# Traffic modeling, forecasting and assignment in large-scale networks

THÈSE Nº 6853 (2015)

PAR

## Mehmet YILDIRIMOĞLU

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2015

# Acknowledgements

First and foremost, I am extremely grateful to my supervisor Nikolas Geroliminis for his invaluable guidance, everlasting enthusiasm, continuous encouragement and full support. I am particulary thankful to him for trusting me and offering me an opportunity to pursue a Ph.D. at LUTS. He is not only a great scientist, but also a good friend. I feel extremely lucky to know him and work with him.

I would like to thank my committee members: Michel Bierlaire, Ludovic Leclercq, Ben Heydecker and Eugen Brühwiler for taking time to review this dissertation, for interesting discussion and valuable comments. Their comments helped me significantly improve the manuscript.

It was a great pleasure to be a member of LUTS. I have enjoyed the good mood, the friendly and motivating environment at the working place during my Ph.D. Special thanks go to our lovely secretary Christine Droguet who runs all the important practical administrative duties in our laboratory. I would like to thank previous and current LUTS colleagues: Anastasios Kouvelas, Burak Boyaci, Claudia Bongiovanni, Isik Sirmatel, Jack Haddad, Konstantinos Ampountolas, Leonardo Bellocchi, Mikhail Murashkin, Mohammadreza Saeedmanesh, Mohsen Ramezani, Nan Zheng, Raphael Lamotte, Wei Liu and Yuxuan Ji. It has been a great pleasure to know you and work with you. I would like to particularly thank Raphael for helping me translate the abstract to french and Wei for delivering the manuscript when I was away. I would like to express my sincere gratitude to Mohsen Ramezani and Ypatia Limniati, with whom I had the opportunity to work together on some parts of this dissertation.

Many thanks to my friends in Lausanne for making my life enjoyable and exciting. My special thanks to Baris Karpuz, Burak Boyaci, Nihal Ozen, Onur Ozen and many others with whom I had great memories. I hope they always stay in my life.

My deepest gratitude goes to my parents, Esma and Resat, and my sister, Bilge, for their unconditional love and continuous support not only during my Ph.D. but my whole life. Thank you for being there any time I need you.

Most importantly, I would like to express my heartfelt gratitude to my wife, Çağla, for her understanding, patience and support. In fact, the story of my Ph.D. is the story of an amazing journey through life we have embarked on together. I have met her in Lausanne few weeks after I started my Ph.D., and here we are... Thank for for all that we had, and we will have together.

*Lausanne, 20 November 2015* M. Y.

# Abstract

TRAFFIC congestion has become a disease of modern life in big metropolitan areas. Today's society suffers from wasted time, wasted fuel, air pollution, stressed drivers, etc. due to increasing traffic congestion every day. A 'quick straightforward' response to the congestion problem is to build new infrastructure and increase capacity. Nevertheless, the expansion of transportation system often results in urban sprawl and induced demand. A sustainable solution to alleviate congestion could be achieved through the development of Intelligent Transportation Systems (ITS) technologies with novel modeling, forecasting and routing methods that predict traffic states in real-time and implement adaptive guidance schemes.

Today, the development and evaluation of traffic management strategies heavily relies on microscopic traffic simulation models. In case detailed input (i.e. od matrix, signal timings, etc.) is extracted and incorporated in these simulators, they can provide valuable traffic state predictions. However, as this type of information is almost never available at the large-scale and traffic represents chaotic behavior in saturated networks, microscopic simulation models remain intractable and unstable. An alternative is a recently discovered network traffic model; macroscopic fundamental diagram (MFD). Nevertheless, large-scale traffic management strategies remain a big challenge partly due to unpredictability of choices of travelers (e.g. route, departure time and mode choice). Part I of the thesis is an attempt to fill this gap. Chapter 2, 3 and 4 elaborate new aspects of large-scale traffic modeling, and integrate route choice behavior into the modeling. Chapter 2 proposes a dynamic traffic assignment (DTA) model to establish equilibrium conditions in multi-region urban networks where the modeling is done through MFD dynamics. The method handles the stochastic components of the aggregated model through a sampling approach. In addition, the assignment model enables us to consider the response of drivers to changing traffic conditions in an aggregated modeling framework. Chapter 3 extends the DTA model presented in Chapter 2 to a route guidance system, where drivers are given a sequence of subregions to follow. Two aggregated models