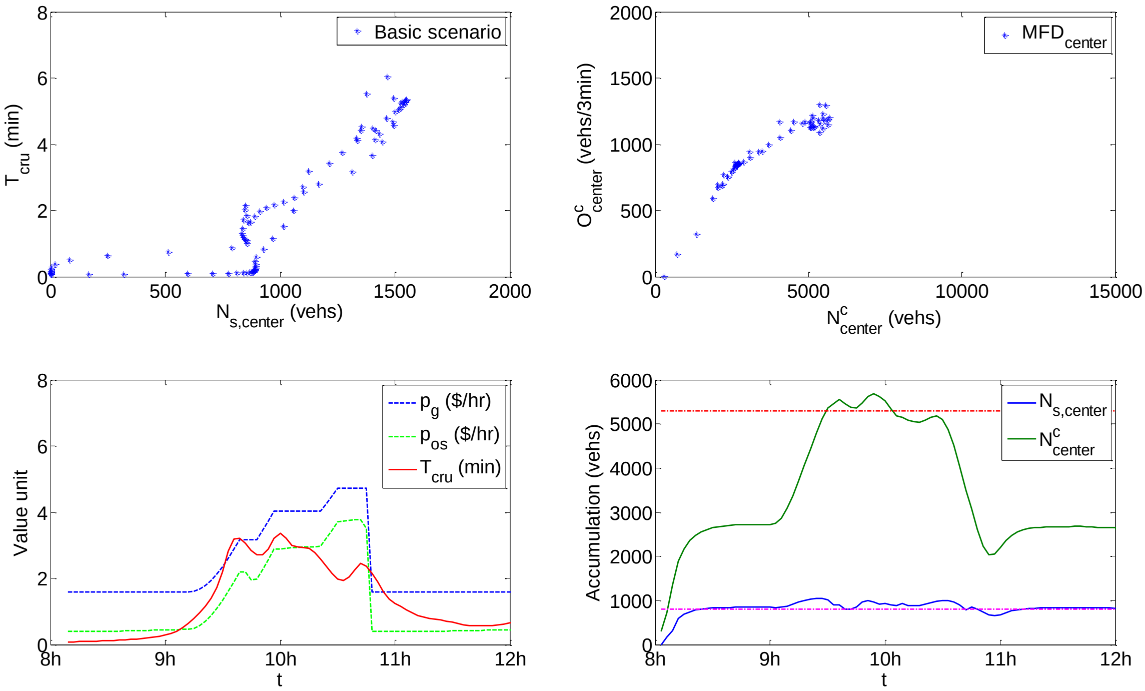

Fig. 3.4 (a-top left) Relationship between cruising delay T_{cru} and cruising vehicles $N_{s,center}$ of the basic scenario; System performance by Strategy P1: (b-top right) the MFD of the center region car network, (c-down left) the prices of p_{os} and p_g, and cruising-for-parking time T_{cru}, and (d-down right) the evolution of accumulations with the desired states in dotted lines: upper red for critical accumulation N^{cr} and magenta for the cruising threshold N^{ST}.

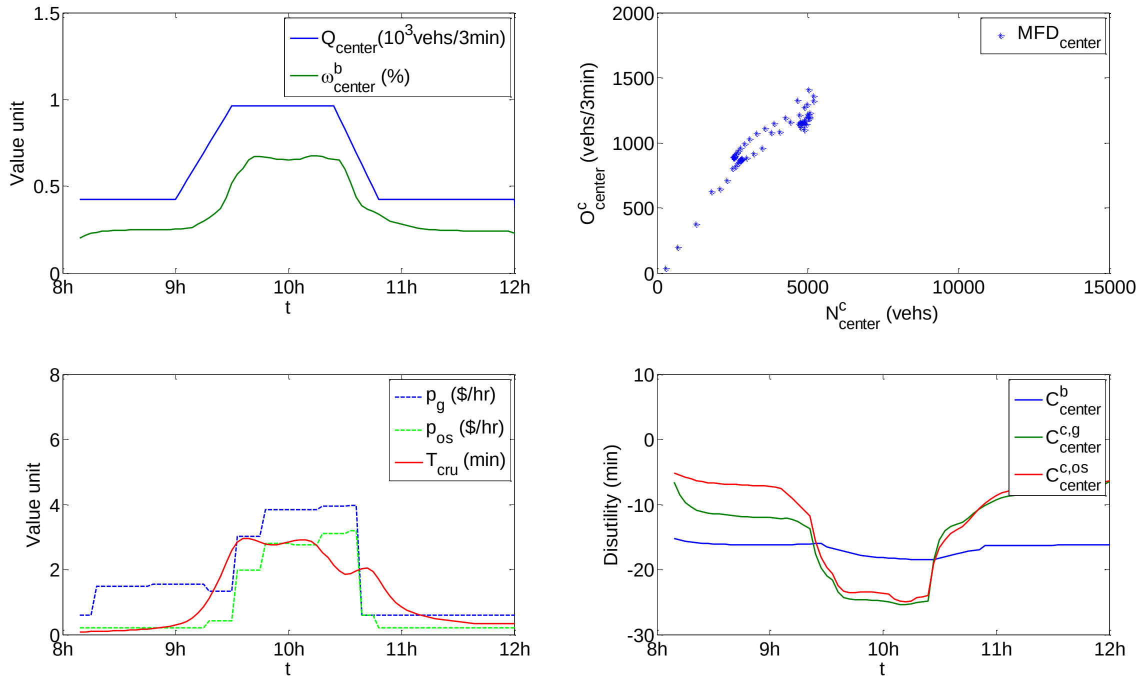

Fig. 3.5 System dynamics and traffic performance under the optimal Strategy P2. The same graphs are reproduced as in Fig. 3.3.

Now let us examine the performance under Strategy P2. The same graphs of Fig. 3.3 are reproduced and displayed in Fig. 3.5. Comparing to Strategy P1, the cruising delay and accumulation further decrease. Higher maximum outflow can be observed from the MFD. With a careful investigation on the resultant mode shares of bus, we conclude that the improvement of performance under Strategy P2 is due to its capability of triggering an earlier mode shift from cars to buses during the on-set of the peak-hour. Remarkably, such small change (roughly a mode share difference of 2% during 20min) in the mode share creates significantly different traffic performances. Furthermore, it should be highlighted that less total pricing is charged on the users in strategy P2.

3.3.2 Pricing scheme evaluation

The proposed modeling approach is capable of producing the physics of overcrowding in general traffic and on-street parking, the effect of cruise-for-parking, and the aggregated behavioral change (mode choices) given parking space limitation and cost. The second part of this case study is to test and attain efficient pricing strategies. We compare the resultant system costs among the 4 scenarios by the following performance indicators: the savings in total person hours travelled (PHT) comparing to the base scenario, the total toll paid (TTP, converted to time unit with a mean value-of-time of 16$/hr), the toll efficiency which is the ratio between the savings in PHT and the corresponding TTP, and the average cruising delay which is the total delay time divided by total amount of on-street parkers. The statistics are listed in Table 3.2.

Table 3.2 Comparison of the performances of different pricing schemes.

	PHT Savings (hrs)	Total toll (hrs)	Toll efficiency (PHT.sav / TTP)	Ave. cruising delay (min)
Base scenario	PHT=45337	660	0	3.7
Constant pricing	1211	2641	45.8%	1.9
Strategy P1	5944	7207	82.5%	1.4
Strategy P2	6990	5857	119.5%	1.2

The base scenario applies a flat rate to garage parking only. In this case, on-street parking is highly desired at first, as parking on-street experiences less cost until the cruising time cost is equivalent to garage pricing rate. However due to the large demand for on-street parking, on-street parking becomes full and results an average 3.7min cruising delay for the later travelers. For the studied network, the average travel time per trip is 6min under free-flow condition and

15min under congestion. A 3.7min cruising time is a relatively considerable amount. Small improvement in PHT can be found when flat-rate tolls are implemented and optimized for both garage and on-street parking. From an efficiency point of view, however, the constant-pricing strategy makes users pay two times more than they gain in the reduction of travel time, reflected by the toll efficiency value of 45.8%.

Applying Strategy P1, it is observed that the PHT significantly decreases. To achieve this, however, high parking tolls have to be charged. The resultant toll efficiency for achieving this PHT saving is improved to 82.5% albeit still indicating an insufficient toll scheme. Stimulating result is obtained by the implementation of Strategy P2, where the saving in PHT is higher than the required amount of toll payment.. Moreover, the cruising delay by Strategy P2 outweighs Strategy P1 by 15%. The difference in efficiency between the two pricing strategies P1 and P2 can be explained as the following. Since Strategy P1 is responsive, controllers (3-9) and (3-10) do not explicitly consider congestion evolution. While benefiting from the optimization, Strategy P2 is able to predict and adapt traffic conditions and set optimal tolls to trigger an early mode shift. As mentioned-above, a slight mode shift during the on-set of peak-hour could lead to substantial improvement. While strategy P2 is difficult to be implemented in reality as it requires future predictions, a careful analysis of historical data can possibly make this feasible through a model predictive control approach where uncertainty will be treated in the optimization horizon. This is a challenging but interesting future direction.

3.3.3 Parking policy indication

Time-dependent parking-rate represents the current mainstream parking price management in urban areas, though the rates may be determined based on different criteria. For instance, the public-owned parking facilities in the city of San Francisco operate demand-responsive parking pricing which updates the prices at a monthly or shorter basis to reduce cruising time. Strategies P1 attempts to employ such pricing mechanism with a second objective to reduce general congestion. The result shows that the strategy indeed effectively reduces the total PHT and average cruising time. Strategy P2 demonstrates higher efficiency thanks to the opportunities of adapting the changing conditions of both the users and the parking prices. Such opportunities in practice can be offered through online information system.

Note that in this study, we assume policy-makers operate the two parking facilities and have full control of the pricings p_{os} and p_g. In practice, the authorities belong to different operators, e.g. city policy-makers operate on-street parking while real-estate companies operate garage parking.

These operators can be considered that have different objectives, e.g. city tries to minimize the generalized cost of all users, TPC, while garage operator maximizes its profit. Then Problem (3-11) can be re-formulated as the follows:

$$\min_{p_{os}(t)} TPC = \sum_{t,i,m} (PHT_i^m(t) + Tos_i^c(t) + Tg_i^c(t)) \tag{3-12}$$

$$\max_{p_{og}(t)} BG = \sum_{t,i} Tg_i^c(t) \tag{3-13}$$

where $Tos_i^c(t), Tg_i^c(t)$ is the total toll collected at time t in region i from car users using on-street parking and garage parking, respectively.

It can be seen that direct conflict of interests, represented by the term $Tg_i^c(t)$ in Equations (3-12) and (3-13), exists between the two operators. Competition behaviour thus should be expected. Let us provide some preliminary result on the system performance under conditions with the existence of competition, while more detailed investigation will be reported in a later version of the paper.

Assume now that the two operators are cooperative with each other and they compete to maximize the common profit. Then solving Problems (3-12) and (3-13) can be considered as a bi-objective optimization problem. We apply standard optimization procedure, to obtain the efficient frontier. Two scale parameters valued (0, 1) are given to the two objective functions respectively, where the sum of the two parameters equals to 1. With the scale parameters, the problem is transformed into a single-objective optimization and different Pareto optimal solutions are produced. Fig. 3.6(a) displays the efficient frontier of this problem. With our system model, the potential combination of management pricing policies can be readily estimated.

Assume a second scenario that the two operators are selfish. The pricing competition is a responsive and negotiate-alike process, where each operator changes the pricing strategy after recognizing the impact of the other party's action. We consider solving such competition by a leader-follower optimization procedure, where the policy-maker as the higher level leads the process and the real-estate companies as the lower level follows. At each competition round, each operator tries to optimize its own objective given the action of the previous one. To avoid oscillatory behavior of the optimization procedure, we add constraints that the actions of the one optimization problem should not worsen the other problem by more than 20% compared to the previous step. Comparing to the existing applications of bi-level optimization in transport-related researches, such as in Yin (2000), the challenges here are two-fold: (i) the two objective functions have direct conflict, and (ii) the existence of equilibrium or a competition efficient frontier where the two parties cannot improve their profit. Result of a simulated competition is displayed in Fig. 3.6(b), which illustrates that a long-term equilibrium can be reached for the pricing strategies between the two operators. On-going work makes further effort in the investigation of convergence towards this direction.

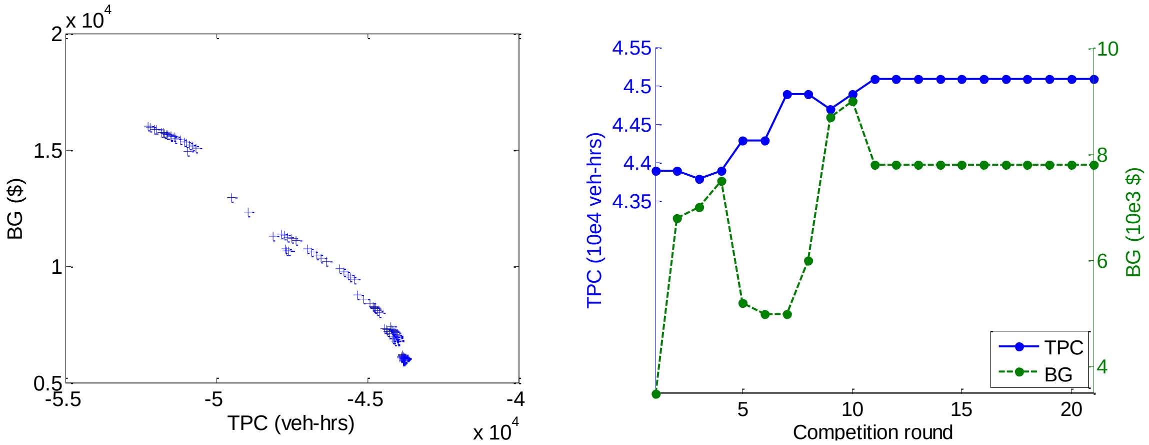

Fig. 3.6 (a) The efficient frontier between maximizing the system performance (minimizing total cost TPC) for the on-street parking operator and maximizing BG the revenue for garage parking operator; (b) The evolution of the costs of TPC and BG over 20 rounds of leader-follower pricing competition

3.4 Summary

Following Chapter 2, we proposed in Chapter 0 a macroscopic modeling approach for modeling multi-modal traffic system with parking limitation and cruising-for-parking flow. Parking limitation was integrated in the developed multi-modal system model, where vehicles need to cruise for parking before reaching destination. The time of cruising was estimated by assuming the probability of finding a parking space follows a geometric distribution and depends on the dynamic parking availability. The effect of cruising on the global performance, e.g. the average speed, was also captured, by the MFD dynamics. Case study was carried out in the same two-region bi-modal network as in Chapter 2. Two parking choices were added: limited on-street parking requiring cruising, and unlimited garage parking with higher parking fee but no cruising cost. The resultant system behavior under parking limitation and pricing were consistent with the common expectations. We then used the system model to test two network-level parking pricing strategies. One strategy adapted a feedback-type controller for determining the parking price, which was congestion- and parking availability-dependent. Applying this pricing strategy, traffic performance was maintained at desired (controlled) levels. The second strategy was obtained through optimization of the total cost (PHT + parking fee). Applying this strategy, the total cost was further reduced, as the prices were determined with long-term impact taking into account. Inspired by this result, we investigated the impact of competition on the performance of the pricing strategy, assuming the authorities of on-street and garage parking belong to different parties who manage prices with different goals. We presented preliminary results of cooperative-competition via a bi-objective optimization, while a bi-level optimization framework was proposed to simulate a responsive and negotiate-alike parking pricing market.

4 A Three-dimensional Macroscopic Fundamental Diagram for Mixed Bi-modal Networks

In this chapter, we shift our focus from system modeling and optimization to the analysis of traffic flow characteristics, and seek for an approach for capturing the bi-modal traffic congestion patterns and the individual modal impact on the global network performance. Furthermore, the approach is expected to provide insight into the dynamics of passenger flows and identify types of heterogeneity that may exist. If MFD-type of models would hold for further scrutiny in multimodal networks and enable the aforementioned investigations, network-level control strategies can be developed to maximize urban mobility. Despite the dynamics of traffic flow in bi-modal networks are more complicated due to the operational characteristics of buses and the interactions between cars and buses, simulation studies from previous study (Gonzales et al. (2011)) and the results of Chapter 2 on small networks showed that a classical MFD applies (under certain conditions) for bi-modal urban networks. However, the influence of each mode in the network dynamics and performance cannot yet be quantified. Understanding this relationship will facilitate the development of control strategies at various levels, e.g. bus signal priority controls, or redistribution of urban space as proposed in Chapter 3.

This chapter is organized as follows. Section 4.1 investigates the existence and the properties of a three-dimensional vehicular MFD (3D-*v*MFD) relating the accumulation of cars and buses, and the total circulating vehicle flow in a bi-modal network based on data collected from the simulation of the San Francisco network. A three-dimensional passenger MFD (3D-*p*MFD) is obtained as well, relating the accumulation of cars and buses and the passenger flow. Then, an analytical form of the 3D-*v*MFD is estimated. A Bus-Car Unit equivalent value representing the bi-modal mode conflict pattern is also derived from the function form of the 3D-*v*MFD. Section 4.2 presents an analytical approach to approximate the shape of the 3D-*p*MFD, given the 3D-*v*MFD. In Section 4.3, the existence of heterogeneity with respect to mode composition and congestion distribution is investigated. A network partitioning algorithm is proposed to deal with such heterogeneity. In Section 4.4, the application result of a 3D-MFD based perimeter flow control is illustrated and discussed. The last section summarizes the findings of this chapter.

A nomenclature is provided in the table below for the main variables and parameters used in this chapter.

Table 4.1 Nomenclature of the main variables and parameters used in Chapter 4.

Variables	Description
$Q_m \mid P_m$	Circulating vehicle (passenger) flow in a mixed network by mode m
N_m	The accumulation of mode m currently in the network
$Q_m(N_c, N_b) \mid P_m(N_c, N_b)$	Vehicle (passenger) flow of mode m in function of the bi-modal accumulations
$q_{im}(t) \mid k_{im}(t)$	Flow (density) of mode m on link i in the network at time t
l_i	Link length of link i
L	The total network length (lane-kms)
L_i^m	The total network length (lane-kms) of partitioned region i for mode usage m
$\tilde{n}_c$	The critical value of car accumulation (given a specific bus operation)
$\hat{k}_b$	The critical value of bus density for dedicated-bus-lane network
$h_m(t)$	The average passenger occupancy on mode m in the network
$BCU(n_c, n_b)$	Bus-Car Unit equivalent value in function of the bi-modal accumulations
δ_i	Mode composition ratio (for a bi-modal network) in the partitioned region i

4.1 A City-scale Three-dimensional MFD for Bi-modal Traffic

In this section, we investigate the relation among accumulation of cars and buses and circulating flow (both vehicle and passenger) in bi-modal traffic networks via simulation experiments. We show that area of Downtown San Francisco exhibits a city-scale three-dimensional vehicle MFD (3D-*v*MFD) relating the accumulation of cars and buses to flow with moderate scatter under different bi-modal demand patterns. We further show that our test site exhibits a city-scale three-dimensional passenger MFD (3D-pMFD) relating accumulation of cars and buses, to the passenger flows in the bi-modal network.

An empirical study on the 3D-*v*MFD and 3D-pMFD is hard to carry out due to the difficulty on data availability. We therefore perform our analysis with data collected from a microscopic traffic simulation environment AIMSUN. The test site is a 3 km^2 area of Downtown San Francisco including about 100 intersections and 430 links of total 100 lane-kms. The number of lanes varies from 2 to 5 lanes and the free flow speed is around 45 km/h. Traffic signals are all multiphase fixed-time operating on a common cycle length of 100s for the west boundary of the area and 60s for the rest. The traffic flow in the (bi-modal) network comprises two vehicle classes, i.e., passenger cars and buses. Bus routes and frequencies for lines in the studied network have been obtained from the San Francisco Municipal Transportation Agency (SFMTA).

4.1.1 A three-dimensional MFD for vehicle flows (3D-*v*MFD)

Let us denote by n_c the accumulation of cars and n_b the accumulation of buses at a specific time interval, Q_m is the total network circulating flow for each mode respectively, $m \epsilon \{b, c\}$ and Q the total network flow (in vehicle per unit time), which is the sum of car and bus circulating flows, all estimated during the same interval. Q_m is defined in accordance to Edie's definitions as $Q_m = \sum_i l_i q_{im} / L$, where l_i is the length of link i, q_{im} is the flow of link i at interval t for mode m and L is the average link length (Edie, 1963). For the developed model, the flow Q in the bi-modal network is considered to be a function of n_c and n_b given by (time interval t is omitted from the equations for simplicity)

$$Q = Q(n_c, n_b) = Q_c(n_c, n_b) + Q_b(n_c, n_b) \tag{4-1}$$

To obtain the shape of (4-1) we perform extensive simulation experiments in the test network with time-dependent asymmetric origin-destination tables, starting from different initial compositions of the bi-modal traffic (pairs of n_c and n_b). The initial profile for cars is a typical peak-hour demand with a trapezoidal shape. For buses, the demand is determined by the number of lines and their operational frequency using data from SFMTA. Higher demand scenarios are also analyzed to generate various mode compositions, i.e. n_c/n_b ratios. The simulation horizon for each experiment is 5.5 h and pairs of data (n_c, n_b) are gathered every 5 min from the simulator. For each 5min interval, Q_b, Q_c are estimated as described above. It should be noted that more than 20 scenarios with different dynamic demand characteristics are carried out to generate the corresponding traffic performance Q for a wide range of n_c and n_b. Some scenarios will be carefully analyzed later.

Fig. 4.1(a) illustrates the 3D-*v*MFD for bi-modal traffic, $Q(n_c, n_b)$. As a first remark, Fig. 4.1 (a) confirms the existence of a 3D-*v*MFD like-shape for bi-modal networks, whose shape is seen to depend on the accumulation of both cars and buses. To enable a better understanding of this figure, Fig. 4.1(b) displays the contour plot of the 3D-MFD on the (n_c, n_b) plane, using Delaunay triangulation/interpolation algorithm (De Berg et al., 2008) to estimate flow in a continuous space of the accumulation plane. The triangle in this figure indicates (n_c, n_b) pairs, where the network operates close to the maximum throughput of the 3D-*v*MFD (values within the 80% range of the maximum flow). In particular, it captures the "optimal operational regime" of bi-modal traffic. City managers and practitioners could seek to derive management strategies to operate at this optimal regime. It can be seen that the flow Q decreases monotonically as n_c and n_b increases, albeit with different slopes. Remarkably, the slope of buses is higher that the slope of cars. This indicates that effect of an additional bus in the network is much different than an additional car. A simple explanation is that buses make additional to traffic signal stops for passengers, and negatively influence traffic and create stop-and-go phenomena. The figure also depicts critical accumulations of cars $\tilde{n}_c$ (rising line in the triangle) where Q reaches its maximum for different values of n_b. The slope of the rising dashed line reflects that the capacity to serve cars has to be compromised in order to serve more buses, approximately $\tilde{n}_c \cong 3500 - 0.5n_b$. As a general remark, the 3D-*v*MFD can be used by policy makers to exploit the trade-off between the operation of buses and cars and design more sustainable cities. Note also that the maximum value of the network flow occurs for $n_b = 0$ and $n_c = 3500$ because of the effect of bus stops. If one ignores the higher passenger bus occupancy, an optimal strategy to maximize vehicle flow is to operate at the (3500,0) pair. As we will show in the next section, a consideration of different vehicle occupancies for buses and cars and the estimation of network passenger flow will produce a completely different result.

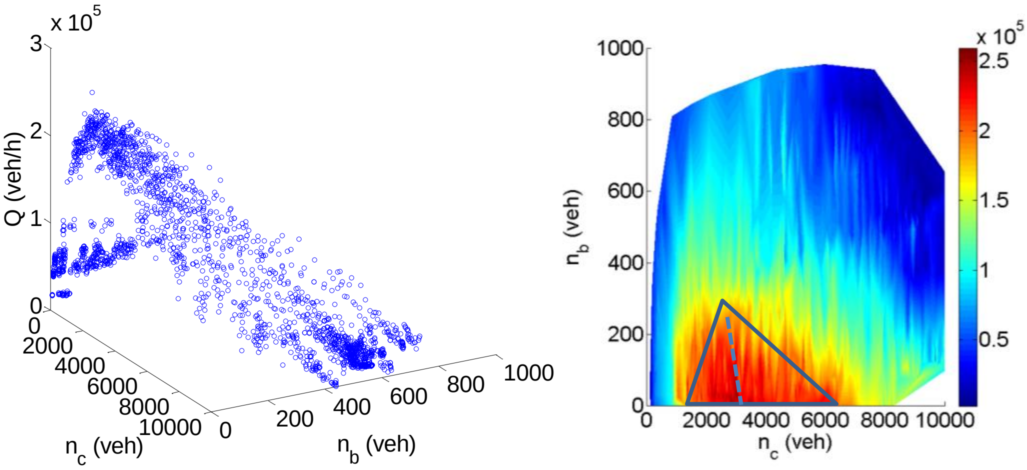

Fig. 4.1 (a) The 3D-*v*MFD points for bi-modal traffic and (b) contour plot of the 3D-*v*MFD after interpolation by Delaunay triangulation. The color represents the value of circulating flow $\boldsymbol{Q}$ (vehs/h).

4.1.2 A three-dimensional MFD for passenger flows (3D-*p*MFD)

Given the 3D-*v*MFD in Fig. 4.1, incoming flow can be controlled at the boundary of the network in order to direct the network operates at its "optimal operational regime". Existing methodologies for car-only perimeter control can be found in Keyvan-Ekbatani et al. (2012), Geroliminis et al. (2012), Haddad and Geroliminis (2012), Aboudolas and Geroliminis (2013) and elsewhere. However, these types of strategies are not able to consider that buses are more efficient modes due to higher passenger occupancy and they might restrict bus flow from entering the network. Before developing traffic management strategies for bi-modal systems with preferential treatment of buses, dynamics of passenger flows have to be investigated.

Denote P the passenger flow in the bi-modal network, with n_c and n_b. We are interested in observing $P = P(n_c, n_b)$. Denote h_m the average number of on-board passenger occupancy per vehicle, $m\epsilon\{b, c\}$. We assume that car occupancy h_c is constant $h_c = 1.3$, while we estimate bus occupancy h_b as a function of the dwell times. Dwell times are time-dependent, with three periods: on-set, during and off-set of peak hour. Dwell times are stochastic following a normal distribution, so to capture some uncertainty in passenger demand (in the simulation, only dwell times are available). Detailed description can be found in Section 2.1. Then, by definition, total passenger flow P can be expressed as

$$P(n_c, n_b) = h_c Q_c(n_c, n_b) + h_b Q_b(n_c, n_b) \tag{4-2}$$

We estimate passenger flow via the simulated data, with the measurements of car flows Q_c, bus flows Q_b and the estimated bus occupancy h_b. We use the same simulated flow data as the ones used to construct Fig. 4.1(a). Fig. 4.2(a) illustrates the 3D-*p*MFD relating accumulation of cars and buses to passenger flow. This figure presents an 3D-MFD like-shape for passenger flow, which looks similar with Fig. 1(a), but with higher maximum value, due to the h_b term in Equation (4-2). Nevertheless, a more careful look through the contour plot will reveal significant changes.

Fig. 4.2(b) depicts the resulting contour plot of P on the (n_c, n_b) plane after applying the same triangulation algorithm as in Fig. 4.1(b). The polygon in this plot captures the "optimal operational regime" of passenger flow of bi-modal traffic. It can be seen that the shape of the region is significantly different from the one observed in Fig. 4.1(b). More precisely, for a given n_c passenger flow P first monotonically increases as n_b increases to a critical $\tilde{n}_b$ and then decreases for $n_b > \tilde{n}_b$. Thus, having buses in the network will significantly increase the efficiency of the system but overloading buses will eventually cause delays for all vehicles and reduce passenger throughput. The figure also displays that $\tilde{n}_c$ in this case (rising dotted line in the polygon) has a clear tendency of becoming smaller as n_b increases, $\tilde{n}_c \cong 3500 - 6n_b$. The slope of the rising line reflects that only a slight increase of buses can allow a large reduction of cars to maintain the same passenger flow. The maximum network flow occurs now for a positive n_b, the (2800, 120) pair. It can be foreseen that more buses can be deployed in the network to succeed a higher passenger flow if dedicated bus lanes are provided.

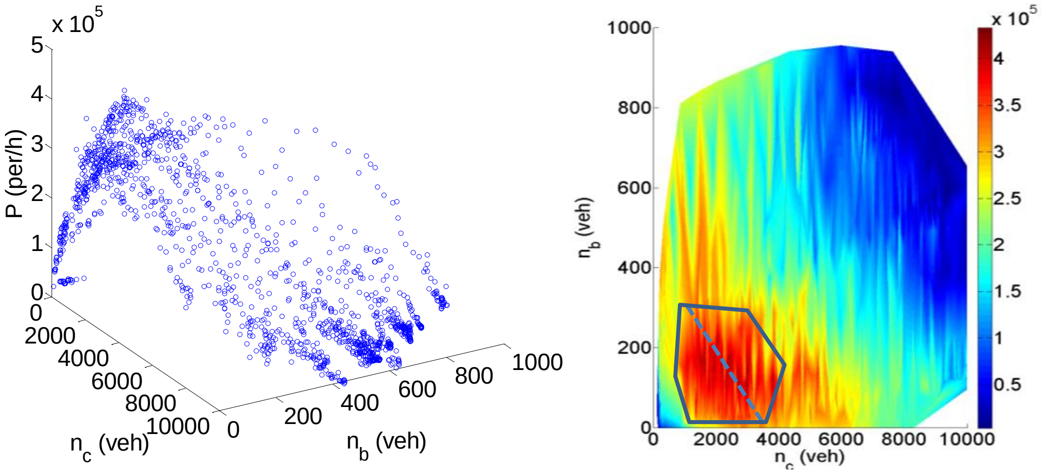

Fig. 4.2 (a) The 3D-*p*MFD for bi-modal passenger traffic, and (b) contour plot of the 3D-*p*MFD, the color represents the value of flow ***P*** (pers/h).

4.1.3 A functional form of the 3D-MFDs

Given that evaluating $Q(n_c, n_b)$ in Equation (4-1) for many pairs of n_c and n_b is tedious, we propose instead using an analytical function that approximates the 3D-*v*MFD in Fig. 4.1(a). Given the properties of $Q(n_c, n_b)$ described in the previous section, we consider an exponential-family function, which has useful algebraic properties. In general exponential families are in a sense very natural distributions to consider with plenty of applications in physics and statistics. Note that parabolic families do not provide a good fit given their single value of the 2nd derivative. To this end, we consider the following exponential flow function for data fitting:

$$Q(n_c, n_b) = a(n_c + n_b)e^{bn_c^2 + cn_b^2 + dn_c n_b + en_c + fn_b} \tag{4-3}$$

where a, b, c, d, e, f are model parameters. The parameter values should be specified so as to minimize the deviation of model (4-1) from the corresponding measured values $\boldsymbol{Q'}$. This function can be considered as a generalization of the Drake's exponential function for a single-mode fundamental diagram between flow and density. An unconstrained estimation will not be consistent with the physics of traffic thus constraints (4-4b) are added and described later. To this end, we estimate the values by Least-Squares parameter estimation for the given simulated data Q' in Fig. 4.1(b). The parameter estimation problem is formulated as follows (P1):

$$\min_{a,b,c,d,e,f} Z = \|\boldsymbol{Q} - \boldsymbol{Q'}\|^2 \tag{4-4a}$$

subject to

$$\begin{cases} Q \geq 0, \;\; for \quad 0 \leq n_c \leq max(n_c) \;\; and \;\; 0 \leq n_b \leq max(n_b), \\ \qquad \dfrac{\partial V(n_c, n_b)}{\partial n_c} \leq 0, \qquad for \;\; 0 \leq n_b \leq max(n_b), \\ \qquad \dfrac{\partial V(n_c, n_b)}{\partial n_b} \leq 0, \qquad for \;\; 0 \leq n_c \leq max(n_c), \end{cases} \tag{4-4b}$$

where V is the space-mean speed in the network, $\boldsymbol{Q}$, $\boldsymbol{Q'}$ are vectors with elements of the model (4-3) and the simulated data for each 5 min sample interval, respectively. Variable $V(n_c, n_b) \overset{\text{def}}{=} QL/(n_c + n_b)$, where L is the average link length of the network. The first constraint in (4b) guarantees non-negative flows for values of accumulation between zero and jammed. The second and the third constraint emphasize that the space-mean speed of all vehicles should decrease monotonically as n_c and n_b increase.

The parameter estimation problem (P1) is nonlinear and is solved through the Sequential Quadratic Programming (SQP) approach. SQP method solves a sequence of optimization sub-problems, each of which optimizes a quadratic model of the objective subject to a linearization of the constraints. We apply this algorithm for multiple initial values (around 1000) to avoid convergence to local minima, which might be the case for non-smooth data. The estimated parameters resulting from the optimization problem are: $\{a, b, c, d, e, f\} = \{1.95 \cdot 10^2, -2.34 \cdot 10^{-9}, 5.28 \cdot 10^{-7}, 6.34 \cdot 10^{-8}, -2.92 \cdot 10^{-4}, -1.50 \cdot 10^{-3}\}$. Note that as accumulation terms n_c and n_b have a unit scale of 10^3, the parameters b, c, d, e, f appear to have low values. Result of T-test indicates that all six parameters are statistically significant (p-value almost zero). Fig. 4.3(a) illustrates the results of fitting model (P1) with the estimated parameters to the simulated data. An R-square value* of 0.91 (close to 1) indicates that the resulting 3D-*v*MFD fits well with the data and all physical constraints (4-4b) are satisfied. We also verify the fitting performance of the model by 3 additional datasets, which were not used for model estimation and have different demand profiles. The average R-square value for these 3 datasets is 0.85. Fig. 4.3(b) depicts the contour plot of the 3D surface on the (n_c, n_b) plane. Comparing Fig. 4.3 (b) with Fig. 4.1(b), we can see that most patterns observed closely matches each other except the area for very high values of buses $n_b > 600$ due to: (i) the 3D surface does not reach zero when the values of n_b are extremely high, a typical phenomenon with the Drake's exponential formulation, and (ii) lack of simulated data for this area. Nevertheless, the traffic states reflected by the area cannot be easily observed in real systems. Moreover, it can be seen that the "optimal operational regime" (triangle in Fig. 4.1(b)) of bi-modal traffic is reproduced in a very similar way.

We perform a similar optimization for fitting the measured passenger flow P with parameters $a', b', c', d', e', f', g'$ for $P(n_c, n_b) = a'(n_c + g'n_b)e^{b'n_c^2 + c'n_b^2 + d'n_c n_b + e'n_c + f'n_b}$. Denote this parameter estimation problem as (P2). Solving (P2), we get $\{a', b', c', d', e', f', g'\} = \{3.46 \cdot 10^2, 6.41 \cdot 10^{-10}, -2.27 \cdot 10^{-6}, -1.14 \cdot 10^{-7}, -3.77 \cdot 10^{-4}, -5.30 \cdot 10^{-4},\ 3.66 \cdot 10^0\}$. Fig. 4.3(c) depicts the fitted model. The R-squared value of 0.91 (close to 1) indicates a decent model fitting, as well. Fig. 4.3(b) depicts the contour plot of the 3D surface on the (n_c, n_b) plane. Nevertheless, by comparing the parameters of the two models, no physical interpretations can be made. In the next section we will provide an analytical derivation to estimate $P(n_c, n_b)$ from $Q(n_c, n_b)$, without the need for curve-fitting, that will unveil interesting properties of network traffic flows.

---

* The R-square value is calculated with its general form $1 - SS_{Res}/SS_{Tot}$, where SS_{Res} is the residual sum of squares $SS_{Res} = \sum_i (Q_i' - Q_i)^2$ and SS_{Tot} the total sum of squares $SS_{Tot} = \sum_i (Q_i - \bar{Q_\iota})^2$.

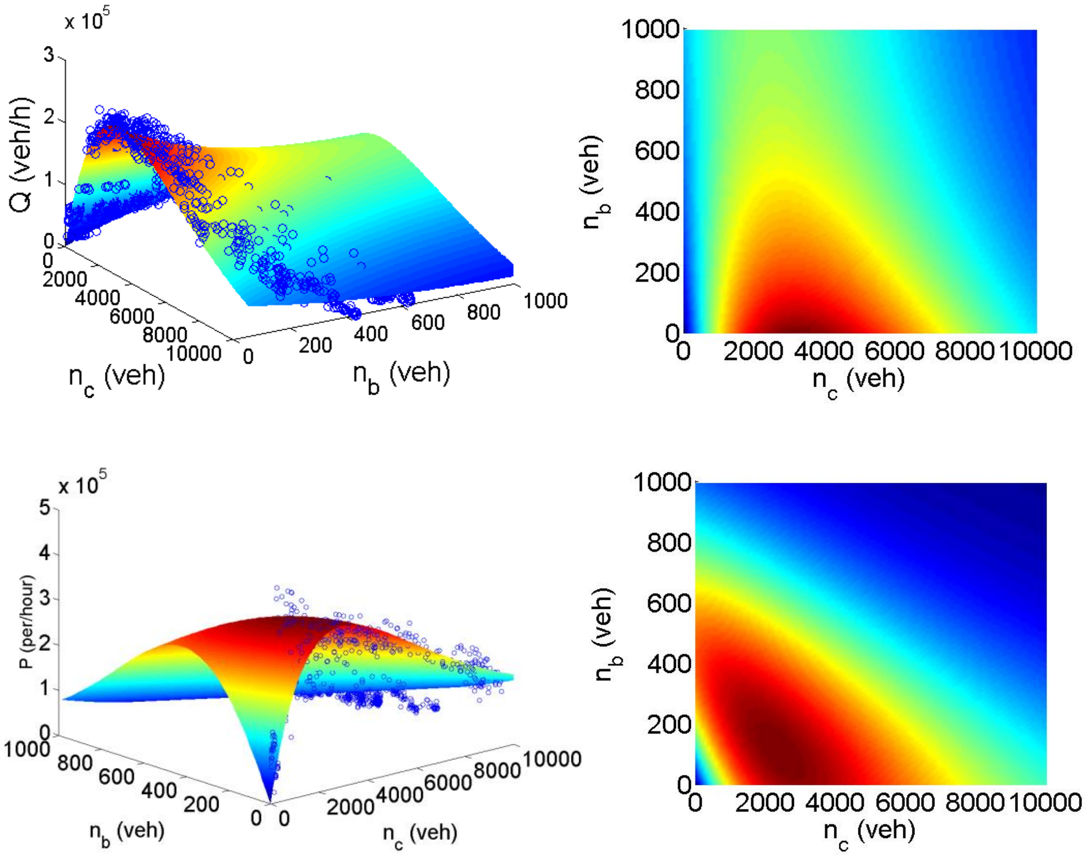

Fig. 4.3 (a) The exponential 3D surface plot of $Q(n_c, n_b)$, (b) the contour plot of $Q(n_c, n_b)$; (c) The exponential 3D surface plot of $P(n_c, n_b)$ and (d) the contour plot of $P(n_c, n_b)$.

4.1.4 Bus-Car Unit (BCU) Equivalent

In case of bi-modal systems, it is a common approach to assume that each bus is equivalent to a constant number of cars independent of the level of congestion and the densities of each mode and simplify the derivations of different performance measures. Nevertheless, it is clear from the explanation of the previous section that the influence of buses in the network performance depends on the state of the system (n_c, n_b) and cannot be considered constant. Thus, a static BCU consideration is insufficient to reflect the dynamic relationship between bus/car accumulation and congestion. For this reason, we propose here an analytical estimation of the BCU value as a function of the accumulation of both modes.

Let's consider that average network speed is $V(n_c, n_b) \overset{\text{def}}{=} Q(n_c, n_b)L/(n_b + n_c)$. We assume that for a given n_b, $V(n_c, n_b)$ is monotonically decreasing function of n_c and the same assumption for n_c (see the 2nd and 3rd constraint of Equation (4-4b)). Given that $V(n_c, n_b)$ is a continuous and differentiable function, we can estimate the effect that buses and cars have in speed V for a given (n_c, n_b) pair. Mathematically speaking, BCU is defined and estimated as the ratio of partial derivatives with respect to $n_c$and n_b

$$BCU(n_c, n_b) = \frac{\frac{\partial V(n_c, n_b)}{\partial n_b}}{\frac{\partial V(n_c, n_b)}{\partial n_c}} \tag{4-5a}$$

By assuming the closed form solution of $Q(n_c, n_b)$ in (3) and after some manipulations we obtain

$$BCU(n_c, n_b) = \frac{2cn_b + dn_c + f}{dn_b + 2bn_c + e} \tag{4-5b}$$

Fig. 4.4 plots the $BCU(n_c, n_b)$, a monotonically decreasing function of vehicle accumulations Physically speaking, Fig. 4.4 highlights that the effect of buses in the network speed is smaller as the network is more congested. This is because dwell times at bus stops are a smaller fraction of car travel time compared to free-flow conditions. For example, in case of 10% bus accumulation, BCU values range between 5.0 for free-flow conditions to 3.5 for highly congested conditions.

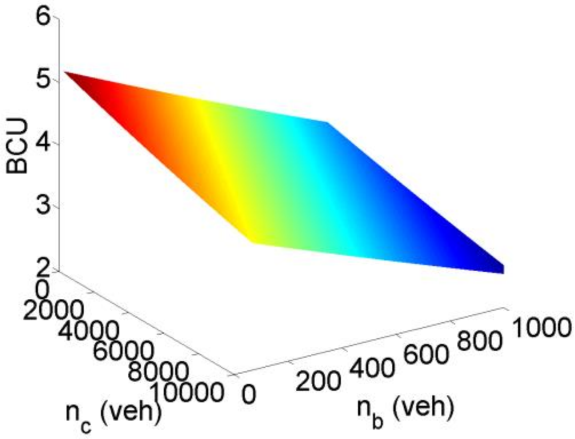

Fig. 4.4 A 3D diagram relating accumulation of cars and buses with BCU.

Nevertheless, Equation (4-5a) might not be straightforward to estimate with real data without curve fitting, due to the scatter of the MFD measurements. An approximation of (4-5a) based on engineering principles, $BCU^*(n_c, n_b)$, can be obtained by (4-5b) if we consider for a given pair of (n_c, n_b), an equivalent system with cars only that satisfies $V(n_{c0}, 0) = V(n_c, n_b)$, where n_{c0} is the accumulation of cars for a car-only network, i.e. $n_b = 0$. Thus, $BCU^*(n_c, n_b)$ is the solution of the equation $n_{c0} = n_c + BCU^* \cdot n_b$. In other words, BCU expresses an equivalent system with cars only that has the same space-mean speed with the mixed-traffic network. Then the average speed can be rewritten as:

$$V(n_c + BCU^* \cdot n_b, 0) = V(n_c, n_b) \tag{4-6a}$$

Combining the functional form (4-3) and Equation (4-6a), we obtain an analytical solution of BCU as a solution of the quadratic equation $(bn_b BCU^{*2} + (2bn_c + e)BCU^* - cn_b - dn_c - f = 0)$

$$BCU^*(n_c, n_b) = \frac{-(2bn_c + e) - \sqrt{(2bn_c + e)^2 + 4bn_b(cn_b + dn_c + f)}}{2bn_b} \tag{4-6b}$$

Given that the value of b is relatively small, BCU estimation can be approximated (error less than 0.01 by the solution of $eBCU - cn_b - dn_c - f = 0$ and therefore gives the BCU^* a plane shape). These results also indicate that the BCU^* value for buses becomes smaller as congestion increases. While this result might be considered as counter-intuitive, it is consistent with the physics of traffic because the effect of bus related stops is a smaller part of the total travel time for congested conditions.

4.1.5 An MFD for buses operating in dedicated lanes

As dedicated bus lane allocation becomes an important strategy for multimodal urban transport system with preferential treatment of higher occupancy modes of transport, it deserves the effort to look into a city-wide MFD for buses running on dedicated bus lanes. While a detailed analysis is beyond the scope of this thesis, we perform an analysis where all the bus lines perform in dedicated lanes. To this end, the existing test site is updated and one lane in each road of the city center with public transport lines is dedicated to buses if space is available. The resulted bus-lane network is well connected and includes about 5 km of dedicated bus lanes. To derive and investigate the shape of the MFD for buses, simulations are performed with a field-applied fixed-time signal control plan and fixed number of bus lines. To account for demand effects of the bi-modal traffic composition, ten runs were carried out for a 4-h time-dependent scenario with different car demand and bus frequency. Additionally, two scenarios (with five runs each) based on different dwell time of buses were defined in order to investigate the impact of the dwell time in the shape of the MFD for buses. Fig. 4.5 displays the MFD for buses resulting for the aforementioned runs. This figure plots the network flow-density relationship (buses/5min vs. buses/km) in the network for the whole simulation time period. Each measurement point in the diagram corresponds to 5 minutes. Fig. 4.5(a) confirms the existence of an MFD for buses with moderate scatter across different scenarios. Note that the value of critical density and maximum flow are considerably smaller compared to the ones of the vehicle MFD (critical density=7 buses/km and capacity=70 buses/hr).

It can be seen that the free flow speed in the bus lanes is around 25 km/h and the maximum flow occur in a density range from 5 to 10 buses per km. If the density is allowed to increase, then the dedicated bus lanes become severely congested. The flow decreases with density (negative slope) and the network can lead to gridlock. This MFD like-shape for buses operating in dedicated lanes is quite conforming to the MFD for cars and mixed traffic. An interesting observation is that the diagram indicates a high flow scatter for the critical bus density $\hat{k}_b$. This difference is attributed to different traffic patterns for the cars and the interaction of the bi-modal traffic at the intersections that could lead to partially block of the dedicated bus lanes due to spillbacks. Besides bus frequency, the effect of the dwell times in the shape of the MFD is analyzed as longer dwell times reduce the average speed and flow. Fig. 4.5(b) illustrates the impact of different dwell time of buses on the MFD, plotting the MFD for buses resulting for multiple scenarios, where half scenarios have average dwell time of 25 s and the other 45 s. It can be seen that the higher dwell times lead to smaller flow capacity (in buses/hr), by about 15%. But, if we consider that bus occupancy might be almost double with higher dwell times, then the efficiency of the specific dedicated bus lanes in terms of passengers significantly increases. Optimal redistribution of road space between buses and cars has been introduced in in the previous chapters.

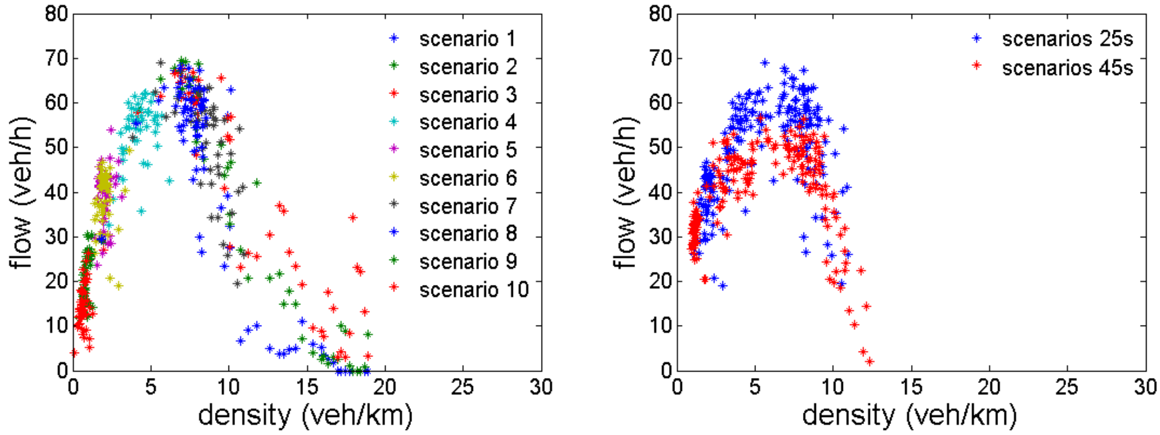

Fig. 4.5 (a) The MFD for buses for ten runs with different bi-modal traffic patterns and (b) the impact of different average dwell times of buses on the MFD.

4.2 An Analytical Passenger Flows Three-dimensional MFD

To maximize the passenger flow of a network and thus realize the efficiency of a multimodal network, we shall need a 3D-pMFD that can capture the dynamics of passenger flows but also can be derived from model (4-3) rather than curve fitting. To this end, we describe a derivation of the analytical form of the passenger flow 3D-pMFD based on the vehicle flow 3D-vMFD in this section. We further show that by partitioning the network into regions with homogeneous traffic and mode composition, the analytical model is able to reproduce the passenger flow pattern as observed from the data.

Let us investigate the analytical derivation of P given the analytical form of Q in (4-3). Our first objective is to estimate flows Q_c and Q_b for each mode by utilizing Q, n_c and n_b. To this end, we first consider that $V = (v_c n_c + v_b n_b)/(n_c + n_b)$ where v_c is the speed of car and v_b the speed of bus, and:

$$Q = Q_c + Q_b = (v_c n_c + v_b n_b)/L \tag{4-7}$$

To be able to estimate the individual flows of each mode, a relationship between their speeds is necessary, $v_b = v_c(n_c, n_b)$. To obtain v_b as function of v_c, we utilize a first order approximation, a linear model between v_c and v_b

$$v_b \cong \theta v_c + \beta, \tag{4-8}$$

where θ and β are parameters that can be estimated with real data and depend on the operational characteristics of the buses, the network topology and the ratio of cars and buses in the network. Alternatively, one can use a more detailed speed model, as for example the one proposed in Section 2.1 of Chapter 2, where the average distance between successive bus stops and the average dwell time at a bus stop influence the relation between v_b and v_c. We will show later that model (4-8) can give a decent approximation of v_b. Nevertheless, parameters θ and β might not be universal and depend on the ratio of buses and cars in the network, n_c/n_b. The aforementioned models might be less accurate in case of bus bunching, where multiple buses are queued in a bus stop with limited berths for boarding and alighting. A more detailed queuing model can be derived in such a case. By introducing (4-8) in (4-7), we can obtain the analytical form of v_c. Then after some manipulations we estimate Q_c and Q_b as follows ($Q(n_c, n_b)$ is noted as Q):

$$Q_c = v_c n_c = \frac{QL - \beta n_b}{(n_c + \theta n_b)} \frac{n_c}{L} \tag{4-9}$$

$$Q_b = v_b n_b = \left(\frac{QL - \beta n_b}{(n_c + \theta n_b)} \theta + \beta\right) \frac{n_b}{L} \tag{4-10}$$

A physical interpretation of such equations is that as bus is a slower mode, the fraction of bus flow is smaller than the fraction of bus accumulations. Then we derive $P(n_c, n_b)$ by combining (4-2), (4-9) and (4-10).

$$P(n_c, n_b) = h_c \frac{Q(n_c, n_b)L - \beta n_b}{(n_c + \theta n_b)} \frac{n_c}{L} + h_b \left(\frac{Q(n_c, n_b)L - \beta n_b}{(n_c + \theta n_b)} \theta + \beta\right) \frac{n_b}{L} \tag{4-11}$$

Note that the 3D-pMFD is a function of the 3D-vMFD, the individual accumulations of buses and cars, (n_c, n_b) and the average link length, L. A validation of the analytical model is provided later.

4.3 Network Partitioning for Mixed Bi-modal Network

4.3.1 Homogeneity in mixed bi-modal networks

By using a single value of θ and β in Equation (4-8) for the whole network, one assumes that the interaction between cars and buses is homogeneous in space and time. Nevertheless, the analytical estimation in (4-11) cannot accurately reproduce $P(n_c, n_b)$ if $Q(n_c, n_b)$ is assumed to follow the exponential model of (4-3). The reason is simple, bus lines are not evenly distributed in the studied network. The ratio δ_i between the density of bus k_{bi} and density of cars k_{ci} (where i is a link) is much higher in the center of the network, where there is higher frequency and density of bus lines. Fig. 4.6(a) displays the bus routes of the simulated network, which match the real ones. Note that some lines share links with the others in the center part of the network and that not all the lines can be visualized in the figure. We can observe that, the center area of the network ($x\epsilon$ [2500, 4000], $y\epsilon$ [3500 4500]) includes more bus lines than the other parts of the network. Let us take δ_i at a certain time interval during the simulation, when the network is operating at its maximum flow rate. Fig. 4.6(b) illustrates the scale distribution of δ_i by a grey-scale plot, where a darker color indicates a higher value of δ_i. This result shows that there exists a strong heterogeneity in the spatial distribution of δ_i, especially when comparing the mode composition between the center area (roughly identified) and the rest of the network.

To this end, we partition this network for the sake of homogeneity of δ_i within each partitioned regions. This partitioning builds in a clustering algorithm proposed in Ji and Geroliminis (2012) for heterogeneously congested networks. The objectives of partitioning are to obtain (i) small variance of link ratio δ_i within a cluster and (ii) spatial compactness of each cluster, which makes feasible the application of perimeter control strategies. The partitioning mechanism consists of three consecutive algorithms. Firstly, an over segmenting of the network is provided by a sophisticated algorithm (Shi and Malik, 2000). Secondly, a merging algorithm is developed based on initial segmenting and a rough partitioning of the network is obtained. Finally, a boundary adjustment algorithm is designed to further improve the quality of partitioning by decreasing the variance of δ_i while keeping the spatial compactness of the clusters. The similarity between links i and j, $w(i,j)$, is given by

$$w(i,j) = \begin{cases} exp\left(-\left(\delta_i - \delta_j\right)^2\right), if\ r(i,j) = 1 \\ \qquad\quad 0, \qquad o.w. \end{cases} \tag{4-12}$$

where $r(i,j)$ is the spatial distance between two links denoted by the length of the shortest part between i and j. Equation (4-12) simply explains that if two links are not adjacent, their similarity is negligible. Such an approach, can guarantee that the developed clusters will be well-connected (compact) and the topological characteristics influence links in the same cluster. For a detailed description, the reader should refer to Ji and Geroliminis (2012).

The studied network is divided into two regions (clusters) as shown in Fig. 4.6(c) after the application of the algorithm. Let us denote these two regions the “center region” and the “outside region” respectively, where the center region has more bus lines. We can see that this result closely matches what we have observed from the data in Fig. 4.6(b). Note that, the partition algorithm by its nature tries to give a compact shape with smooth boundaries of the partitioned regions. This explains why the exact shape of the area with high value of δ_i is not fully obtained. Nevertheless, the objective of partitioning is satisfied as the mean value of δ_i in the center region is much larger than the one in the outside region, $(\delta_{center}, \delta_{outside}) = (0.40, 0.04)$. The mean value of δ_i in the whole network is 0.20 (around 20% of vehicles are buses). The standard deviation of δ_i is 0.45 for the center region and 0.12 for the outside region, which is smaller compared to 0.76 for the whole network. This means that the algorithm succeeds to cluster together links with more similar δ_i. Note that standard deviations are high because in both sub-regions there are plenty of links with no buses, $\delta_i = 0$ and in the center region there are plenty of cases with bus bunching (and high δ_i). We are interested in investigating if this partition reduces the heterogeneity of congestion as well. Therefore, we calculate the total density k_i ($k_i = k_{ci} + k_{bi}$) of link i in each partitioned region and comparing the variance of k_i within each region. The mean values of k_i are 15 for the whole network, while 18 for the center region and 13 for the outside region. The standard deviation of k_i are 9 for the center region, and 6 for the outside region, which is again smaller compared to a value of 16 for the whole network. It shows that a mode-composition-based partitioning is sufficient to reduce heterogeneity of 3D-MFD. In reality this may not always be the case, if the scale of flow densities is not similar among links. In those cases, one may apply a partition algorithm based on both δ_i and k_i. As we will show in the next section, the resulting partitioning is sufficient for understanding the traffic heterogeneity and provides a good approximation of the 3D-pMFD.

Furthermore, we verify the usefulness of the partitioning from a traffic point of view, by examining the time when each region reaches congestion. Here, the “congestion” is referred as the start of network flow drop after the network reaches its maximum flow. To compare regions of different sizes, instead of accumulations we utilize densities. Fig. 4.7(a) shows the MFD (vehicle flow $Q = Q_b + Q_c$ vs. vehicle density $k = \sum_i k_i$) of the center region, the outside region and the whole network respectively, for one scenario. Density of buses and cars (k_b, k_c) are estimated by dividing accumulations with the total lane kilometers utilized by each mode, $L_b^{center} = 24$ ln-km, $L_c^{center} = 37$ ln-km, $L_b^{outside} = 26$ ln-km, $L_c^{outside} = 64$ ln-km, $\{b, c\}$ for bus and car respectively. The start of congestion in each region is indicated by an arrow with the corresponding timestamp attached. We can observe that the center region reaches congestion first, half hour earlier than the outside region. Fig. 4.7(b) displays the time series of the average density of the two regions respectively. It is clear that traffic condition in the center region is always heavier than the outside region. This information is important, as it indicates that different traffic management measures should be applied for the two regions based on their individual traffic states.

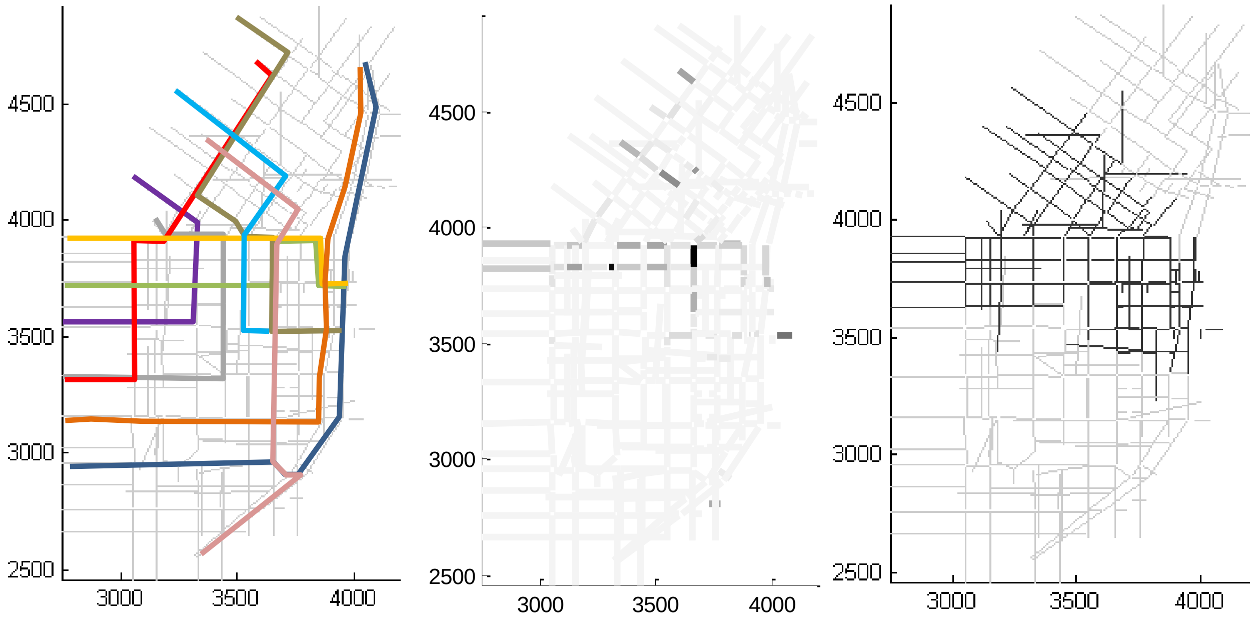

Fig. 4.6 Illustration of the partitioned network: (a-left) Layout of the bus lines over the network, (b-middle) contour plot of link ratio δ_i, and (c-right) result of partitioning.

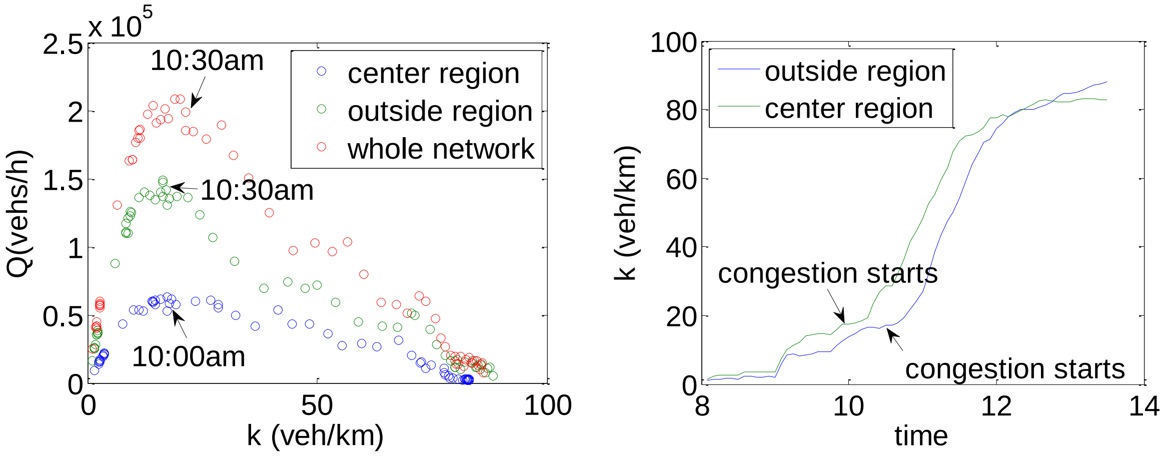

Fig. 4.7 Comparing the time each region reaches congestion in (a) network flow-density relation and (b) density time-series.

4.3.2 Vehicle 3D-*v*MFD after partitioning

Let us now show the 3D-vMFD for the partition regions. We utilize the data of the same scenarios as used for Fig. 4.1. To have the scale of the MFDs of the two regions comparable, we normalize the values of accumulations and reproduce densities. Denote density of cars k_c and density of buses k_b. Densities of buses, have been estimated by normalizing accumulations with the length of roads in the sub-networks that buses are moving on. The normalized 3D-*v*MFDs for the two partitioned regions are plotted in Fig. 8(a). k_b of the center region has a much higher value, as the region is covered by more bus lines. Nevertheless, the general pattern observed in Fig. 8(b) and (c) is similar to the 3D-*v*MFD as observed in Fig. 3, especially the values of critical densities (for example 35vh/km * 100 ln-km =3500 veh which is the critical accumulation of Fig. 4.3(b)). Given the 3D-*v*MFD of each region, specific management measures such as perimeter signal control can be applied with the objective of operating at the region's optimal regime.

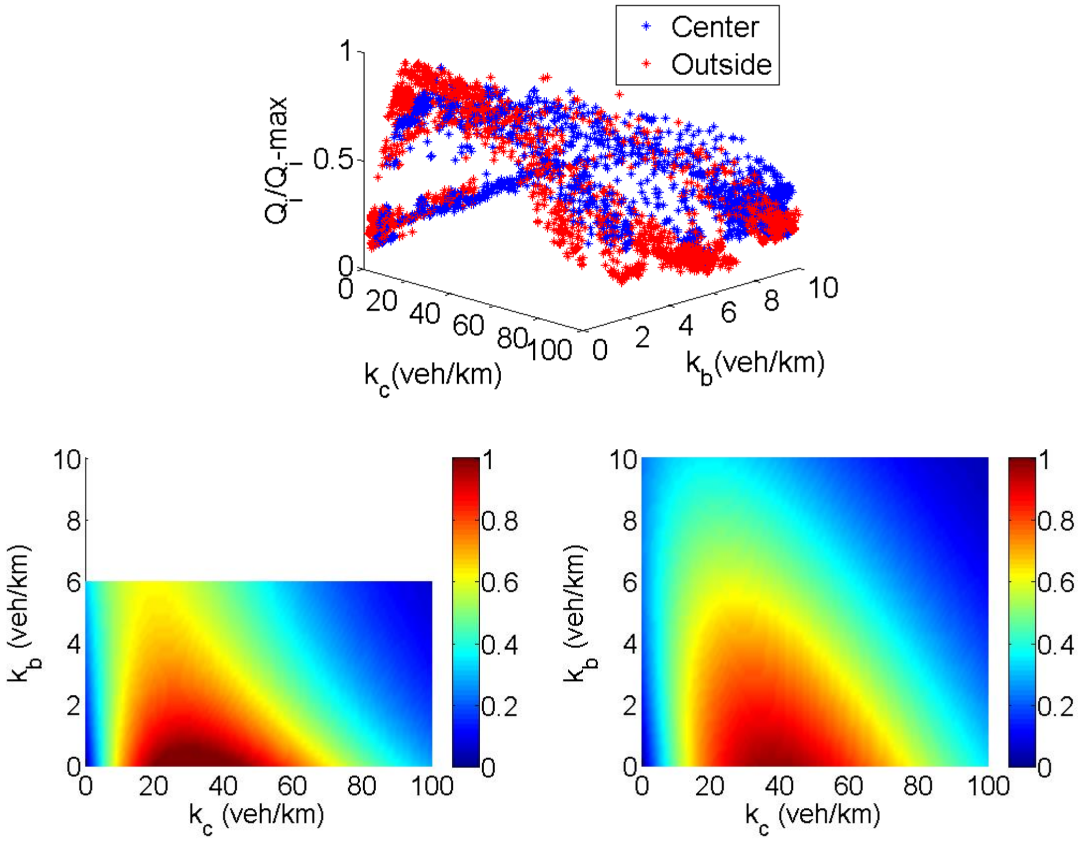

Fig. 4.8 (a-top) The normalized 3D-*v*MFD of the partitioned regions; the contour plots of the 3D-*v*MFD on (k_c, k_b) plane for (b-down left) the center and (c-down right) the outside region.

4.3.3 Passenger 3D-*p*MFD after partitioning

The 2-D accumulation-flow relations (5min) of one single scenario are displayed in Fig. 4.9(a), for both vehicles and the passengers. Even without partitioning, the network has a low scatter MFD with a well-defined maximum flow and a clear value of critical accumulation. As the accumulation ratio δ is not significantly varying during the simulation, passenger flow $P(n)$ vs. total accumulation $n = n_b + n_c$ follows an MFD-shape as well. Both MFDs have similar shape, but the value of critical accumulation of $Q(n)$ is larger.

Now let us compare the resulting passenger dynamics before and after the partitioning. Fig. 4.9(b) plots n_b vs. n_c with time for the two partitioned regions. It can be seen that the value of δ for the center region is about 9%, which is 6 times larger than δ of the outside region. Note that δ is the aggregated ratio between $\sum_i n_{bi} / \sum_i n_{ci}$. Fig. 4.9(c) depicts the space-mean speed relation for each region between cars and buses. The dashed lines are the fitting results of model (4-8) on the speed relation scatters. Parameter θ for the center region has a value of 0.3 and for the outside region is 0.5, while β is 0 in both cases. If Equations (4-8) to (4-11) were applied to estimate $P(n)$, errors would not be negligible because of the differences in δ and θ between the two regions. An example is given in Fig. 4.9(d), which displays the estimated $P(n_c, n_b)$ for the whole, the center and the outside regions respectively. Evidently, the dynamics are distinct for the three regions.

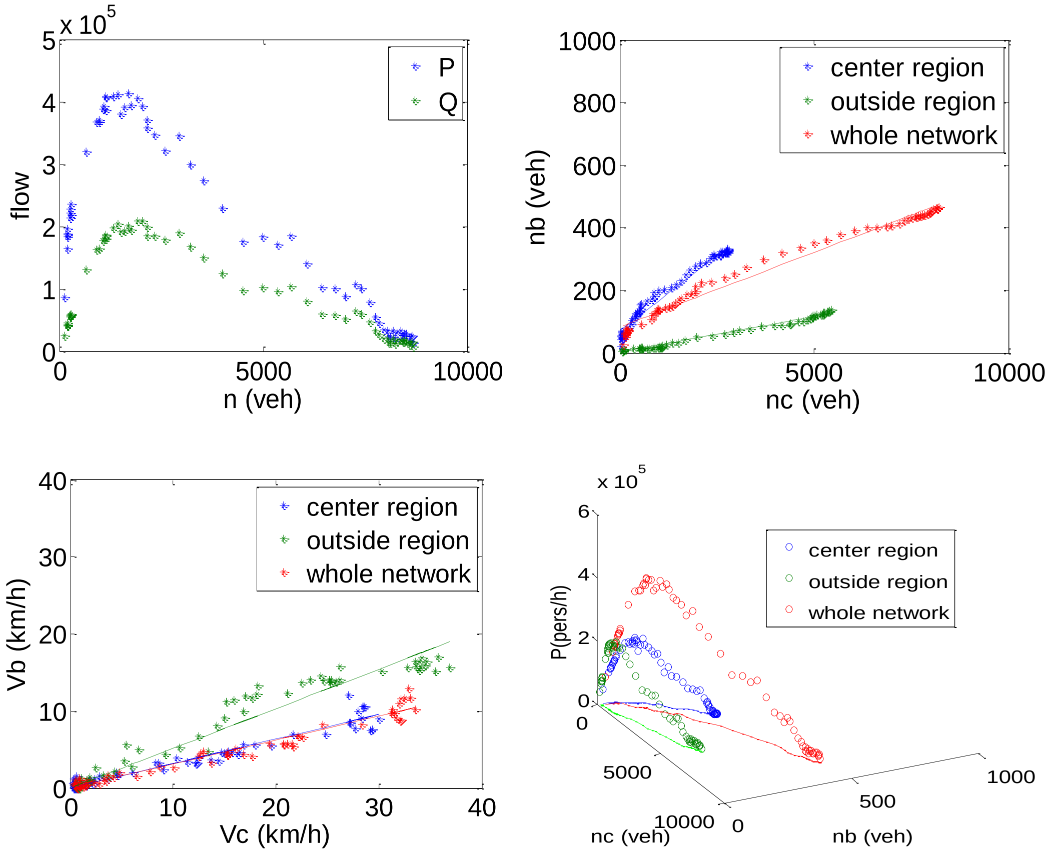

Fig. 4.9 (a-top left) Passenger MFD from the simulation, (b-top right) n_c vs. n_b, (c-down left) v_c vs. v_b, and (d-down right) the estimated passenger MFD for the two partitioned regions and the whole network.

Let's now utilize model (4-8)-(4-11) and estimate the 3D-*p*MFD for three different scenarios with various bus demands. Denote the three scenarios as s1, s2 and s3 respectively. After applying the partitioning algorithm, δ and θ are estimated for the center and outside region, and the passenger MFDs are estimated. The total network passenger flow is then estimated as

$$P(n_c(t), n_b(t)) = P_1(n_{c1}(t), n_{b1}(t)) + P_2(n_{c2}(t), n_{b2}(t)) \qquad (4\text{-}13)$$

where indices 1 and 2 represent the two partitioned regions, center and outside.

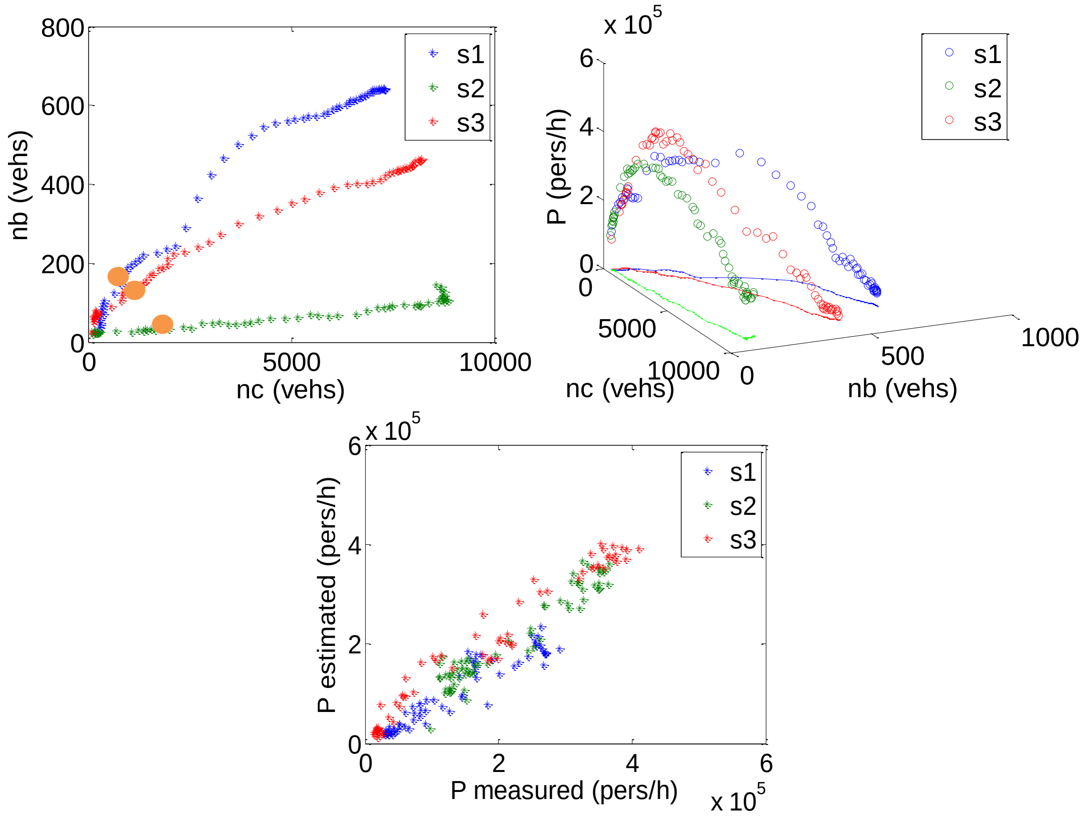

Fig. 4.10 Comparison of (a-top left) n_c vs. n_b for the entire network, (b-top right) the estimated 3D-*p*MFD and (c-down) the measured P vs. the estimated P by model (4-13) for the three scenarios.

Fig. 4.10(a) shows the accumulation relationship (5min) for the entire network for the three scenarios. The red dotted line (middle) represents the same scenario as of Fig. 4.9. The yellowing circles highlight the accumulation pairs where maximum flow is achieved in each scenario. Fig. 4.10(b) depicts the corresponding $P(n_c, n_b)$ of the three different scenarios, and the δ of each scenario can be approximately estimated given the projection of the 3D curves on the (n_c, n_b) plane. By applying partitioning and identifying the region-specific θ, the analytical model is able to reproduce the important patterns as observed in Fig. 4.2(b): (i) P increases and then decreases as n_b increases, (ii) the critical accumulation largely shifts towards the left, and (iii) there exists an “optimal operational regime” where the same value of maximum P (close to critical value) can be achieved with different combinations of n_b and n_c. Fig. 4.10(c) compares the estimated P from the model with the measured P from the simulation for the whole network (the closer the linear rate to 1, the closer the estimation to the original values). With the partitioning, the estimation result of P shows a decent accuracy. Furthermore by comparing the estimated passenger flow with and without partitioning, we find that the mean absolute error has a 10% improvement (the higher the value of δ, the higher the improvement).

4.4 Application of the 3D-MFD and Partition: Real-time Perimeter Flow Control

It is worth mentioning here that the 3D-MFD and the partitioned 3D-MFD can be readily utilized to develop perimeter flow control strategy. In this section, we will briefly introduce the control logic and the improved system performance under two perimeter flow control schemes: namely a single-region control and a two-region control. The performance of these two control strategies will be compared to the embedded control plan of the studied network, which was implemented in the field network (will be referred as the "pre-timed control" in later text).

The dynamics for a perimeter controlled network can be illustrated as follows. Denote the time-dependent inflow of the network $\beta(t) \in [\beta_{min}\ \beta_{max}]$, controlled at the perimeter of the network; $d(t)$ as the uncontrolled demand and disturbances. Buses enter the network with slow-varying service frequencies. The outflow of the network $O(n(t), \delta)$ is described by the 3D-MFD. Mode composition δ depends on the operational characteristics of buses. Assume we can monitor the mixed traffic accumulation $n(t)$, may be estimated via detectors or GPS in real-time. Then the state equation can be written by

$$\frac{dn(t)}{dt} = \beta(t) + d(t) - O(n(t), \delta) \tag{4-14}$$

Equation (4-14) is highly non-linear, given the non-linear property of the 3D-MFD. Provided the function form of the 3D-MFD, linearization can be performed at the desired point $(\hat{n}, \hat{\beta}, \hat{d})$ with first-order Taylor approximation. Recall that the desired state is the critical state given δ. Briefly speaking, our perimeter flow control has the form in Equation (4-15), aiming to mitigate congestion in the bi-modal network by maintaining the above system to operate around the desired steady-state $(\hat{n}, \hat{\beta})$ for given $\delta \in \Omega$, while the system's throughput (as expressed by the 3D-MFD) is maximized.

$$\beta(t) = \hat{\beta} - K[n(t) - \hat{n}] \tag{4-15}$$

where $\beta(t)$ is the flow of vehicles allowed to enter the network if the current state $n(t)$ is observed. This controller is a classical Linear Quadratic Regulator (LQR). In control theory, such controllers can be analytically obtained given the dynamics of the target system

(Equation (4-14)), and they are proved with high stability in maintaining the target system at desired states (Papageorgiou and Kotsialos, 2000). For detailed information on the design of the controllers (the linearization of the system dynamic equation, the derivation of Equation (4-15), the determination of control gain K, the combination of Equation (4-15) for the two partitioned regions in the two-region control case), an interested reader should refer to Ampountolas et al. (2014a), (2014b), in which the author of this thesis serves as the second author that mainly contributes in the application of the developed controller and performing the numerical analysis.

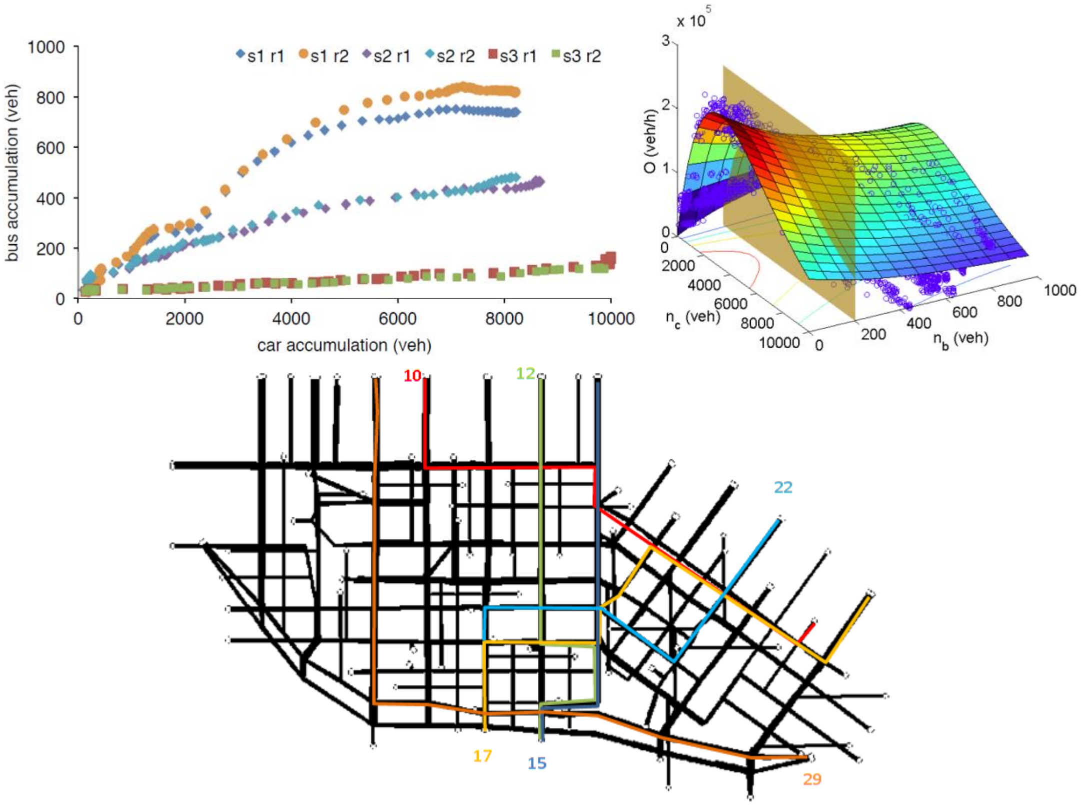

Fig. 4.11 (a-top left) The composition rate δ for three representative scenarios, Ω =[0.02 0.15] (b-top right) A cross-section of the 3D-MFD for a constant bus accumulation demonstrates the typical dependence of flow with δ, and (c-down) Network characteristics: 2 square-kms, 100 intersections, 430 links of total 100 lane-kms, 8 major bus lines (in total 29 lines with service frequencies between 3-20mins).

4.4.1 Single-region control

Fig. 4.11 shows the relevant properties of the studied network. Let us first present case study results on the single-region controller in Fig. 4.12 and Fig. 4.13, by comparing scenarios without applying the controller (denoted as pre-timed control, fixed pre-defined signal control plan), and with applying the perimeter controller (denote as perimeter control). Fig. 4.12(a) and (b) depict the resulting MFD of five scenarios under pre-timed control and perimeter control cases. When perimeter control is applied, the network operates under efficient traffic conditions and states in the decreasing part of the MFD are not observed; under pre-timed control, the network becomes severely congested with states in the congested regime of the MFD. Moreover, the outflow is maintained to high values around the set point. We can also observe that the hysteresis formed in the offset period of congestion is reduced significantly, especially for the traffic congested scenarios S2 and S4.

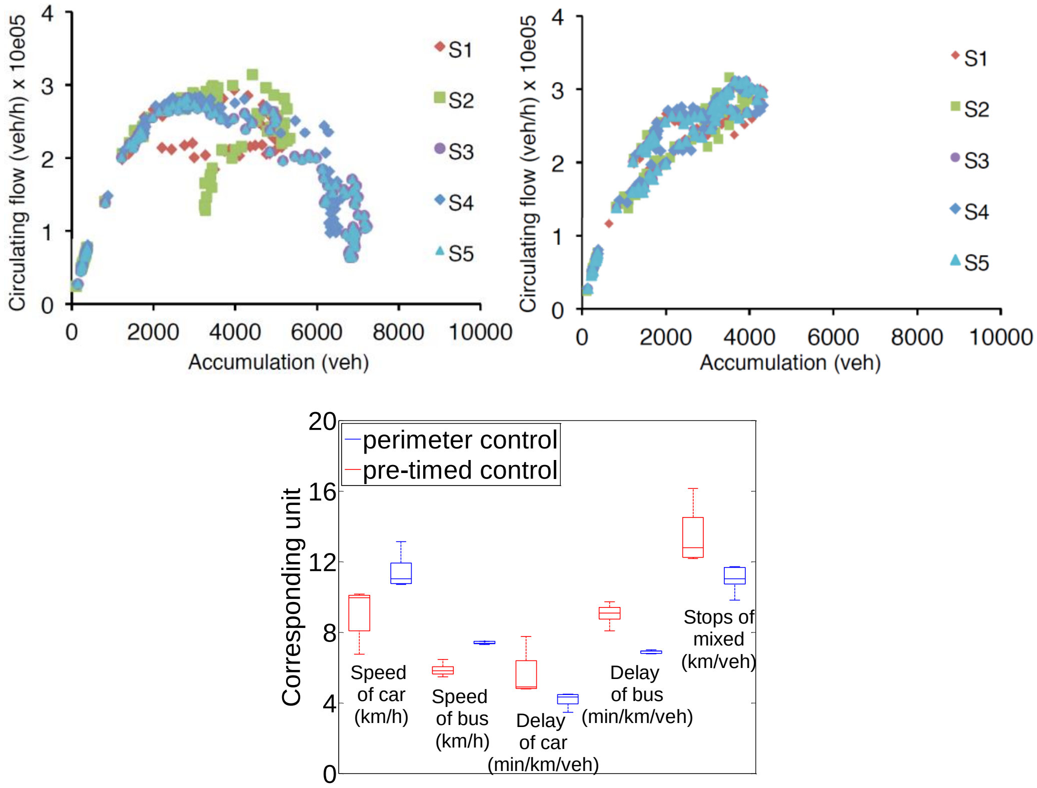

Fig. 4.12 Comparing overall traffic performances for 5 different scenarios: (a-top left) 2D-MFD with pre-timed control and (b-top right) 2D-MFD with the proposed perimeter control (the overall congestion is eliminated); (c-down) Boxplot of performance indicators: Travel speed and delay improved 40% for cars and 30% for the buses.

Fig. 4.12(c) depicts the resulting average performance (speed, delay, stops) of each mode of 6 replications. Clearly the proposed perimeter control increases the speed and decreases delays and number of stops in both modes of traffic (in average from 40% to 50%). A further analysis of the spatial dimension of traffic congestion in the central avenue (Market Avenue) of the network and its upstream links (southeast) can shed more light in the perimeter control actions within the transport public lines. The considered path includes the entire route for public lines 15 and 19 and six other bus lines that overlap parts of the path such public lines 5, 11, and 13, to investigate the interaction among conflicting public transport lines.

To gather the bus trajectories that traverse this path, we simulate buses equipped with GPS-based mobile sensors that reporting their location every 3 seconds. Fig. 4.13 displays the gathered bus trajectories for eight public transport lines (each with different colour) during the heart of the rush (11:00 am to 13:00 am), when pre-timed control and perimeter control are applied. In these time-space diagrams, the x-axis reflects the simulation time, while the y-axis reflects the one-dimensional distance travelled. Given that the studied network is a grid, the two-dimensional road distance is transformed into one-dimensional by calculating the Manhattan distance between the GPS-reported location of a bus and the starting point of the path. The horizontal time distance between consecutive bus trajectories with the same colour indicates the headway between two buses servicing the same public transport line. The location of junctions and bus stops are also reported (see caption for details) to allow a better understanding of bus bunching and stop-and-go phenomena within the public transport lines. Fig. 4.13 underlines the superiority of perimeter flow control over pre-timed control to maintain public transport lines normal time schedule. A time-space diagram of bus trajectories for one of the high demand corridors (across Market Avenue) is shown for pre-timed and dynamic control. Traffic conditions are almost identical for both control cases from 11:00 am to 11:20 am, as time goes on, in the pre-timed control case, buses entering their transport lines (upstream traffic) suffer increasing delays waiting other buses and cars in the centre of the network between 700 m and 1200 m (downstream traffic) to be served. Then traffic condition becomes deteriorated in the center of the network, link queues start spilling back and blocking upstream junctions, thus the entering traffic approximately matches the speed of the downstream traffic. This creates multiple backward moving shockwaves with negative speed that are illustrated with arrows in Fig. 4.13. Clearly when perimeter control is applied the network operates under free-flow traffic conditions and buses are able follow their normal time schedule (with slight travel delays). It can be seen that buses only experience delays between 11h50 and 12h30 at the same spatial distance. To further investigate what caused these delays, the traffic conditions in bus line 11 (among others) were carefully analysed. The inspection of different replications eventually shown that the delays are mainly caused by a sudden increase of left turn demand of cars and buses at a specific junction close to the protected network. Note that the existence of such cases can be possible under perimeter control, since we only control junctions at the perimeter of the network.

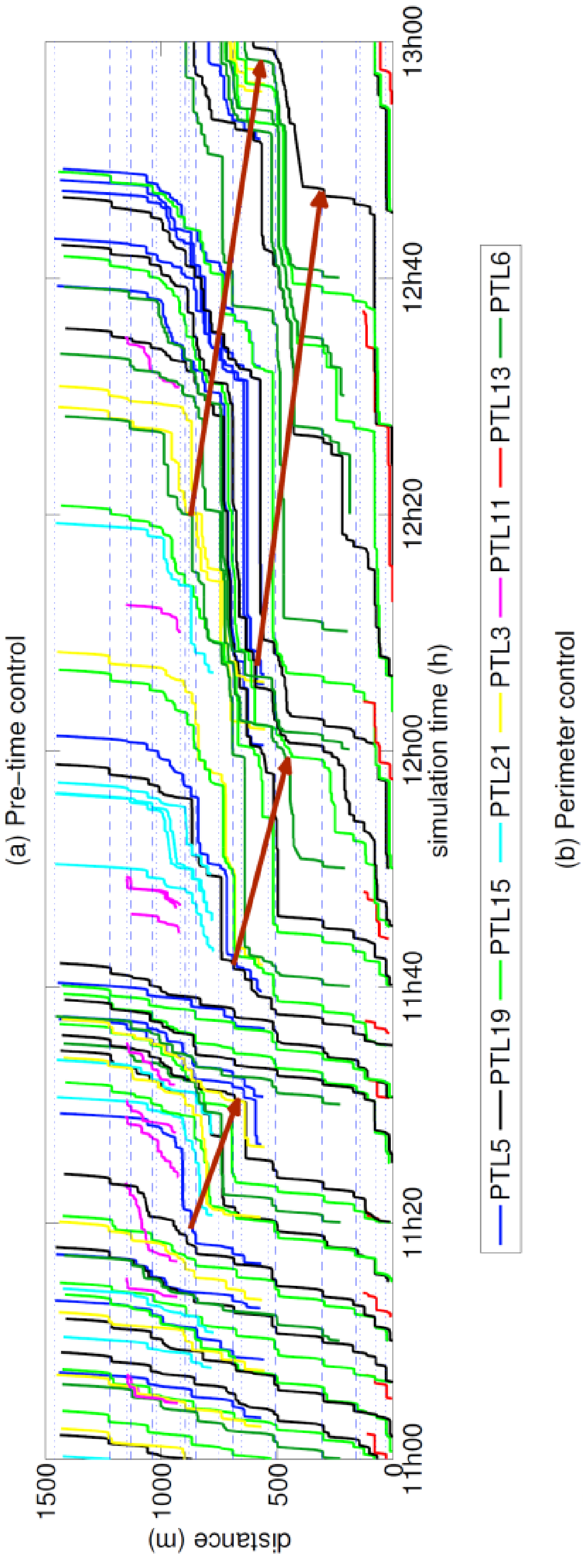

Fig. 4.13 Comparing bus trajectories in several bus lines in the network during the heart of rush between (a) pre-time control and (b) perimeter control. Dotted lines indicate the location of intersections. Dashed lines indicate the location of bus stops. Arrows indicate shockwaves.

4.4.2 Two-region control

The results presented in Fig. 4.15 and Fig. 4.16 are based on the two-region clustering in Fig. 4.14(a), which is used to exploit and illustrate the benefits of two-region over single-region (whole network) perimeter control. The single-region control remains the same. We apply both controllers in scenarios where demands are higher than the ones used in the previous section. Fig. 4.15 depicts the average performance (speed, delay, stops) of each mode of traffic for 10 replications of a demand scenario with strong heterogeneity in the spatial distribution of mode composition and congestion. It can be seen that two-region robust perimeter control increases the speed and decreases delays and number of stops in both modes of traffic in average by 10%. It should be noted that for scenarios with small variability of mode composition the performance of the two-region over single-region control is slightly deteriorated, as expected. Thus, multi-region perimeter control must be carefully designed and applied to a heterogeneous network.

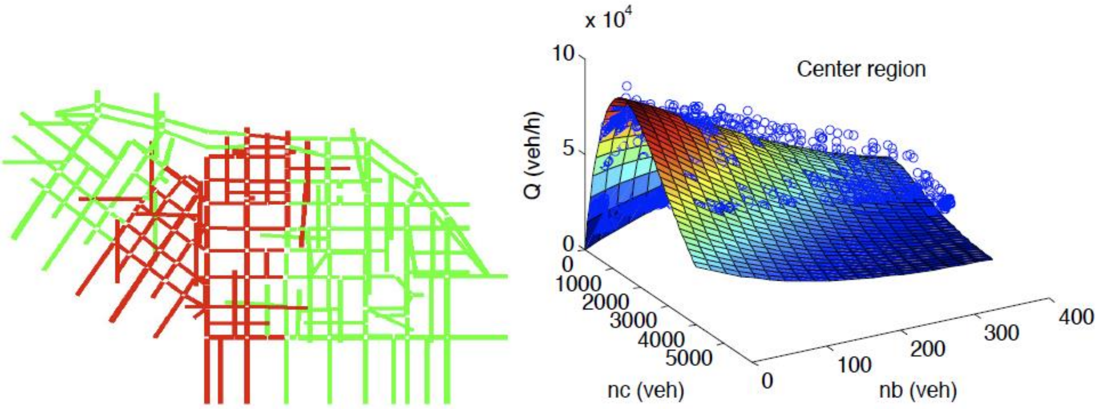

Fig. 4.14 (a) Snapshot of Downtown San Francisco clustering into two regions; red color on the city center; green color the rest of the network, and (b) the 3D-MFD of the center region relating accumulation of cars n_c and buses n_b with circulating flow, respectively.

A further analysis of the spatial dimension of traffic congestion within the two regions can shed more light of the effect of perimeter control to the Public Transport Lines (PTL). The considered spatial distance includes eight PTL that overlap parts of the two regions, to investigate the interaction among conflicting public transport lines. To not repeat ourselves, we carry out the analysis on the efficiency and equity properties of both modes of traffic under single-region and two-region perimeter control. Travel Time per Kilometre-distance Travelled (TTKT) is chosen as performance indicator. Considering TTKT allows us to compare travel times of the same scale in case they those travel times vary significantly due to different trip lengths. Different trip lengths are natural because the size of the two regions is uneven and the two modes of traffic generate different trips. Fig. 4.16 displays the cumulative probability distribution of the TTKT for each region and mode (cars and buses). TTKTs are collected from 5 scenarios, every 5-min for a 2.5-hour peak period. The two curves (with blue and red colour) in each of the four subplots depict commutative distributions of TTKT under single-region and two-region control, respectively. The median and the standard deviation of each distribution is calculated and displayed. It can be seen that the two-region control performs better than single-region. The median value of the TTKT for cars improves by about 10% for both regions. For buses the improvement in terms of median is quite small (about 3%), but reliability increases (standard deviation is 20% smaller). Remarkably, the two-region controller increases the reliability of the network as its cumulative curve of TTKT is less spreading and the variance is smaller. Finally, it is stimulating to observe that the two-region control provides fairly equal improvements to both regions, albeit the center region seems to benefit slightly more than the outside region.

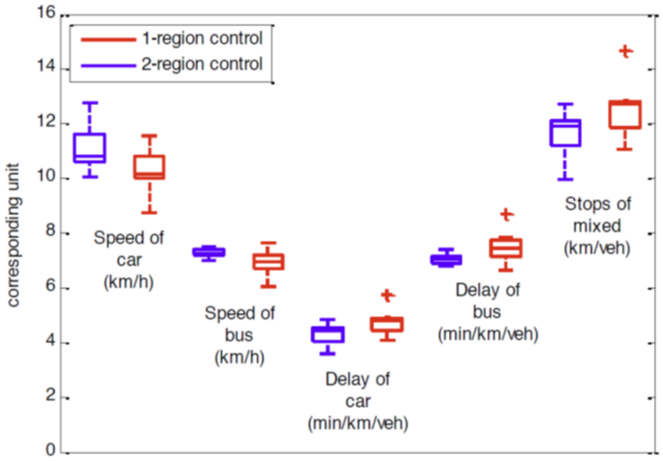

Fig. 4.15 Minimum, maximum, and median values of different performance indices for the two modes of traffic under single-region (1-region) and two-region controls (2-region).

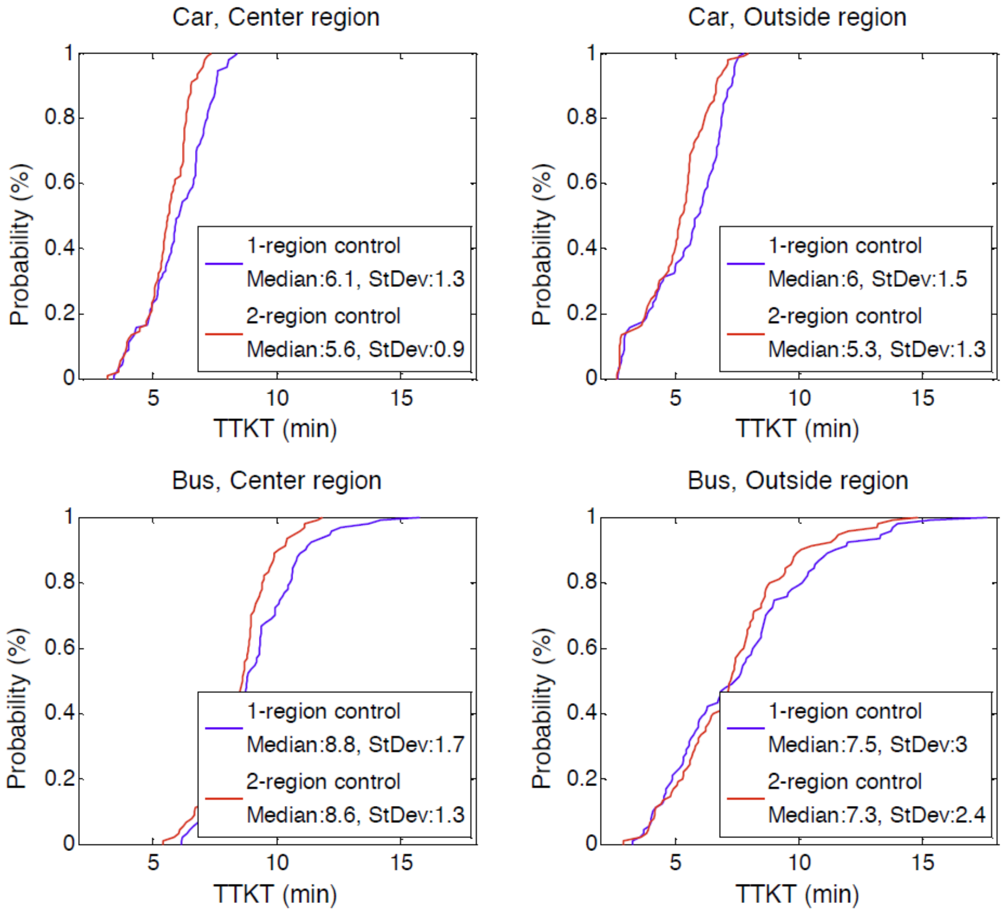

Fig. 4.16 Cumulative distributions of TTKT for the two modes of traffic in the two regions under single-region and two-region controls.

4.5 Summary

To identify and quantify individual modal impact on the global traffic performance, we investigated a new type of MFD model in Chapter 4. The existence of a three-dimensional multimodal MFD (3D-MFD), relating the accumulation of cars and buses, and the circulating vehicle flow in a network, was demonstrated via simulation experiments. An exponential-family function was proposed for the analytical form of the 3D-MFD, where the individual modal and the joint impact on global performance were directly observed. To further investigate the modal impact, the Bus-Car Unit equivalent value was estimated and found state- and mode-composition-dependent rather than deterministic. Then, we derived the passenger-flow 3D-MFD with an elegant analytical model, which provided a different perspective of bi-modal flow characteristics. We applied a partitioning algorithm to cluster a center and an outside region of the network according to the similarity of the car/bus density ratio. It was found that such partition enhanced the accuracy in the estimation of the 3D-

MFDs, which revealed the impact of heterogeneity of mode composition on the bi-modal modeling. Furthermore, we utilized the 3D-MFD to develop perimeter flow control strategies. We presented the results of two controllers, a single-region controller and a two-region (center-outside) controller. Congestion was significantly reduced for the whole network, while the performance of buses in terms of travel delays and schedule reliability was improved without even giving bus priority. Queues and gridlock were avoided on critical paths of the network.

The multimodal urban modeling should be also investigated by considering additional heterogeneity (i) among users, with respect to their mode choices and the trip length and (ii) among different regions of a city. We will show how the findings of this part in the following chapter.

5 Development of Dynamic Congestion Pricing Schemes with Macroscopic Fundamental Diagram in an Agent-based Simulation

The previous chapters have demonstrated the efficiency of the MFD-based strategies for improving mobility. In this chapter, we aim to investigate how such strategies, in particular the congestion pricing schemes, can be designed with the MFD if complex behavioral dynamics are present (comparing to Chapter 2 and 3 where mode choices are treated, and Chapter 4 with route choices). We will test different pricing schemes in an agent-based model where travel behaviors are reproduced comprehensively for resembling the reality. Comparing to traditional approaches, the advantages our pricing approach lie in that it has lower collection and transaction costs than the link-based or area-based tolling, and it is based on traffic models utilizing input data that are readily observable with existing monitoring technologies.

To develop efficient pricing schemes, behavioral adaptation and heterogeneity should not be overlooked. Questions such as how to adjust pricing rate based on user's behavioral changes or what are the impacts of an incentive program of using public transport (PT) on the performance of pricing should be addressed. It has to be highlighted that despite the vast literature in pricing, field tests and implementations are quite limited and this is among other reasons of user acceptability. Thus, providing incentives for taking other modes of transport (e.g. return a fraction of the tolls paid to users that switched to public transport mode) can make pricing policies more attractive for real cases.

This chapter is structured as the follows. Section 5.1 provides a general description of the properties and simulation algorithm of the utilized agent-based model. Before presenting in detail our MFD-based pricing control schemes in Sections 5.3 and 5.4, Section 5.2 examines the existence of the MFD in the simulation output and other traffic flow characteristics that are relevant to support the development of MFD-based control strategies. Case studies on the two proposed pricing schemes are given in Section 5.3 and 5.4 respectively. The performance of pricing with respect to effectiveness, efficiency and user adaptation is evaluated. Section 5.5 takes one step further and investigates the impact of pricing if heterogeneity among users is considered. A summary of the chapter can be found in Section 5.6.

5.1 The Multi-agent Based Traffic Simulator

The agent-based model MATSim has been widely applied for transport and land use studies and travel behavior modeling. The simulation model integrates activity-based demand generation with dynamic traffic assignment.

Demand generation is embedded in a concept of daily activity sequence from which the need for transport is derived. In the context of activity-based demand, the entire activity plan (mode choice, departure time choice and the activity sequence) is the unit of decision to iterate route assignments.

Random utility theory is applied to generate plans of daily activities. Each agent in the simulation is assigned with different utility functions when performing different activities. A typical utility function consists of four items for an agent performing her daily plan: $U_{i,act}$ being the score performing activity i, $U_{i,travel}$ being the score of traveling to activity i, $U_{i,wait}$ being a penalty for waiting instead of performing activity i, and $U_{i,short}$ being a penalty for performing activity i for a too short duration, where $i \in \{1 \dots n\}$ is the number of planned activities. Now let us consider the existence of heterogeneity in the agent population and agents are grouped with respect to VOT (agents within the same group have the same VOT). Denote j as the group index. Equation (5-1) shows the utility calculation for user group j:

$$U_{plan}^{j} = \sum_{i=1:n} U_{i,act}^{j} + U_{i,travel}^{j} + U_{i,wait}^{j} + U_{i,short}^{j} \tag{5-1}$$

Note that in the current work, the influence of VOT is mainly reflected by the physical travel cost term $U_{i,travel}^{j}$, while it is easy to incorporate treatment to other terms, such as earliness and lateness penalty which are embedded in $U_{i,act}^{j}$.

These properties provide more realistic reaction of the users (agents) towards any traffic management strategy, and thus help traffic engineers develop more reliable toll schemes.

As for the simulation, the process can be summarized as follows: Each agent performs its activity plan (daily events), which are simulated along the timeline in the model representation of the physical world. A First-In-First-Out queue traffic model is embedded to determine the travel time of the agents moving from one location to another, with predefined road capacity and consideration of shockwave between vehicles travelling backwards at constant speed in the case of traffic jam discharge. The executed activity plans are evaluated with a measure of general utility, as introduced above in Equation (5-1). Given the evaluation, certain amount of the agents (by default 10%) carries out a re-planning strategy such as changing mode choice or departure time choice. New plan thus is created and added into agents' memory. Agents then decide to either execute the new plan or choose one of the plans from their memories, preferably these with the highest scores. This procedure is iterated via a day-to-day learning process, until conceptually agent-based stochastic user equilibrium is achieved (Nagel and Flötteröd, 2009).

Under such equilibrium, not only travel behavior (e.g. mode choice and route choice between the same origin-destination) is stabilized, but also the utility of the entire daily activities U^{j}_{plan} is indirectly optimized. Please note that as it is complicated to achieve a classical SUE for systems with such multi-dimension behaviors, trade-off needs to be made between the computational cost of iterations and the magnitude of behavioral stabilization. For detailed information on MATSim, the readers may refer to Meister et al. (2010).

5.2 The Existence of the MFD in the Agent-based Simulation

The main motivation of this section is to t examine the traffic outputs of the agent-based simulator MATSim, before carrying out the analysis on pricing. We will show that although MATSim is designed and calibrated for activity-based simulation without modeling in details disaggregated characteristics of traffic (such as car-following and queue dynamics), the outputs can reproduce the physics of traffic reasonably at an aggregated level, as expressed by an MFD. Therefore, the effect of traffic management on performance can be quantitatively observable independently, without utilizing the high-level detailed information of MATSim. MFD-based strategies can be developed and tested in this simulation environment of complex behavioral adaptations, e.g. the change of departure time and mode choice. Furthermore, we also examine if spill-back phenomenon can be observed in MATSim since spill-back is the key reason of the existence of capacity decrease.

5.2.1 The MFD

For a 4.5km-radius center region of Zurich, two MFDs are shown in Fig. 5.1(a). They show the existence of Regime I and the beginning state of Regime II, indicating this network on a macroscopic level is not heavily congested. To observe a complete MFD with congested states, we zoom into the center region of the area where traffic congestion can be observed. The resultant MFD is shown in Fig. 5.1(b), with the existence of a complete Regime II in which network is operated at its capacity and congestion Regime III where the more agents queue in the network while less reach destination.

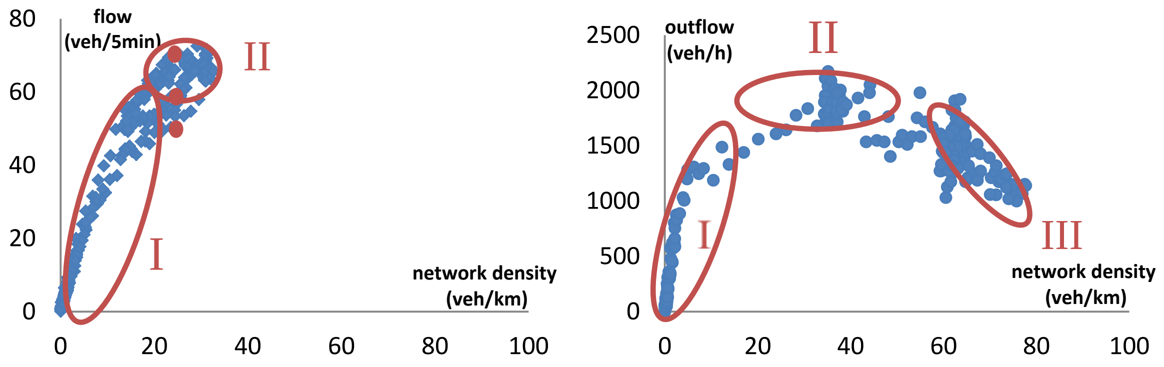

Fig. 5.1 (a) MFD of a 4.5km radius area network of Zurich and (b) of a 1km radius area

5.2.2 An explanation of scatter in the MFD

By examining Fig. 5.1(a), we find loops in the timely connected scatters. Similar phenomena have been observed by Geroliminis and Sun (2011), and defined as the hysteresis phenomena: higher network flows are observed for the same average network density in the onset and lower in the offset of congestion. This is because there are different spatial distributions of congestion, high variance among densities, for the same level of average network density for different times of a day.

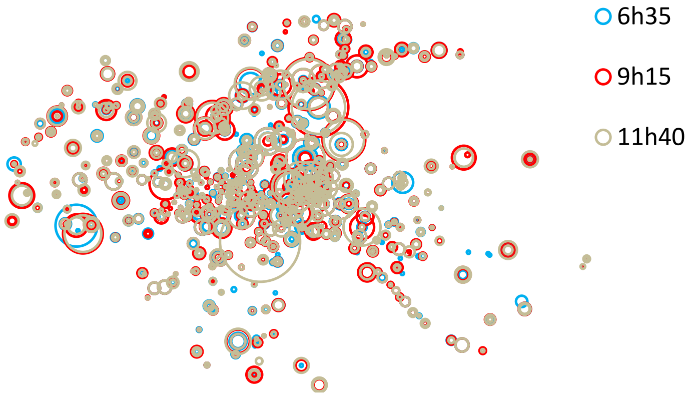

Fig. 5.2 Density distribution of the network at time 6h35, 9h15 and 11h40. Axis represent coordinates of links (x,y-axis represent coordinates)

In Fig. 5.2 we draw link density distributions of all links in the network at three different times, where the network holds the same amount of vehicles. The chosen times are 6h35 (the highest point), 9h15 (the lowest point) and 11h40 (medium point). They refer to the onset of the morning peak from 6h30 to 9h30, and the offset from 9h30 to 11h30. The size of the “bubble” is the value of density and X-Y axis are coordinates of links. It is clear that traffic is more uniformly distributed at 6h35 than at 11h40, while at 9h15 densities are extremely high at some locations. The standard deviations of density are 24, 29 and 32 veh/km respectively, while flow rates are 840veh/h, 720veh/h and 600veh/h.

5.2.3 The Fundamental Diagram (FD) and spill-back effect

We now provide a detailed investigation on how congestion propagates at link level in MATSim. Flow rate and density are calculated for individual link. We study if spillback effects are present and long queues can decrease the output of upstream links. Fig. 5.3 shows the FDs for four consecutive links along one of the arterial roads of Zurich in Scenario 1. The colour of the scatters in the figure corresponds to the link with the same colour. From the figure, we see at density around 10, the flow rate reaches its maximum and remains the same value until density around 30. Then flow decreases as density increases, indicating congestion happen at the latter point. Secondly, we see that congestion spills back from downstream to upstream, as the links in green and yellow operate in Regime I and II while the links upstream experience Regime III. These observations make clear that queues are growing from downstream to upstream and blocking effects are present in the simulator. We also observe that queues propagate from downstream to upstream, as congestion appears upstream at a later time. It is also clear that the individual fundamental diagrams exhibit high scatter, especially in the congested regime.

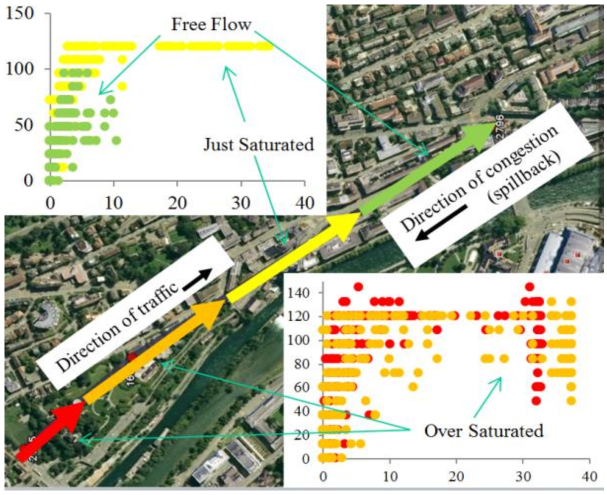

Fig. 5.3 Illustration of congestion propagation along consecutive links. Axis is density of the road and axis y is the flow rates. Note that the scenario utilized 25% of the population

5.2.4 The MFD-controlled pricing schemes

Let us now introduce the control algorithm of the pricing scheme. Denote n as the simulation and the toll adjustment index, $n = 1,2 \ldots \ldots n$, the toll at the (n)-th toll adjustment for time interval t, $Toll_t(n)$, and the network density $K_t(n)$. Suppose we have an initial scenario for a certain urban network, without any pricing. We observe the congestion level (historical data) and determine an initial pricing (offline). Then a new simulation is executed where travelers make changes to optimize their new travel cost. Once the travelers fully adapt their behaviors under the pricing and stabilization is reached, we evaluate and adjust the tolls. Note that the toll is kept fixed during each simulation n(if prices are adapted in every iteration, then the travelers do not have enough time to adapt their decisions and convergence might never be reached. To be able to simulate traveler adaptation in a realistic way while dynamic pricing is applied, pricing is updated based on the traffic conditions after a number of iterations where the system is close to convergence: the agent-based equilibrium as described in Section 5.1).

Given the traffic conditions (converged) under the n-th toll, we obtain the resultant MFD and identify $K_t(n)$ which exceeds the critical density K_{cr} (where maximum network capacity is reached and the network production drops down), and control the toll for time t to prevent the network from falling beyond K_{cr} A typical result is that traffic density will maintain a state around K_{cr}. Different from online traffic control strategies, where the effect of control on traffic is immediate and car users might not have the ability to adapt, the effect of pricing can be long-term and depends on how users change their travel behavior. Pricing could make users to change their departure time from the origin, or mode of transport. Considering this adaptation in the development and design of efficient pricing schemes is highly desired, while challenging. As the adaptation is quite difficult to model in an analytical way, we consider that an agent-based framework influences their decisions in an unknown way to system operators (traffic manager). Nevertheless, the operators can monitor the aggregated behavior as network densities at different times $K_t(n)$ are known. This MFD-controlled pricing adjustment process will be terminated until all $K_t(n)$ are found below K_{cr}.

The algorithm is illustrated in the flow chart below.

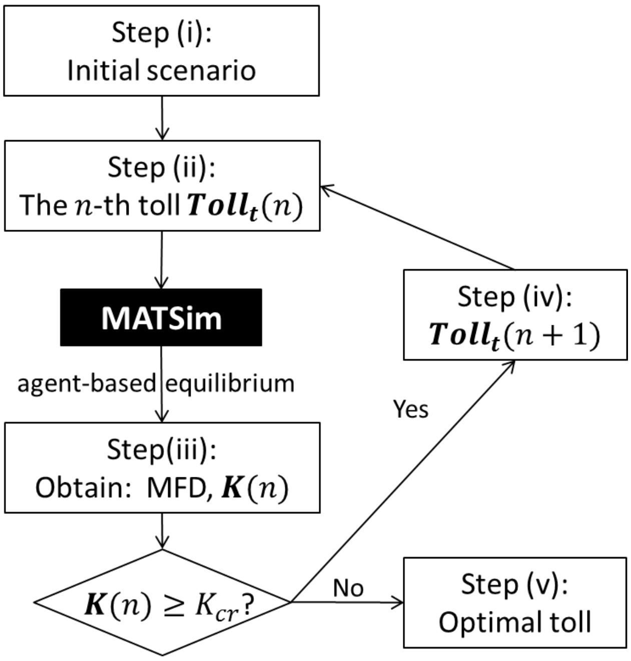
Step (i):
Initial scenario
Step (ii):
The n-th toll $\boldsymbol{Toll_t}(n)$
MATSim
agent-based equilibrium
Step(iii):
Obtain: MFD, $\boldsymbol{K}(n)$
$\boldsymbol{K}(n) \geq K_{cr}$?
Yes
No
Step (iv):
$\boldsymbol{Toll_t}(n+1)$
Step (v):
Optimal toll

Fig. 5.4 An illustration of the optimization algorithm of pricing

Detailed description of the algorithmic steps is provided below:

(i) Initial scenario ($n = 0, Toll_t = 0$).

(ii) Perform the nth simulation with the updated pricings $Toll_t(n)$ until stabilization

(iii) Obtain the resultant MFD. If there exists congested states where $K_t(n) \geq K_{cr}$, go to step (iv), otherwise to step (v).

(iv) Update $Toll_t(n)$ (detailed approach will be elaborated in the next sections). Pass the updated toll $Toll_t(n+1)$ to the $(n+1)$-th simulation.

(v) Terminate the toll adjustment and obtain the optimal pricing.

5.3 A Feedback-type Pricing scheme

5.3.1 The pricing controller

Firstly we propose a proportional controller to update the pricing. This controller is a classic linear feedback control strategy. It is known in control theory that dynamical systems with well-defined properties (e.g. small errors in the state description) can be stabilized with a feedback strategy to a desired state by choosing an appropriate value of parameter c in Equation (5-2):

$$Toll_t(n+1) = max(0, Toll_t(n) + c(\overline{K}_t(n) - K_{cr})) \tag{5-2}$$

Detailed information on proportional controllers can be found in (Otaga, 2001). Basically the controller states that if the average density of the peak hour (pre-defined) $\overline{K}_t$ exceeds K_{cr} after the nth round of pricing adjustments, an additional toll is charged in the $n+1$-th round. This additional value is proportional to the difference between $\overline{K}_t(n)$ and K_{cr}. For practical consideration, time index t here refers to morning and evening peak and two different tolls are estimated accordingly. Parameter c is the constant proportion, which influences the rate of achieving the optimal toll. K_{cr} is a constant as well and estimated from the MFD directly, recalling that K_{cr} is a property of the network. These types of controllers have been widely-applied in freeway traffic flow management, e.g. the well-known ramp-metering algorithm ALINEA utilizes it as the core control law to control flow entering from on-ramp to main freeway. Stability analysis has shown the global convergence of this law: For all choices of controller parameters the strategy preserves closed-loop stability and forces the actual K_t to reach the desired one (Kosmatopoulos and Papageorgiou, 2003). Note that the pricing control we perform here is offline. The objective is to determine an optimal pricing based on day-to-day feedback, and reduce the congestion level for the considered time slot t. The principle remains the same, although the application is different from e.g. ALINEA.

The choice of the value of K_{cr} in Equation (5-2) can be a policy decision. Some might argue that if K_{cr} belongs in regime I the toll is very strict and over-charging might occur, as the system operates in a state less than its capacity. Nevertheless, this does not necessarily indicate over-charging, unless the total toll paid outweigh the total travel time savings. The reason is that the length of the toll period is shorter than the length of the congestion period without toll. This is because the network capacity, as expressed by the MFD, decreases for high values of network density and the system is not operating at the maximum network flow during congestion (this is not the case in the classical bottleneck problem of morning commute (Vickrey, 1969), where toll period is equal to the congestion period). Nevertheless, the system operator can choose the desired toll in a way to maximize the network flow and as a result to transfer the highest possible number of users. The value of toll which results in points in the MFD just without Regime III can be set as the lower bound, while the one which results the total toll paid just equal to total travel time savings can be set as the upper bound in order to avoid over-charging. This higher toll will push the system to operate at a smaller than the maximum outflow during the toll period, but with higher average speed and some potential savings in travel delay. This also implies that we should use a smaller value for K_{cr} in Equation (5-2). For example, state for K_{cr} =20vh/km in Fig. 1.1 is a more reliable and less equitable state, because the average speed is higher (and more likely to be stable), but the system operates at a network outflow below the maximum. Thus, fewer people (presumably those with a higher value of time) pay the toll and travel in the rush hour. State for K_{cr} =35vh/km is more equitable (with a higher flow) but has a slower speed, so the total welfare may be smaller depending on the distribution of the value of time within the population. Also, state with smaller critical density is more reliable because given that users reaction in toll changes is not part of the model, a toll may not be efficient at all times, and allow the system to reach congested states with $\overline{K}_t(n)$ > K_{cr}. In which state of the MFD a city should operate is a policy decision. Also, note that the city operates in the part of MFD with critical density. This means that most of the links during the peak hour operate in Regime II (capacity) and a few in Regime I and III. During the same period without pricing most of the links were in Regime III. While the objectives of existing operational pricing strategies are similar, the MFD gives the quantitative tools to meet these objectives and identify the efficiency and operability of a network.

5.3.2 Case study setup

We now test the effectiveness of the MFD-based pricing scheme proposed in Section 4.1. The data of Scenario 2 are used, which experiences congestion and some scatter. In this way, we expect to provide a stricter test for the effectiveness/robustness of a cordon-based pricing scheme, given that our macroscopic approach is less accurate. The targeted cordon area is a circle of 1km-radius, the area inside the red solid ring in Fig. 5.5. Agents who cross the border of the area, the red line, will pay a toll. Before we apply any toll on the cordon, we run one simulation until it achieves equilibrium. Network density over time is plotted in Fig. 5.6(a), while the resulting MFD is plotted in Fig. 5.6(b) as the “no pricing” curve in these two figures. The following information is calculated, given as input to Equation (5-2): The critical density K_{cr} is 28veh/km; The periods for charging a toll from 7:30am to 9am and from 4pm to 8pm.

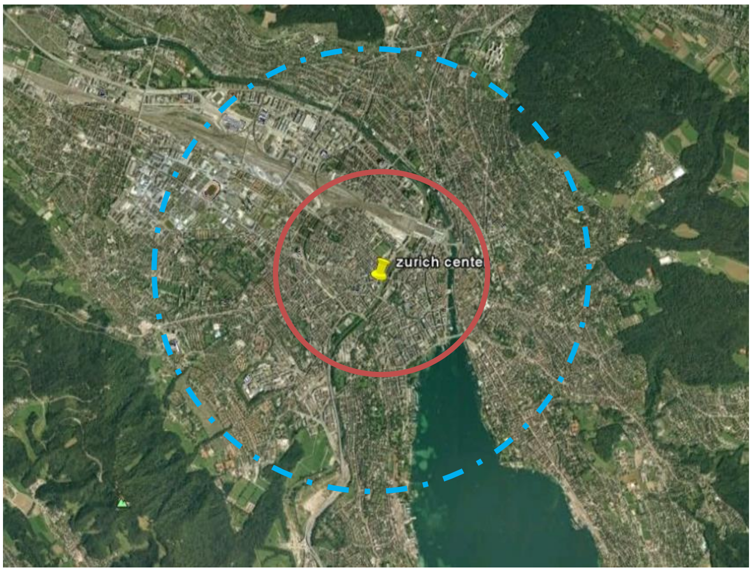

Fig. 5.5 The targeted cordon area (area insider the solid line) and the neighbour area (between the solid line and the dotted line), the city of Zurich

5.3.3 Performance of the pricing scheme

An initial toll of 1€ for the morning while 4€ for the evening are applied in the first trial. A value of 1euro/density for the regulator parameter $\boldsymbol{c}$ is used. The optimal toll is achieved after four updates: a 2€ toll is charged for the morning peak, while 10€ toll is charged between 18:30 and 19:30 and an 8€ is charged for the rest part of the evening peak). We see in Fig. 5.6 (a), traffic congestion, drops as the amount of toll increases; and in Fig. 5.6(b), congestion states in Regime III disappear. Thus this aggregated approach for pricing, which does not consider individual link behaviour, produces the desired results to reduce congestion and identifies the appropriate value of pricing tolls to meet the desired mobility goals. Furthermore, we look at the same graphs for the area outside the charging zone as defined in Fig. 5.5 (1.5km distance from the cordon line, concentric). The motivation is to check if the traffic conditions in the periphery become worse. Fig. 5.6(c) shows that the density of the neighbour area slightly increases. The explanation is that some agents who travelled through the cordon area now choose to not enter it but detour in the neighbour area, in order to avoid the toll. But if we look at the MFD for this area, which is shown in Fig. 5.6(d), the entire area is operating in Regime I and II. Traffic volumes on the links in the cordon area are illustrated in Fig. 5.7, in which red colours represent congested links and green colour uncongested or at capacity. By applying a toll, traffic decreases during peak hours and travel conditions are improved. We now provide a quantitative analysis to estimate delay savings after the implementation of tolls.

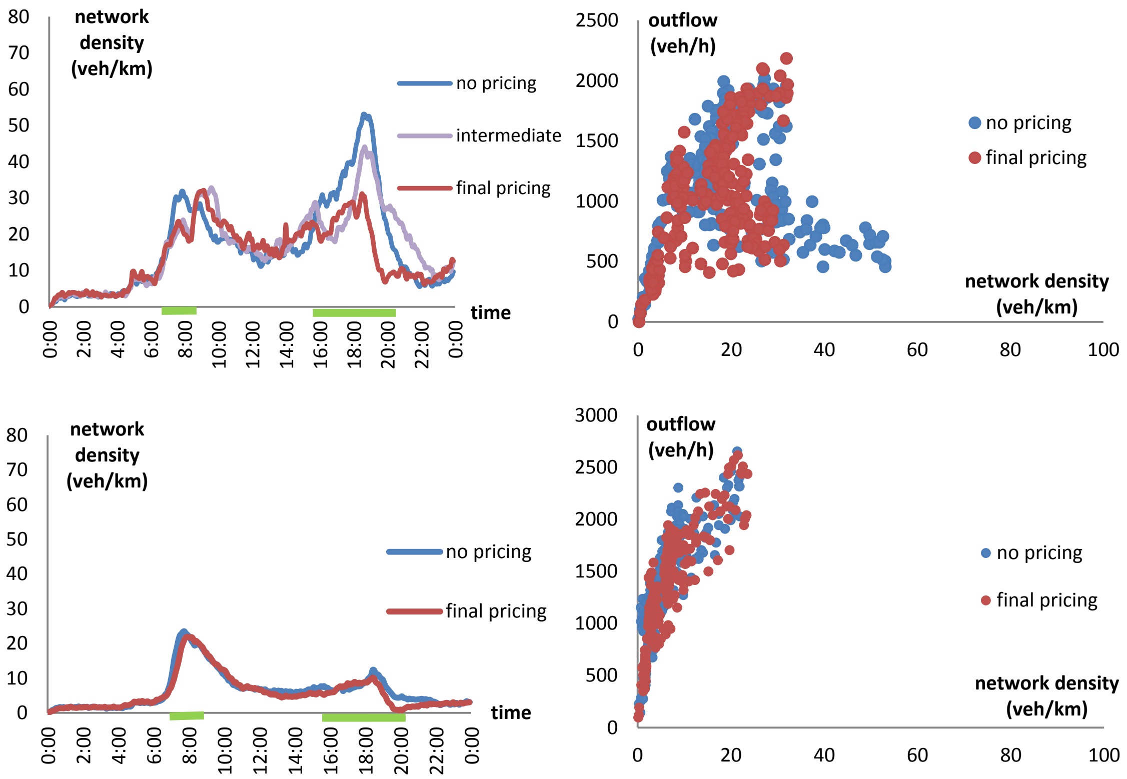

Fig. 5.6 For Cordon area (a- top left) network density time series before and after final pricing; (b-top right) the MFDs before and after the final update of pricing; For outside Cordon (c-down left) density time series before and after the final update of pricing; (d-down right) the MFDs before and after the final update of pricing. The green bars indicate toll period.

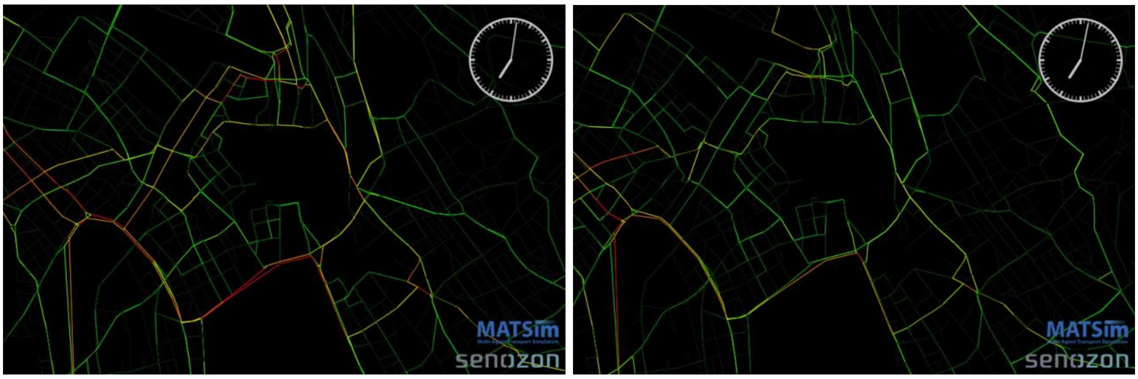

Fig. 5.7 Comparison of the link traffic volumes in (a) the no-toll scenario and (b) the 4th-toll scenario at 19pm

From an economic point of view, we look at travel time savings (as expressed by vehicle hours travelled) and total toll costs. Total travel time savings are estimated as the reduced travel time of the cordon area minus the increased travel time of the periphery. Results are summarized in Table 5.1. Considering the average value of travel time savings (VTTS) of the agents is 15€/hour (Axhausen et al. (2007)), the total savings in the study network are larger than the total toll paid, by an amount of about 20%. Note that most congestion charging methodologies charge a toll equal to the delay cost, while in our case savings are significantly higher. This is a promising result at an aggregated level. Furthermore, we also investigate the savings at an individual user level, which is calculated as travel time savings per km travelled per trip. We see that the effectiveness ratio between travel savings per km travelled per trip inside cordon (positive) and outside cordon (negative) is 7, much higher than the same ratio at the aggregated level (in veh-hours), which is around 1.5. This indicates that a low amount of additional delay (0.13 min per km per trip) which generates almost no impact for the outside cordon, creates significant savings inside the cordon (almost 1 min per km per trip). Zhang and Levinson (2005) among the others argue that the value-of-time should be a non-linear function of travel delay, i.e. the effect of a longer delay (in €/min) is much more significant than the effect of a short delay. Furthermore, comparing the Zurich results to the cordon pricing case of London, the improvement in travel time savings in London is 0.7 min per km per agent (information from Transport for London, http://www.tfl.gov.uk), which shows that our pricing scheme is more effective.

Table 5.1 Summary of the social of the proposed pricing

	TTT Savings (In cordon)	TTT Savings (Out cordon)	Effectiveness Ratio
Aggregated Social Gain	6356 veh-hours	- 4541 veh-hours	1.5
	Travel Savings per Km Travelled per trip (In Cordon)	Travel Savings per Km Travelled per trip (Out Cordon)	Effectiveness Ratio
Disaggregated Social Gain	0.94 minutes	-0.13 minutes	7

5.3.4 Behavioral changes

We analyse the impact of pricing on shifting the time of departure of agent trips. One direct impact is that agents change the time they pass the cordon line. Fig. 5.8 shows a time series of the flow of agents crossing the cordon line with and without pricing. It is clear that some amount of agents avoided passing the cordon line during the tolling-period, which is shown in green colour. In the morning peak, agents tend to switch to later times while in the evening peak agents change their crossing times in both sides of the toll period. We now see how and to what extent agents change their behaviour because of tolls.

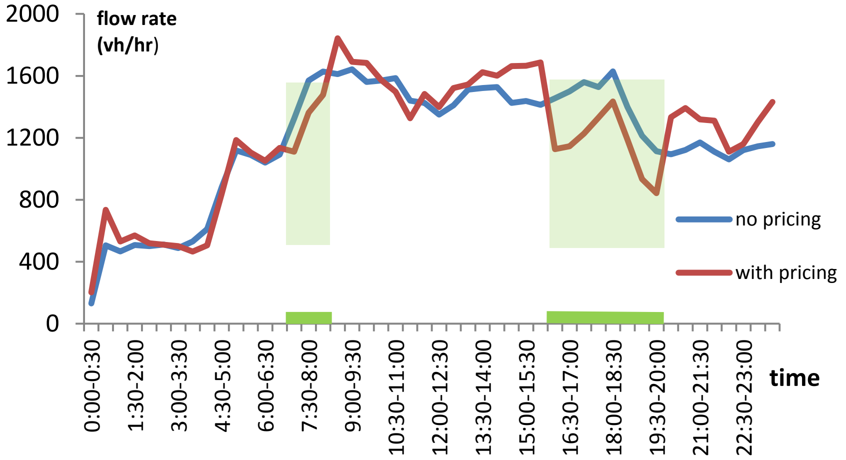

Fig. 5.8 Time series of the flow of agents passing the cordon

In the default version of MATSim, heterogeneity only exists among agents when they perform different activities. This heterogeneity is determined by the duration of the performed activity. Given these information, we can identify the impact of pricing on different activities. We classify activities into two groups: purpose of going-to-work and leaving-from-work, as work-related activities (WA); and purpose of going-to-leisure as non-work-related activities (NWA). Fig. 5.9 shows the comparison of time shift of performing WA and NWA trips in the “no pricing” and the “with pricing” scenarios. The green bars mark the toll periods. Both groups of agents experience a time shift. Agents performing WA tend to switch the starting time of their trip to an earlier time, while agents performing NWA tend to switch to both earlier and later times. The explanation is that for work-related activities the penalty of earliness is lower than this of lateness. About 5% of WA trips shift during the morning toll period while the amount for NWA is 15.7%. During the evening toll period, 16.1% of WA and 19.5% of NWA shift their departure time. We can observe that WA are much less sensitive to small value of toll, as adjusting WA is less flexible than NWA. For higher tolls, the impacts are more substantial. An additional explanation for the observation that the morning shift is less than evening shift, is that starting time of the activities is fixed in the morning, therefore agents have to be on time to avoid a large penalty of not performing their trip (do nothing); while in the evening there is more flexibility. Given the optimal pricing, the number of agents who perform NWA by car outside toll period reduces, which is due to mode change. A deeper investigation on behavioural shift and on equity can be done, if individual data such as income or value of time of agents are available, similar to Axhausen et al. (2007). We will show in later sections, with the integration of the heterogeneity in the distribution of Value-of-time, the behavioural difference can be analysed.

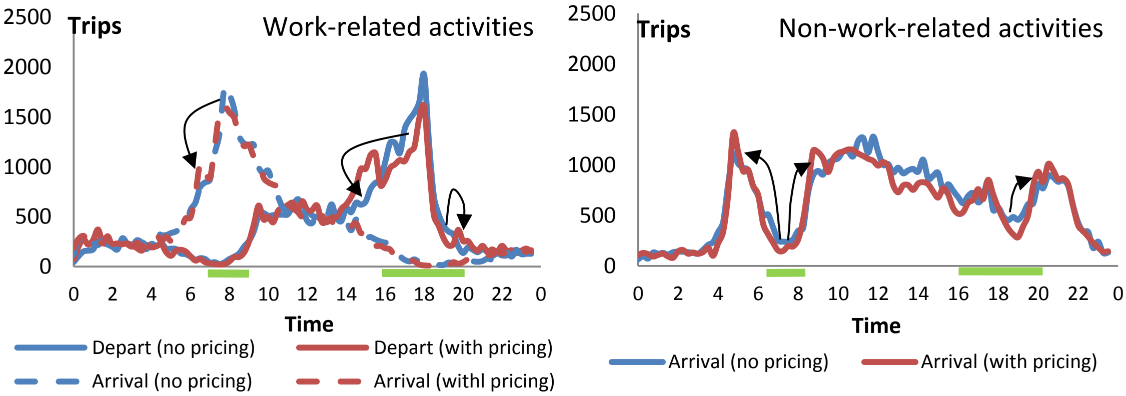

Fig. 5.9 Time shift of WA (a) and NWA (b) in the no pricing and the final pricing scenarios. Y-axis unit, Trips: trip completion rate.

5.4 A Feedback-type Pricing Scheme with User Adaptation

5.4.1 The pricing controller

Secondly, we extend the fixed-peak-hour feedback pricing scheme to a time-dependent one. Furthermore, we aim to include the adaptation of travelers to the change of pricing into the pricing determination process. Once the performance of a network falls into Regime III, control actions should guide the network recover to Regime II around the critical density K_{cr}. Different from online traffic control strategies, where the effect of control on traffic is immediate and drivers might not have the ability to adapt, the effect of pricing can be long-term and depends on how users change their travel behavior. Pricing could make users to change their departure time from the origin or change mode of transport. Considering this adaptation in the development and design of efficient pricing schemes could be challenging. As the adaptation is quite difficult to model in an analytical way, we consider that an agent-based framework influences their decisions in an unknown way to the system operators. Nevertheless, the operators can monitor the aggregated behavior as network vehicle accumulations are known.

The algorithmic steps are the same as given in Fig. 5.4. And controller (5-2) is extended in Equation (5-3). The toll at the $(j+1)$-th toll adjustment for time interval $t, Toll_{j+1}(t)$, is proportional to the magnitude that the average network density $K_j(t)$ exceeds K_{cr}, and the difference between the resultant densities under the current pricing $K_j(t)$ and its previous one $K_{j-1}(t)$. c_1 and c_2 are the proportional and integral gain parameters, respectively.

$$Toll_t(n+1) = Toll_t(n) + c_1(K_t(n) - K_{cr}) + c_2(K_t(n) - K_t(n-1)) \qquad (5\text{-}3)$$

The toll-updating scheme employs a Proportional-integral (PI) type controller which is a classic feedback control strategy. It was demonstrated that dynamic systems with well-defined properties (e.g. small errors in the traffic states) can be stabilized with this type of feedback strategy to a desired state, as well. Comparing to the toll controller (5-2), controller (5-3) distinguishes temporal difference in congestion level, and gives higher flexibility in toll adjustment based on user's adaption to the toll. In particular, this schemes allows reduction of toll if $K_j(t)$ is smaller than $K_{j-1}(t)$ which benefits the users from being not overcharged. c_1 and c_2 are constant and positive parameters. The values are chosen offline via a trial-and-error process to avoid oscillations. Details on the design of the controller will not be discussed here. We would emphasize that this pricing strategy is robust to moderate parameter changes and has fast and global convergence.

5.4.2 Case study setup

This case study is carried out in a similar scale network, the well-known Sioux Fall urban network, though the activity patterns are quite different from the Zurich case. The structure of the network captures the major arterial roads of the real city: highways with 3 lanes per direction at the perimeter of the network and urban roads with 2 lanes in the city center. The average link length is 0.4km. The total length of the network is 150 lane-kms. As a general remark, the scale of the network is suitable for applying MFD-based analysis. The embedded public transport system consists of 5 bus lanes crossing the network and serving 10 bus routes. The spacing between bus stops is 0.6km and the frequency of service is in between 5 to 15min.The studied network area holds a total amount of 110000 travelers. The daily plans of the travelers are created based on detailed regional census data. The generated traffic demand exhibits typical morning- and evening-peak characteristics. For more information on the design of the Sioux Fall simulation scenario for this study, readers may refer to Chakirov and Fourie (2014).

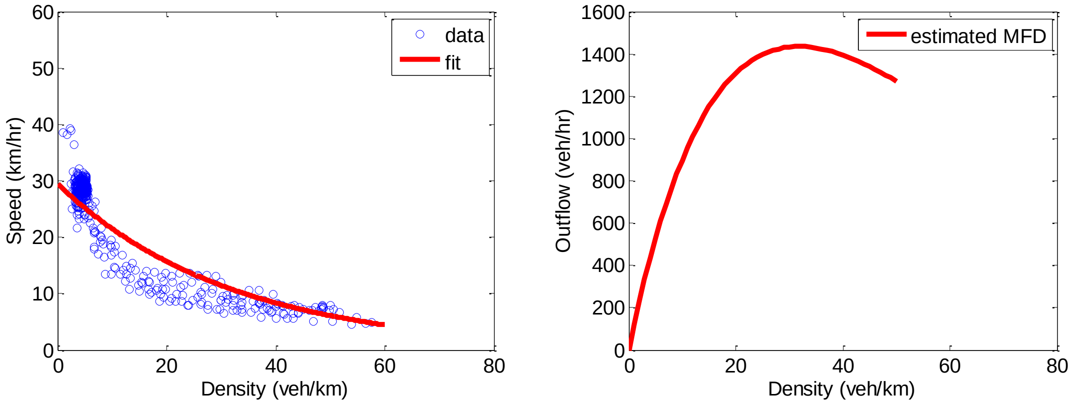

Fig. 5.10 The MFD of the study site: (a) network speed-density data scatter and the fitted curve, $V = 29.48e^{-0.032K}$ (R-square test of the fitting is 0.85) and (b) best-fit network flow-density plot

Parameter K_{cr} is obtained by simulation data of this network. Speed and density data are collected from scenarios with different demand profiles and plotted in Fig. 5.10(a). Each data point corresponds to 5-min interval. This speed-density curve is fitted by an exponential-family function, which is consistent with the classical form of the fundamental diagram and the one proposed in Section 4.2. Given the fitted function and the flow dynamic equation $Q = kv$, where Q, K, V are network space mean flow, density and speed respectively, the MFD of Q and K can be obtained. Fig. 5.10(b) displays this MFD. K_{cr} is estimated analytically based on the exponential family curve and is approximately 35veh/km.

We will look at the resultant network density level, and the total person-hours travelled (PHT), the impact of pricing on multimodal mobility, the efficiency and welfare gain where welfare is defined as the difference between the savings of PHT and the total toll paid (TTP). To have the same unit of the two items, TTP is converted to time unit by value-of-time, denoted as vTTP. The value-of-time is 16$/hr. A positive gain is desirable because it may not be efficient if a pricing strategy reduces congestion at the high expense of users. Moreover, we examine traveler’s behavioral adaptation to the pricing.

5.4.3 Performance of the pricing scheme

Let us first look at the traffic performance of the network without and with pricing. Simulations are executed and an optimal toll is found from the toll scheme described in the previous section. Fig. 5.11 displays the density time series without and with pricing (left axes). The time series of the final optimal pricing is plotted in the same figure (right axes). It gives a 0.8$/trip toll for the morning peak between 7h30 and 8h30, and a higher step toll for the evening peak between 17h-19h (a value of 0.5$/(veh/km) and 0.3$/(veh/km) is used for $\boldsymbol{c_1}$ and $\boldsymbol{c_2}$ in Equation (5-3)). For computational feasibility consideration, the tolling interval is set to 0.5 hr.

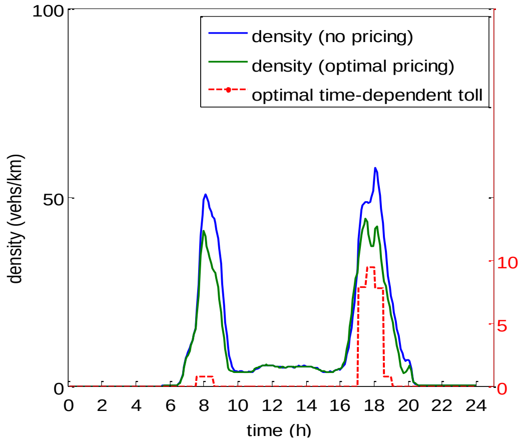

Fig. 5.11 Density time series without and with optimal time-dependent pricing. The dashed line indicates the final toll rates over the day

We can observe that the congestion level of the network, as expressed by density, decreases significantly for both the morning and the evening peaks, albeit with different toll rates. It is interesting to see that roughly similar amount of car usage reduction can be achieved in the morning with much smaller toll rate than in the evening. To understand this result, let us have a closer look on how densities and the corresponding pricing rates change over simulations. Fig. 5.12 illustrates the evolution of densities and toll rates over three simulations. Fig. 5.12(a) shows that the initial tolls are equally high for both peak periods. The morning car traffic adapts quickly to the increased cost due to pricing, and the density level decreases to slightly above K_{cr}. Since our pricing scheme adjusts the tolls according to user's adaptation, the morning toll is thus decreased so that car users are not overcharged. We can also observe that there is almost no induced demand for the morning peak, even though the toll rates become less comparing to the initial toll. While for the evening peak, higher toll must be given and kept for two reasons: (i) travelers have to travel during the peak hour as the peak hour overlaps the closing time of nearly all the facilities where they perform activities. Leaving activities early would experience "schedule penalty" which is larger than the cost of pricing; and (ii) the PT mode is not sufficiently fast to trigger mode shift until the cost of pricing is greater than the difference in travel times. Therefore many travelers choose to travel anyway which weakens the elasticity of demand towards pricing.

Pricing adjustment to user's adaptation can be observed for the evening peak, as well. As the implementation of the initial toll decreases network density, a considerable drop of the toll rate can be found at the 3rd toll in Fig. 5.12(b). Given this adjustment, it seems that travelers begin to switch back to the peak period and take advantage of the improved network condition. Our pricing scheme identifies this trend. The same figure shows that a higher toll is re-implemented from the 6th pricing adjustment. Note that a pricing adjustment takes place approximately every 50-iterations of the agent-based simulation. A relevant question now is that how sensitive are the travelers to the change of tolls. We will address this later.

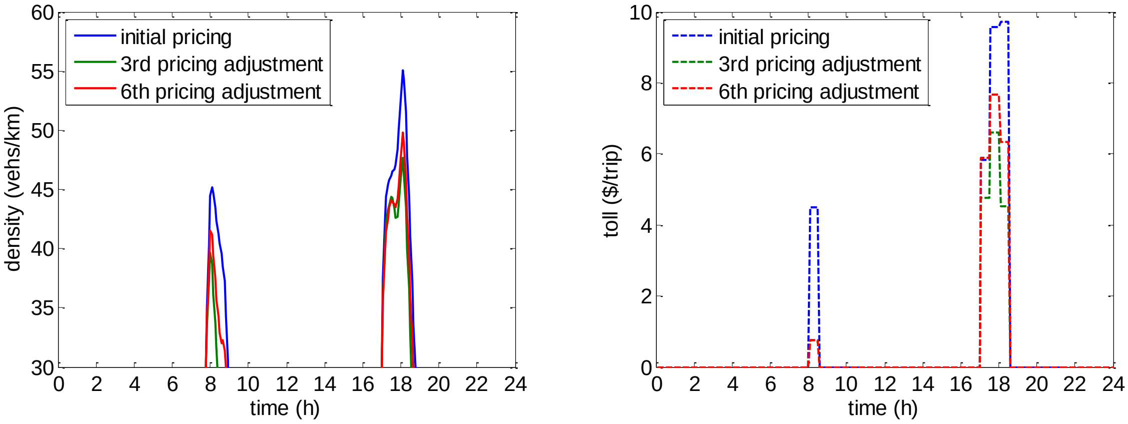

Fig. 5.12 (a) The density series and (b) toll rates of three simulations before reaching the optimal pricing

Table 5.2 Comparison of total and average travel time after pricing

	Total PHT (hrs)	Ave. TT. (hrs/trip)	
		Car	PT
Without pricing	73840	0.58	0.87
Optimal pricing	68433	0.47	0.88

To quantify the impact of pricing on traffic performance, we calculate the improvement in travel time over toll adjustments, at aggregated and disaggregated scales between "Without pricing" and "Optimal pricing". The statistics are summarized in Table 5.2. The total PHT decreases by 7%, and a 17% for average travel time by car. There is however no improvement for traveling by PT, which is understandable as the same level of PT service now attracts more users. To improve the mobility of multimodal network, a pricing strategy should promote the development of PT services and stimulate mode shift. To this end, we decide to utilize a fraction of the toll revenue such that it serves as an incentive for higher PT usage. We will show in the next sub-section that substantial welfare gain can be obtained with such incentives.

5.4.4 PT benefit and accessibility improvement

Continuing the discussion above, we seek to enlarge the efficiency of the pricing scheme by promote PT-oriented strategies. We propose to distribute the collected TTP in two ways: (i) to expand the number of bus lines and bus networks so that the accessibility cost (e.g. access time from home to bus stops) decreases, and (ii) to provide a one-time award money for travelers who shift the mode from cars to buses. In practice, the design and operation of such incentive programs are complex and highly dependent on political, financial and other purposes. We do not wish to deepen ourselves into this direction of discussion, but to demonstrate via compact analysis that proper treatment of toll revenue on PT system and PT users can produce critical impact on the efficiency of the pricing schemes.

In Table 5.3, the total cost and welfare gain are estimated for the three pricing strategies: (i) the pricing scheme defined by Equation (5-3), denote as “pricing only”; (ii) a fraction of pricing revenue is invested to improve PT accessibility, e.g. the resultant average access time from home to bus stops are assumed to decrease 15%. Denote this scheme as “pricing+BA”; and (iii) in addition to strategy (ii), 50% of the toll revenue is rewarded one-time and evenly distributed to the mode shifters. Denote as “pricing+BAI”. Since a fraction of the collected toll is returned to users, vTTP shall not be entirely viewed as cost term for strategies (ii) and (iii). In Table 5.3, the pricing efficiency is calculated (recall that this is the ratio of total savings in PHT compare to that of without pricing (73840 in Table 5.2) and vTTP). It is evident that congestion pricing can be remarkably efficient if the service of PT is improved using the toll revenue. With a full investment of the toll revenue, the “pricing+BAI” strategy may expect 100% efficiency.

Table 5.3 Comparison of PHT savings and pricing efficiency

	Savings in PHT (hrs)	vTTP (hrs)	Pricing efficiency (%)
pricing only	5627	14790	38.1%
pricing+BA	9673	13437	72.0%
pricing+BAI	9090	9177	99.0%

With the PT benefit, congestion level of the network can be controlled at K_{cr} by a smaller amount of toll. For example, Fig. 5.13 (a) and (b) compare the time series of density and toll between the "pricing only" strategy and the "pricing+BA" strategy. The "pricing+BA" succeeds in maintaining network density under 40vehs/km with 20% less toll charge. For mode share of PT users under different pricing strategies, it is as expected that the number of PT users increases with the magnitude of PT benefits. Strategy "pricing BA" for instance, results a 5% increase in mode shift than "pricing only" and a nearly 10% increase to "without pricing".

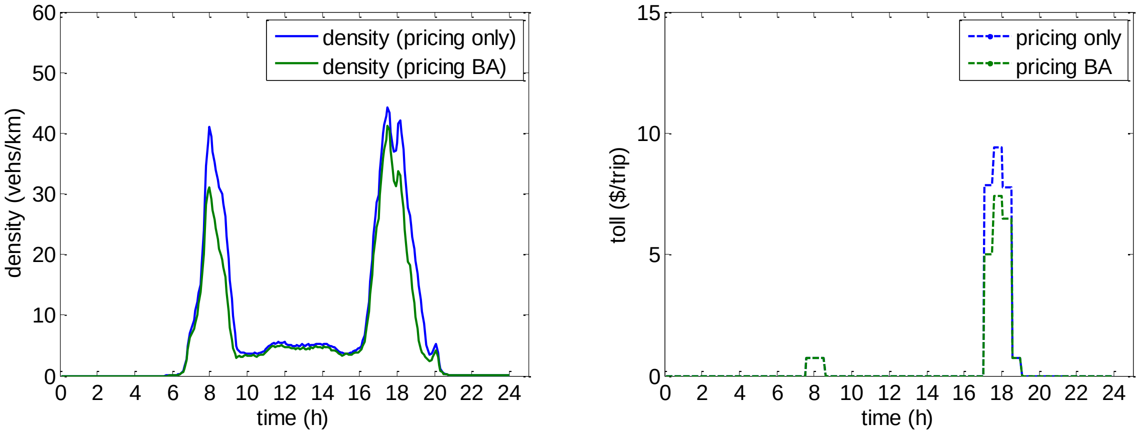

Fig. 5.13 Comparison of (a) density series and (b) resultant toll between "pricing only" and "pricing BA"

5.4.5 User adaptation

Finally, let us address the question on how sensitive the travelers react to pricing and if the proposed pricing scheme reaches equilibrium. To this end, we investigate the shape of the total toll paid (TTP) and the total number of mode shifters over toll adjustments. Fig. 5.14(a) plots TTP for strategies "pricing only" and "pricing+BA" over the entire iterations. The first pricing is implemented at the 100th iteration. Since then, the tolls are adjusted after every 50 iterations (recall that 50 iterations per simulation). It can be seen that travelers struggle to adapt the tolls when the tolls are initially introduced, reflected by the huge jumps at the 150th and the 200th iterations. Similar phenomenon can be observed in Fig. 5.14(b), where the number of mode shifters (from car to bus) is plotted over iterations for the two cases. After the 3rd adjustment of the toll (the 250th iteration), however, both TTP and mode shifts vary slowly over toll adjustments. With sufficient amount of iterations, it can be expected that a user-equilibrium could be achieved. These are intriguing results, as they indicate that our pricing scheme shows high convergence in the long-term operation of pricing. Small level of induced demand might appear from time to time, similar to we have observed in Fig. 5.12, nevertheless neither congestion would not return nor travelers would perform drastic behavioral change as the pricing scheme of Section II is designed to be behavioral adaptive.

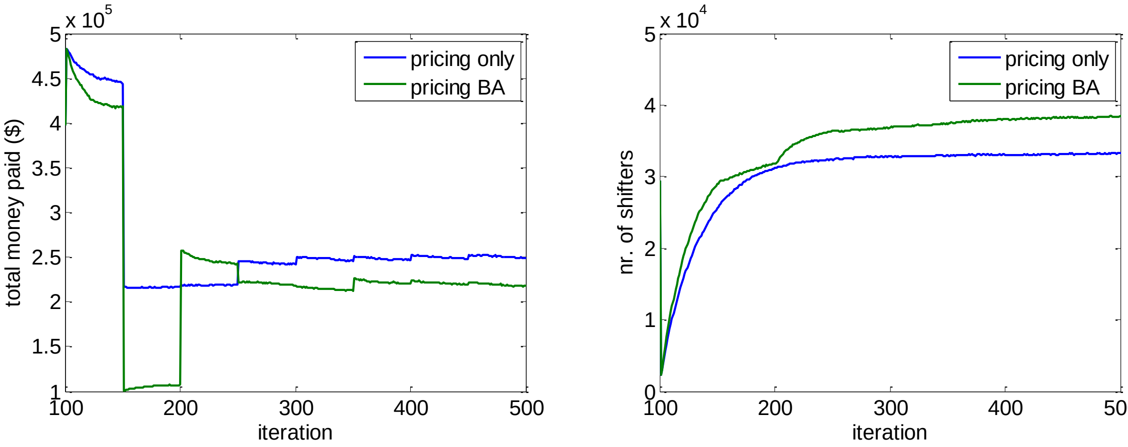

Fig. 5.14 Equilibrium condition of the pricing schemes: (a) total money paid over iterations, and (b) mode shifters over iterations

5.5 Impact of User Heterogeneity

So far we have focused our discussion on the global performance. In this section, we will investigate the disaggregated impact of pricing on different groups of users in the system. There exists works in literature on the treatment of heterogeneity of travelers. The distribution of VOT among travelers has been the main focus, for instance in Lu et al. (2008), van der Berg and Verhoef (2011), and Qian and Zhang (2013). Given the advantage of the agent-based approaches and the efficiency of the MFD-based pricing, we extend the homogenous single-user-group MATSim to a two-group one. Thus for Equation (5-1), index j equals to 2. There are a group of users with a high value of VOT, which takes one third of the whole population, and a group with low VOT for the rest of the population. We consider the mean VOT for the entire population has a value of 16$/hr, and determine the VOT for the two groups as 27$/hr and 11$/hr respectively. The VOTs are part of travelers' decision-making and planning process in Equation (5-1).

5.5.1 Behavioral difference

Let us first show the behavioral difference after introducing heterogeneity, without implementation of pricing. The mode share of buses is displayed in Fig. 5.15(a) for scenario without user heterogeneity (a single user group) and the scenario of two groups defined earlier.

Denote the user group with high VOT the “R group” and the one with low VOT the “P group” (as rich vs. poor groups). It is observed that the mode choice behavior is quite different when heterogeneity is taken into account. While the mode share of buses is identical for the single-group scenario, there is nearly 10% difference between the R group and the P group when two VOTs are active. The R group tends to prefer traveling with cars, as it is less costly and travel time are valued much higher for the users of this group than those of the P group. Since the P group now has a smaller VOT comparing to their VOT in the single-group scenario, more users are willing to travel with bus even if the utility difference remains the same between travelling with buses and with cars.

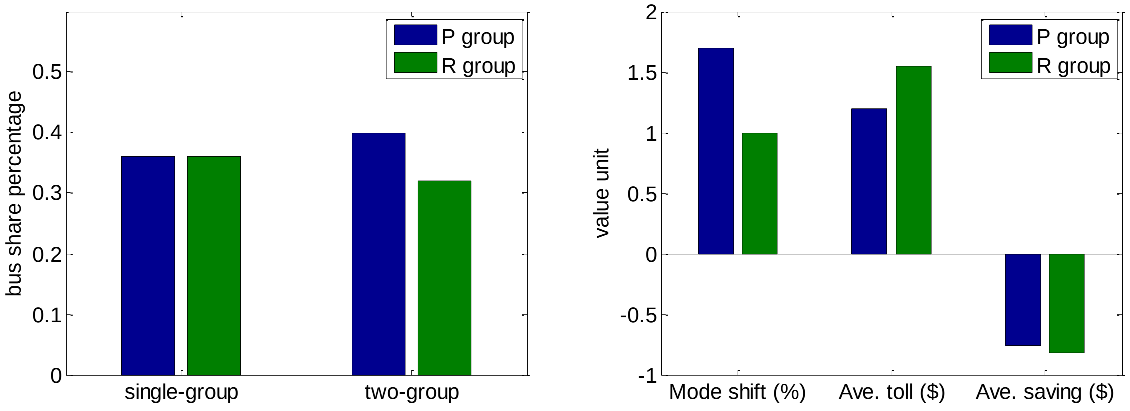

Fig. 5.15 (a) Mode choice difference of the two user groups, without congestion pricing, and (b) comparison of mode shift, toll paid and savings between the two user groups under strategy “pricing”.

When congestion pricing is applied, the two user groups exhibit distinct reactions as expected. Fig. 5.15(b) displays the comparison of the resultant mode shift (the percentage of users shift from cars to buses after pricing), average congestion toll paid per person (total toll paid divided by the total amount of the users), and the average savings per person (welfare divided by the amount of the users) under strategy “pricing”. For mode choice behavior, more users from the P group tends to switch to buses though the numerical result suggests a minor difference between the two groups. The resultant average toll indicates that the R group users are willing to pay 0.3$ more toll for traveling faster with cars. The resultant saving is negative for the entire population, and almost equivalent for the two groups of users. Nevertheless, different behavior for the two groups of users is observed in the other pricing schemes.

5.5.2 Distributional effect under different pricing strategies

The results on VOT heterogeneity shown in the previous sub-section are consistent with findings in literature summarized in Levinson (2010). We will now investigate the resultant equity of the pricing strategies. Fig. 5.16(a) illustrates the mode shift by strategies “pricing”, “pricing BA” and “pricing BAIp” (recall that strategy “pricing BAIp” distributes subsidy only for the mode shifters from the “P group”). Mode shift here is the mode shift to buses under each of the three strategies compared to the “no pricing” case. As the incentive of using buses arises, there is an increasing demand for both groups to travel with bus, because of pricing and the attraction by PT benefit. For the R group, it is evident that users are reluctant to change mode even though traveling with bus becomes less costly. Mode shift is below 1% under strategy “pricing” while less than 2.5% under the other two pricing strategies. This is probably due to the fact that traveling with cars becomes less costly at the same time as more users travel with buses, and therefore traveling with cars becomes attractive for the R group who values the gain in travel time much more important. While for the P group, mode shift shows a clear tendency of growth as strategies provide more PT benefit. Under strategies “pricing” and “pricing BA”, the mode shifts are 1.8% and 4.5% each and nearly doubles the R group; while under “pricing BAIp” the percentage goes to 6.8% which nearly triples the P group. Comparing to strategy “pricing BA”, “pricing BAIp” makes 3% more users of the R group switch their mode. We will show next that this strategy is not only efficient in triggering mode shift but also advantageous in equity.

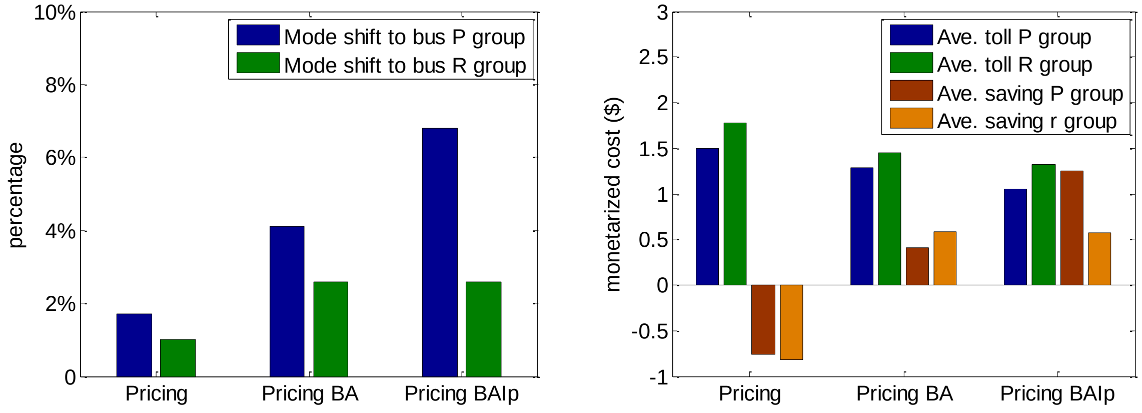

Fig. 5.16 (a) Comparison of the difference in mode shifts and (b) comparison of mode shift, toll paid and savings the two user groups among the three pricing strategies: pricing, pricing BA and pricing BAIp.

Fig. 5.16(b) illustrates the comparison of average toll paid and average savings for the two user groups under the three pricing strategies. In all cases, the R group users are willing to pay higher toll to reserve the right of traveling with cars. It seems both user groups benefit from the PT incentives, since they all find positive savings. The savings of the P group users outweighs the R group by 100% when subsidy only provided to the R group. As the users of the P group is generally considered to represent the users with lower income, strategy “BAIp” is undoubtedly a more equitable pricing scheme. These results indicate that heterogeneity of users and distributional effect should be treated carefully when designing strategies and policies of congestion pricing, so that beneficial and equitable condition can be achieved for everybody.

5.6 Summary

In this chapter, we combined an agent-based approach with MFD to develop and evaluate city-level congestion pricing schemes. We first investigated the feasibility of the combined approach by examining whether the outputs of the agent-based model exhibits traffic patterns as expressed in the fundamental diagram at road section level and the MFD at network level. We found and concluded that they were consistent with the existing traffic flow theories. We then developed cordon-based and area-based pricing strategies for congested multimodal urban networks. Two feedback-type control logics were proposed to determine the pricing schemes. Stimulating results were found for both pricing schemes in congestion reduction and social welfare optimization, with case studies on the Zurich center region network and the Sioux-Fall network respectively. The second pricing scheme considered user’s adaptation to the toll cost, allowing a greater flexibility in toll adjustment. The scheme also dealt with the promotion of PT usage. Integrating incentive programs such as PT accessibility improvement or money award for PT mode shift using the collected toll revenue, this pricing scheme achieved high efficiency as large welfare gain is obtained. Remarkably, smooth user behavioral equilibrium in long-term operation was also found under such pricing scheme. Furthermore, we investigated the impact of pricing on the traffic performance in a system of heterogeneous user groups. Two groups were recognized with respect to their value-of-times. Significant differences in behavioral responses and trip costs were found for the two groups. By realizing user heterogeneity, pricing strategies achieved higher efficiency and equitable result.

6 Conclusions

Developing effective, efficient and sustainable traffic management strategies for multimodal urban networks are challenging tasks given the complex dynamics of the transport system. This thesis is dedicated to develop a network-level approach which unveils the fundamental laws of congestion dynamics in multimodal networks despite of the complexity, and facilitates the design of management strategies for mobility and multimodality optimization. The findings and contributions are summarized in this final chapter.

6.1 Thesis Summary

In Chapter 2, we presented a macroscopic approach for allocating road space among modes of transport with the objective of minimizing the total hours traveled by travelers (PHT). We extended the single-mode MFD to a bi-modal one, where the effect of bi-modal operation in global performance was considered. For example, the effect of dwell time and the share of dedicated bus lanes (of the total space) on network space-mean speed were quantified by the bi-modal MFD. A system model then was constructed for a multi-region network, with a network-level flow conservation model and an aggregated dynamic mode choice model. Given this system model, the performance under different road space strategies can be predicted, provided with data input that can be readily collected such as regional origin-destination tables and road space allocation plans. We tested the rationale of this modeling approach with a two-region bi-modal city case study, and investigated the performance of two space allocation strategies for the center region of the city where demand was high and heavy congestion existed. A static and a dynamic (time-dependent) allocation strategies were obtained through non-linear optimizations respectively. We found that the dynamic allocation strategy managed to minimize PHT in a more efficient way as it utilized the bus lane space during the off-peak period and served a higher amount of passengers during peak period. By implementing pricing during the peak period, further reduction of PHT was achieved, as pricing led a higher demand shifted and increased the utilization of buses which was the faster traveling mode. We also observed the existence of user equilibrium in our macroscopic model, e.g. the utilities of traveling by modes became equivalent during the peak-period. Furthermore, we carried out sensitivity analysis and showed the robustness of our approach towards the fluctuations in demand inputs and model parameters.

Following Chapter 2, we proposed in Chapter 0 a macroscopic modeling approach for modeling multi-modal traffic system with parking limitation and cruising-for-parking flow. Parking limitation was integrated in the developed multi-modal system model, where vehicles need to cruise for parking before reaching destination. The cost of cruising was estimated by assuming the probability of finding a parking space follows a geometric distribution and depends on the dynamic parking availability. The effect of cruising on the global performance, e.g. the average speed, was also captured, by the MFD. A case study was carried out in the same two-region bi-modal network as in Chapter 2. Two parking choices were considered: (i) limited on-street parking requiring cruising, and (ii) unlimited garage parking with higher parking fee but no cruising cost. The resultant system behavior under parking limitation and pricing were consistent with the common expectations. We then used the system model to test two network-level parking pricing strategies. One strategy adapted a feedback-type controller for determining the parking price, which was congestion- and parking availability-dependent. Applying this pricing strategy, traffic performance was maintained at desired (controlled) levels. The second strategy was obtained through optimization of the total cost (PHT + parking fee). Applying this strategy, the total cost was further reduced, as the prices were determined with long-term impact taking into account. Inspired by this result, we investigated the impact of competition on the performance of the pricing strategy, assuming that the authorities of on-street and garage parking belong to different parties who manage prices with different objectives. We presented preliminary results of cooperative-competition via a bi-objective optimization, while a bi-level optimization framework was proposed to simulate a responsive and negotiate-alike parking pricing market.

To identify and quantify individual modal impact on the global traffic performance, we investigated a new type of MFD model in Chapter 4. The existence of a three-dimensional multimodal MFD (3D-MFD), relating the accumulation of cars and buses, and the circulating vehicle flow in a network, was demonstrated via simulation experiments. An exponential-family function was proposed for the analytical form of the 3D-MFD, where the individual modal and the joint impact on global performance were directly observed. To further investigate the modal impact, the Bus-Car Unit equivalent value was estimated and found state- and mode-composition-dependent rather than deterministic. Then, we derived the passenger-flow 3D-MFD with an elegant analytical model, which provided a different perspective of bi-modal flow characteristics. We applied a partitioning algorithm to cluster a center and an outside region of the network according to the similarity of the car/bus density ratio. It was found that such partition enhanced the accuracy in the estimation of the 3D-MFDs, which revealed the impact of heterogeneity of mode composition on the bi-modal modeling. Furthermore, we utilized the 3D-MFD to develop perimeter flow control strategies. We presented the results of two controllers, a single-region controller and a two-region (center-outside) controller. Congestion was significantly reduced for the whole network, while the performance of buses in terms of travel delays and schedule reliability was improved without even giving bus priority. Queues and gridlock were avoided on critical paths of the network.

In the final chapter, we combined an agent-based approach with the MFD to develop and evaluate city-level congestion pricing schemes. We first investigated the feasibility of the combined approach by examining whether the outputs of the agent-based model exhibits traffic patterns as expressed in the fundamental diagram at road section level and the MFD at network level. We found and concluded that they were consistent with the existing traffic flow theories. We then developed cordon-based and area-based pricing strategies for congested multimodal urban networks. Two feedback-type control logics were proposed to determine the pricing schemes. Stimulating results were found for both pricing schemes in congestion reduction and social welfare optimization, with case studies on the Zurich center region network and the Sioux-Fall network respectively. The second pricing scheme considered user’s adaptation to the toll cost, allowing a greater flexibility in toll adjustment. The scheme also dealt with the promotion of PT usage. Integrating incentive programs such as PT accessibility improvement or money award for PT mode shift using the collected toll revenue, this pricing scheme achieved high efficiency as large welfare gain was obtained. Remarkably, smooth user behavioral equilibrium in long-term operation was also found under such pricing scheme. Furthermore, we investigated the impact of pricing on the traffic performance in a system of heterogeneous user groups. Two groups were recognized with respect to their value-of-times. Significant differences in behavioral responses and trip costs were found for the two groups. By realizing user heterogeneity, pricing strategies achieved higher efficiency and equitable result.

6.2 Research Contributions

In this thesis, a dynamic network-level approach is developed for modeling and controlling multi-modal urban mobility. It is demonstrated that the developed approach not only captures congestion dynamics, travel costs and mode conflict patterns under various network structures, road space allocations and multi-modal operations, but also enables the development of efficient and dynamic traffic management strategies, such as dedicated bus lane allocation, parking pricing, real-time perimeter flow control and congestion pricing. The contributions of this thesis therefore are both methodological and practical.

Regarding the methodological part, the major contributions can be listed as follows:

- A bi-modal MFD model is developed for capturing the congestion dynamics, considering the effect of bus operation in the global traffic performance in bi-modal urban road networks.
- An aggregated system model is developed to represent the aggregated flow movements and travel cost in a multi-modal multi-region transport system, where system performance is linked to road space allocation and parking is integrated to the network-level dynamics.
- A 3D-MFD model is proposed for capturing the individual modal impact on the global traffic performance and mode conflict patterns in mixed-traffic bi-modal urban networks. A passenger flow 3D-MFD is also derived and it enables a passenger-based analysis.
- An MFD- and agent-based approach is proposed for developing dynamic city-level congestion pricing schemes for multimodal urban networks with heterogeneous users.

For managing mobility and multimodality in practice, the most significant contributions are:

- An optimization framework is established for the developed bi-nodal system model with the objective of minimizing the total travel cost of passengers. Road space allocation and parking pricing strategies for bi-modal networks can be obtained.
- The 3D-MFD can be utilized to monitor performance, and develop real-time control strategies, such as perimeter flow control, for networks with bi-modal usages.
- The MFD-based congestion pricing schemes increase the multimodality level of the priced networks by integrating the incentive programs of using public transport modes.
- MFD-based management strategies are in favor of field applications given the fact that their efficiency in congestion management is achieved with feasible data requirements and implementation costs.

Other important contributions include:

- Based on the developed bi-modal system model, a bi-objective and bi-level optimization frameworks are proposed to simulate the competition behavior in the parking pricing market.

- Network partitioning is applied to identify the heterogeneity in mode composition and congestion level in bi-modal networks. Recognizing this heterogeneity improves the accuracy of modeling and the efficiency of control.
- The 3D-MFD derives a state- and mode composition-dependent “bus-car equivalent” value, while it is considered constant in the literature.
- An MFD-based pricing scheme can result equitable benefit to users when heterogeneity is taken into account and treated properly.

6.3 Future Work

Thanks to the emergence and advance in the macroscopic traffic flow theory, the bi-modal MFD and the 3D-MFDs that are developed in this thesis, and show a stimulating strength in congestion modeling and mobility management. To the best knowledge of the author, MFD-based monitoring and management are being implemented or planned to be implemented in field tests to deal with mobility problems. We need to point out that there are still aspects deserve and require further research efforts, which are both academically challenging and practically interesting.

To keep the aggregated nature of our model tractable, the dedicated bus lanes and parking spaces are assumed to be distributed evenly over a network. We optimize the percentage of road space allocation while the detailed bus lane assignment is not captured in our model (considered a political decision). If the access cost of different space allocation layouts is provided, it can he inputted directly to the system model. As for cruising, the treatment is dimensionless, meaning that the cruising cost is estimated without identifying the detailed routes of cruising. Such consideration aims at the average cost at system level and ignores the cruising possibilities at disaggregated level. Nevertheless large disaggregated heterogeneity is not expected given the spatial correlations in the distribution of congestion and well-defined bi-modal MFD can possibly be found for more complex city structures.

We will continue to study the effectiveness of the space allocation and parking pricing strategies optimized based on our aggregated system model. The system dynamics under the developed strategies ought to be verified by other type of approaches. For example, empirical studies are extremely desired to demonstrate the relation between the cost of cruising and the parking space availability. Another alternative is to test our result in microscopic traffic models in order to investigate if the findings by macroscopic models are consistent with the outputs by traditional microscopic models. Such accordance will strongly support the soundness of our methodology.

For homogeneous large-scale networks where trips are completed internally or center-periphery structure type networks where trips concern two regions, the MFD representation allows modeling without detailed treatment of route choice. However, for trips that across multiple regions (e.g. due to long travel distance), a region sequence choice should be provided to identify the dynamics. We aim to incorporate such regional route choice model in the dynamic network model with multiple regions. Ongoing work in Lab Urban Transport Systems at EPFL investigates such regional route choice pattern via empirical data and simulation studies.

The proposed system model deals with mode choices of travelers at the aggregated level. By influencing mode choice in a timely dynamic manner, the total cost is minimized. Though only two modes are treated in detail in this thesis, extending the framework to multimodal is not difficult. Furthermore, we have showed in our case studies that under such optimal management, travel costs (utility) by mode can reach convergence. Future work will also try to incorporate departure time choices. The classical Vikrey's bottleneck model may be applied here with MFD as the traffic state and capacity model. Convergence might be an issue and equilibrium may not be obtained straightforward through iterations, given the increased complexity in system dynamics. As for parking facility choices, a further direction is to investigate the effect of parking space capacity. For instance, excessive on-street parking space may lead to cases that fewer travelers go for garage parking even though it is much cheaper.

The existence of the 3D-MFD in this thesis is observed via microscopic traffic simulation due to lack of empirical data from multiple modes for the same network. Evidences from empirical studies on its shape, dynamics nature and properties are highly desired. Though this is a challenging task due to the data availability issue (obtaining the 3D-MFD requires data inputs such as traffic accounts, floating GPS data, which belong to different authorities), we are making significant amount of efforts in pursuing possible data sources. Furthermore, the performance of the 3D-MFD based perimeter flow control will be tested in other type of networks, where the mode composition pattern and congestion distribution are significantly different from the San Francisco network used in this thesis.

As for the pricing schemes, we demonstrate that efficient pricing schemes can be developed based on reliable network-level performance monitoring, even though complex behavior dynamics and uncertainties exist which make it impossible to obtain optimal tolls via traditional pricing approaches (e.g. marginal cost based). We are currently further exploring the efficiency of the pricing schemes with the proposed approach, and welfare distribution among heterogeneous groups, e.g. different trip purposes and users with different value-of-time. These investigations are feasible by utilizing the agent-based model. We will later test the efficiency of the developed pricing schemes for cities where network structure and traffic patterns are more complicated. For such cases, identifying optimal pricing strategies can be challenging as different pricings might be applied for different parts of a city. Furthermore, combing the promotion of bus usage with pricing deserves research attention. It has been shown that this strategy, if applied properly and accepted by users in real-life, can significantly increase mobility and multimodality.

Reference

Aboudolas, K. and Geroliminis, N., 2013. Perimeter and boundary control in multi-reservoir heterogeneous networks. Transportation Research Part B 55: 265-281.

Ampountolas, K., Zheng, N. and Geroliminis, N., 2014(a). Perimeter flow control of bi-modal urban road networks: A robust feedback control approach. In proceedings of the 2014 European Control Conference: 2569-2574. Strasburg, France. 24-27 June, 2014.

Ampountolas, K., Zheng, N. and Geroliminis, N., 2014(b). Robust control of bi-modal multi-region urban networks: An LMI optimization approach. In proceedings of the 16th IEEE Conference on Intelligent Transport System: 489-494. Qingdao, China. 09-11 October, 2014.

Anderson, S. and de Palma, A., 2004. The economics of pricing parking. Journal of Urban Economics 55:1–20.

Anderson, S. and de Palma, A., 2007. Parking in the city. Regional Science 86(4): 621-632.

Forsgerau, M. and de Palma, A., 2013. The dynamics of urban traffic congestion and the price of parking. Journal of Public Economics 105: 106-115.

Arnott, R. and Rowse, J., 1999. Modeling parking. Journal of Urban Economics 45: 97–124.

Arnott, R., 2006. Spatial competition between parking garages and downtown parking policy. Transport Policy 13(6): 458-469.

Arnott, R., 2007. Congestion tolling with agglomeration externalities. Journal of Urban Economics 62(2): 187-203.

Arnott, R. and Rorwse, J., 2009. Downtown parking in auto city. Regional Science and Urban Economics 39(1): 1-14.

Arnott, R. and Inci, E., 2010. The stability of downtown parking and traffic congestion. Journal of Urban Economics 68(3): 260-276.

Arnott, R., Inci, E. and Rowse, J., 2013. Downtown curbside parking capacity. CESifo Working Paper Series No. 4085.

Axhausen, K., Hess, S., Koenig, A., Bates, J., Bierlaire, M. and Abay, G., 2007. State-of-the-Art estimates of Swiss value of travel time savings. Paper presented at the 86th Annual Meeting of Transportation Research Board, Washington D.C., USA, 2007.

Black, J., Lim, P., and Kim, G., 1992. A traffc model for the optimal allocation of arterial road space: A case study of Seoul's first experimental bus lane. Transportation Planning and Technology 16(3): 195-207.

Buisson, C. and Ladier, C., 2009. Exploring the impact of homogeneity of traffic measurements on the existence of Macroscopic Fundamental Diagrams. Transportation Research Record 2124, 127-136.

Bureau of Public Roads, 1964. Traffic Assignment Manual. U.S. Department of Commerce, Urban Planning Division, Washington D.C., USA.

Cameron, I., Kenworthy, J. and Lyons, T., 2005. Understanding and predicting private motorized urban mobility. Transportation Research Part D 8(4): 267-283.

Cao, J., Menendez, M. and Vasileios, N., 2014. The effect of on-street parking on traffic throughput at nearby intersections. Paper presented at the 93rd Annual Meeting of Transportation Research Board, Washington D.C., USA. January 12-16, 2014.

Chakirov, A. and Fourie, P., 2014. Enriched Sioux Falls scenario with dynamic and disaggregate demand. Working paper. Future Cities Laboratory, Singapore ETH Centre, Singapore.

Chen, C., Varaiya, P. and Kwon, J., 2005. An empirical assessment of traffic operations. In proceedings of the 16th International Symposium on Transportation and Traffic Theory (ISTTT): 1-35. Maryland, USA. 19-21 July, 2005.

Currie, G., Sarvi, M., Young, W., 2004. A new methodology for allocating road space for public transport priority. In: Urban Transport X. Urban Transport and the Environment in the 21st Century: 375-388, Dresden, Germany.

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

Daganzo, C., 2007. Urban gridlock: Macroscopic modeling and mitigation approaches. Transportation Research Part B 41(1): 49-62.

Daganzo, C. and Geroliminis, N., 2008. An analytical approximation for the macroscopic fundamental diagram of urban traffic. Transportation Research Part B 42(9): 771-781.

Daganzo, C. and Cassidy, M., 2008. Effects of high occupancy vehicle lanes on freeway congestion. Transportation Research Part B 42(10): 861–872.

Daganzo, C., Gayah, V. and Gonzales, E., 2011. Macroscopic relations of urban traffic variables: Bifurcations, multivaluedness and instability. Transportation Research Part B 45(1): 278-288.

Dahlgren, J., 1998. High occupancy vehicle lanes: Not always more effective than general purpose lanes. Transportation Research Part A 32(2): 99–114.

De Berg, M., Cheong, O., van Kreveld, M. and Overmars, M. (eds.), 2008. Delaunay Triangulations. In: Computational Geometry: Algorithms and Applications. 3rd Edition. Springer-Verlag.

Edie, L., 1963. Discussion of traffic stream measurements and definitions. In proceedings of the 2nd International Symposium on the Theory of Traffic Flow (ISTTT): 139-154. Paris, France. 25-27 July, 1963.

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

Eliasson, J. and Mattsson, L., 2006. Equity effects of congestion pricing: Quantitative methodology and a case study for Stockholm. Transportation Research Part A 40(7): 602-620.

Gallo, M., D’Acierno, L. and Montella, B., 2011. A multiplayer model to simulate cruising for parking in urban areas. Transport Policy 18(5): 735-744.

Gayah, V. and Daganzo, C., 2011. Clockwise hysteresis loops in the Macroscopic Fundamental Diagram: An effect of network instability. Transportation Research Part B 45 (4): 643–655.

Geroliminis, N. and Daganzo, C., 2007. Macroscopic modeling of traffic in cities. Paper presented at the 86th Annual Meeting of Transportation Research Board. Washington D.C., USA.

Geroliminis, N. and Levinson, D., 2009. Cordon pricing consistent with the physics of overcrowding. In: Transportation and Traffic Theory: 219-240.

Geroliminis, N. and Sun, J., 2011. Properties of a well-defined macroscopic fundamental diagram for urban traffic. Transportation Research Part B 45(3): 605-617.

Geroliminis, N. and Boyaci, B., 2012. The effect of variability of urban systems characteristics in the network capacity. Transportation Research Part B 46(10): 1607-1623.

Geroliminis, N., 2014. Dynamics of peak hour and effect of parking for congested cities. Under review.

Godfrey, J., 1969. The mechanism of a road network. Traffic Engineering and Control 11(7): 323-327.

Gonzales E., Geroliminis N., Cassidy M. and Daganzo C., 2010. On the allocation of city space to multiple transport modes. Transportation Planning and Technology 33(8): 643-656.

Gonzales, E., Chavis, C., Li, Y. and Daganzo, C., 2011. Multimodal transport in Nairobi, Kenya: Insights and recommendations with a macroscopic evidence-based model. Paper presented at the 90th Annual Meeting of Transportation Research Board. Washington D.C., USA.

Gonzales, E. and Daganzo, C., 2012. Morning commute with competing modes and distributed demand: User equilibrium, system optimum, and pricing. Transportation Research Part B 46(10): 1519–1534.

Gonzales, E. and Daganzo, C., 2013. The evening commute with cars and transit: Duality results and user equilibrium for the combined morning and evening peaks. Transportation Research Part B 57: 286-299.

Guo, H. and Gao, Z., 2012. Modelling travel time under the influence of on-street parking. Journal of Transportation Engineering 138(2): 229-235.

Haddad, J. and Geroliminis, N., 2012. On the stability of traffic perimeter control in two-region urban cities. Transportation Research Part B 46(9): 1159–1176.

Hadded, J., Ramezani, R. and Geroliminis, N., 2013. Cooperative traffic control of a mixed network with two urban regions and a freeway. Transportation Research Part B 54: 17–36

Helbing, D., 2001. Traffic and related self-driven many-particle systems. Reviews of Modern Physics 73(4): 1067-1141.

Helbing D. and Treiber, M., 1998. Gas-kinetic-based traffic model explaining observed hysteretic phase transition. Physical Review Letters 81(14): 3042-3045.

Herman, R. and Prigogine, I., 1979. A two-fluid approach to town traffic. Science 204(4389): 148–151.

Horni, A., Montini, L., Waraich, R. and Axhausen, K., 2013. An agent-based cellular automaton cruising-for-parking simulation. Transportation Letters 5(4): 167-75.

Ji, Y., Daamen, W., Hoogendoorn, S., Hoogendoorn-Lanser, S. and Qian, X., 2010. Investigating the shape of the Macroscopic Fundamental Diagram using simulation data. Transportation Research Record 2161: 40-48.

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

Kenworthy, J. and Laube, F., 2001. Millenium cities database for sustainable transport: Analysis and recommendations. In: International Union of Public Transport (UITP). Brussels, Belgium.

Keyvan-Ekbatani, M., Kouvelas, A., Papamichail, I. and Papageorgiou, M., 2012. Exploiting the fundamental diagram of urban network for feedback-based gating. Transportation Research Part B 46(10): 1393-1403.

Knoop, V., Hoogendoorn, S. and Van Lint, H., 2012. Routing strategies based on the Macroscopic Fundamental Diagram. Paper presented at the 91st Annual Meeting of Transportation Research Board. Washington D.C., USA.

Kosmatopoulos, E. and Papageorgiou, M., 2003. Stability Analysis of the Freeway Ramp Metering Control Strategy ALINEA. Paper presented at the 11th Mediterranean Conference on Control & Automation, Rhodes, Greece.

Lebacque, J., 1996. The Godunov scheme and what it means for first order traffic flow models. In proceedings of the 13th International Symposium on Transportation and Traffic Theory (ISTTT): 647-677. Lyon, France. 24-26 July, 1996.

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

Leclercq, L., Laval, J. and Chevallier, E., 2007. The Lagrangian coordinate system and what it means for first order traffic flow models. In proceedings of the 17th International Symposium on Transportation and Traffic Theory (ISTTT): 735-753. London, UK. 23-25 July, 2007.

Levinson, D., 2010. Equity effects of road pricing: A review. Transport Review 30(1): 33-57.

Li, J., Song, M. and Zhang, W., 2010. Planning for bus rapid transit in single dedicated bus lane. Transportation Research Record 2111: 76-82.

Li, Z., Huang, H., Lam, W. and Wong, S., 2007. A model for evaluation of transport policies in multimodal networks with road and parking capacity constraints. Journal of Mathematic Model and Algorithm 6: 239-257.

Liu, T., Huang, H., Yang, H. and Zhang, X., 2009. Continuum modeling of park-and-ride services in a linear monocentric city with deterministic mode choice. Transportation Research Part B 43: 692-707.

Little, J., 1961. A proof of the queueing formula $L=\lambda W$. Operations Research 9(3): 383-387.

Liu, Z., Meng, Q. and Wang, S., 2013. Speed-based toll design for cordon-based congestion pricing scheme. Transportation Research Part C 31: 83-98.

Lu, C., Mahmassani, H. and Zhou, X., 2008. A bicriterion dynamic user equilibrium traffic assignment model and solution algorithm for evaluating road pricing strategies. Transportation Research Part C 16(4): 371-389.

Mahmassani, H., Williams, J. and Herman, R., 1987. Performance of urban traffic networks. In proceedings of the 10th International Symposium on Transportation and Traffic Theory (ISTTT): 1-20. Cambridge, USA. 9-10 July, 1987.

Mahmassani, H., Saberi, M. and Zockaie A., 2013. Urban network gridlock: Theory, characteristics, and dynamics. Transportation Research Part C 36: 480-497.

Martens, K., Benenson, I. and Levy, N., 2010. The dilemma of on-street parking policy: Exploring cruising for parking using an agent-based model. Geo Journal Library 99: 121-138.

Maruyama, T. and Sumalee, A., 2007. Efficiency and equity comparison of cordon- and cordon-based road pricing schemes using a trip-chain equilibrium model. Transportation Research Part A 41(7): 655-671.

Mazloumian, A., Geroliminis, N. and Helbing, D., 2010. The spatial variability of vehicle densities as determinant of urban network capacity. Philosophical Transactions of Royal Society A 368(1928): 4627-4648.

Meister, K., Balmer, M., Ciari, F., Horni, A., Rieser, M., Waraich, R. and Axhausen, K., 2010. Large-scale agent-based travel demand optimization applied to Switzerland, including mode choice. Paper presented at the 12th World Conference on Transportation Research. Lisbon, Portugal.

Mesbah, M., Sarvi, M., Ouveyse, I. and Currie, G., 2011. Optimization of transit priority in the transportation network using a decomposition methodology. Transportation Research Part C 19(2), 363-373.

Mohit, S., 2010. Activity-based Travel Demand Modelling including Freight and Cross-border Traffic with Transit Simulation. Arbeitsberichte Verkehrs- und Raumplanung (Project Report Transport and Land Use Planning): 654-683. IVT, ETH Zürich, Zürich.

Nagel K. and Schreckenberg, M., 1992. A cellular automaton model for freeway traffic. J.Phys. I France 2(12): 2221-2229.

Nagel K. and Flötteröd, G., 2009. Agent-based traffic assignment: going from trips to behavioral travelers, Travel Behaviour Research in an Evolving World: 261-293. Emerald Group Publishing, Edition 2012 by R.M. Pendyala and C. Bhat.

Nocedal, J. and Wright, J., 2006. Numerical Optimization, Second Edition. Springer Series in Operations Research, Springer Verlag.

Ogata, M., 2001. In: Modern Control Engineering (4th Edition): 24-27.

Ortigosa, J., Menendez, M. and Tapia, H., 2013. Study on the number and location of measurement points for an MFD perimeter control scheme: a case study of Zurich. EURO Journal on Transportation and Logistics: 1-22.

Papageorgiou, M. and Kotsialos, A., 2002. Freeway ramp metering: an overview. IEEE Transactions on Intelligent Transportation Systems 3(4): 228-239.

Pickrell, D., 1985. Estimates of rail transit construction costs. Transportation Research Record 1006: 54-60.

Pushkarev, B. and Zupan, J., 1977. Public Transportation and Land Use Policy. Indiana University Press.

Qian, Z., Xiao, F. and Zhang, H., 2013. Managing morning commute traffic with parking. Transportation Research Part B 46(7): 894–916.

Qian, Z. and Zhang, H., 2013. The morning commute problem with heterogeneous travelers: the case of continuously distributed parameters. Transportmetrica A 9(2): 178-203.

Radwan, A. and Benevelli, D., 1983. Bus priority strategy: Justification and environmental aspects. Journal of Transportation Engineering 109(1): 88-106.

Saberi, M. and Mahmassani, H., 2012. Exploring properties of network-wide flow-density relations in a freeway network. Paper presented in the 91st Annual Meeting of Transportation Research Board. Washington D.C., USA.

Shi, J. and Malik, J., 2000. Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (8): 888–905.

Shoup, D., 2005. Cruising for parking. Transport Policy 13(6): 479-486.

Small, K., 2004. Urban Transportation. Concise Encyclopedia of Economics, 2nd edition. (Indianapolis: Liberty Fund)

Sparks, G. and May, A., 1971. A mathematical model for evaluating priority lane operations on freeways. Transportation Research Record 363: 27-42.

Van den Berg, V. and Verhoef, E., 2011. Congestion tolling in the bottleneck model with heterogeneous values of time. Transportation Research Part B 45: 60-78.

Van Ommeren, J., Wentink, D. and Rietveld, P., 2012. Empirical evidence on cruising for parking. Transportation Research Part A 46(1): 123-130.

Vickrey, W., 1969. Congestion theory and transport investment. American Economic Review 59(2): 251-260.

Verhoef, E., 2002. Second-best congestion pricing in general networks: heuristic algorithms for finding second-best optimal toll levels and toll points. Transportation Research Part B 36(8): 707-729.

Vuchic, V., 2007. Urban transit systems and technology. Published by John Wiley & Sons. Inc., New Jersey, USA.

Wu, D., Yin, Y. and Lawphongpanich, S., 2011. Pareto-improving congestion pricing on multimodal transportation networks. European Journal of Operational Research 210(3): 660-669.

Yang, H. and Huang, H., 2005. Mathematical and Economic Theory of Road Pricing. Elsevier Science Inc., New York, USA.

Yin, Y., 2000. Multi-objective bi-level optimization for transportation planning and management problems. Journal of Advanced Transportation 36(1): 93-105.

Yildirimoglu, M. and Geroliminis, N., 2014. Approximating dynamic equilibrium conditions with Macroscopic Fundamental Diagrams. Transportation Research Part B 70: 186-200.

Yousif, S. and Purnawan, P., 2004. Traffic operations at on-street parking facilities. Proceedings of the Institution of Civil Engineers-Transport 157(3): 189-194.

Zhang, L. and Levinson, D., 2005. Balancing Efficiency and Equity of Ramp Meters. ACSE Journal of Transportation Engineering 131(6): 447-481.

Zhang, L., Levinson, D. and Zhu, J., 2008. Agent-based model of price competition, capacity choice and product differentiation on congested networks. Journal of Transport Economics and Policy 42(3): 435-461.

Zhou, X., Mahmassani, H. and Zhang, K., 2008. Dynamic micro-assignment modelling approach for integrated multimodal urban corridor management. Transportation Research Part C 16(2): 167-186.

Curriculum Vitae

Nan Zheng

Email: nanzheng1985@gmail.com

Birth Date: 26/03/1985

Citizenship: Chinese

EDUCATION

- 2010-2014, PhD, École Polytechnique Fédéral Lausanne, Switzerland.
- 2007-2009, MSc. Delft University of Technology, the Netherlands.
- 2003-2007, BSc. Southeast University, Nanjing, China.

PEER-REVIEWED JOURNAL PUBLICATIONS

- A dynamic cordon pricing scheme combining the Macroscopic Fundamental Diagram and an agent-based traffic model (2012). N.Zheng, R.Waraich, K.Axhausen and N.Geroliminis. Transportation Research Part A 46(8): 1291–1303
- On the distribution of road space for urban multimodal congested networks (2013). N.Zheng and N.Geroliminis. Transportation Research Part B 57: 326-341.
- A three-dimensional macroscopic fundamental diagram for mixed bi-modal urban networks (2014). N.Geroliminis, N.Zheng* and K.Aboudolas. Transportation Research Part C 57: 168-181.
- A time-dependent area-based pricing scheme for multimodal urban networks with user adaptation: An agent-based approach (2015). N.Zheng, G.Rérat and N.Geroliminis, to be submitted in the IEEE Transaction on ITS.
- Perimeter flow control for bi-modal urban road networks (2014), K.Aboudolas, N.Zheng, and N. Geroliminis, to be submitted to Transportation Research Part B.
- An aggregated dynamic network model for multimodal systems with cruising for parking and pricing (2014), N.Zheng and N.Geroliminis, to be submitted.

PROJECT PARTICIPATION

- 2010-2014, leading researcher, Movement conflicts in multi-modal urban traffic systems: Modeling congestion and developing more sustainable cities, Swiss National Science Foundation project 200021-132501

- 2012-2014, researcher, NetCap intermodal capacity of links, segments and networks, Swiss federal institute ETH-EPFL joint project SVI 2004/032
- 2014-2015, leading researcher, Designing efficient multi-modal traffic management schemes with an for the city of Geneva, project for Canton de Geneva

AWARDS

- 2014, Chinese government award for outstanding doctoral students abroad
- 2012, Golden Award for Doctoral Research in Transportation, awarded at the Young European Arena of Research 2012 competition in Athens, Greece, by the European Commission and five affiliated councils

CONFERENCE PRESENTATIONS

- Variable Analysis for Freeway Work Zone Capacity Prediction, Nan Zheng et al., the 13th ITSC, Funchal, Portugal, 2010.
- A Comparison of Freeway Work Zone Capacity Prediction Models, Nan Zheng et al., the 6th International Symposium on Highway Capacity (ISHC), Stockholm, Sweden, 2011.
- A Dynamic Cordon Pricing Scheme combining a Macroscopic and an Agent-based Model, Nan Zheng et al., the 91th TRB, Washington, the United States, 2012.
- A Dynamic Cordon Pricing Model – A Robust and Effective Road Pricing Model, Nan Zheng et al., the European Conference on Transport Research (Transport Research Arena), Athens, Greece, 2012.
- Re-distribution of urban road space for multimodal congested networks with MFD representation, Nan Zheng and Nikolas Geroliminis, the 92nd TRB, Washington D.C., US, 2013.
- Investigation of a city-scale three-dimensional Macroscopic Fundamental Diagram for bi-modal urban traffic, Nan Zheng et al., the 16th ITSC, Hague, The Netherland, 2013.
- A city-scale macroscopic fundamental diagram for mixed bi-modal urban traffic, Nan Zheng et al., the 93rd TRB, Washington D.C., US, 2014.
- Perimeter flow control for bi-modal urban road networks, Ampountolas et al., the 93rd TRB, Washington D.C., US, 2014.
- Perimeter flow control in bi-modal urban road networks: A robust feedback control approach. Ampountolas et al., the European Control Conference (ECC), Strasburg, France, 2014.

- An aggregated dynamic model for multimodal networks with limited parking. Nan Zheng and Nikolas Geroliminis, Symposium of the European Association for Research in Transportation (hEART), Leeds, UK, 2014.
- A time-dependent area-based pricing scheme for multimodal urban networks with user adaptation: An agent-based approach. Nan Zheng, Guillaume Rérat and Nikolas Geroliminis. The 17th ITSC. Qingdao, China, 2014.

ACADEMIC ACTIVITIES

- Journal Reviewer
  - Transportation Research Part A
  - Transportation Research Part B
  - Transportation Research Part C
  - Journal of Advanced Transportation
- Conference Reviewer
  - Annual meeting of Transportation Research Board (TRB)
  - IEEE Conference on Intelligent Transport System (ITSC)
  - Conference of the European Association for Research in Transportation (hEART)
- Teaching Assistant
  - Linear algebra (Bachelor program)
  - Traffic Engineering (Bachelor program)
  - Fundamentals of traffic operations and control (Master program)
- Supervision of semester/graduation projects:
  - Bi-modal traffic modeling and simulation, Bachelor program, student Lea Kaufmann.
  - Bus bunching problem and passenger flow characteristics: A case study of Fribourg, Bachelor program, students Hadi Karam, Benjamin Schaer and Levy Sharabi.
  - Performance of public transport systems: A case study of Fribourg, Bachelor program, students Selma Lehmann and Nathalie Dufour.
  - Modeling and optimization of bike-sharing system in Barcelona, Bachelor program, student Anna Llopis.
  - Traffic operations on congested freeways, Bachelor program, 1 student Guillaume Kayser.
  - Investigation of traffic performance and multimodality for the city of Geneva with field detector and bus GPS data, Master program, students Nathalie Dufour and Rayane Zahreddine.

- Pricing and population heterogeneity, Master program, student Stefan Naef.
- An agent-based approach for a time-dependent area-based congestion pricing scheme, Swiss National Civil Service program, civil servant Guillaume Rérat.

SOCIAL ACTIVITIES

- 2010-2013, President, Board of management for Chinese Students and Scholars Association Lausanne, Switzerland; since, senior consultant.

SKILLS

- Microsoft Office (Word, Visio, PowerPoint, Excel, Access, Outlook)
- Data and programming tool: MATLAB, SPSS
- Transportation: AIMSUN, VISSIM, CORSIM, TRANSCAD, OMNITRANS
- Chinese: mother tongue; English: proficient in reading, writing, listening and speaking; French: basic in listening and speaking, intermediate in writing and reading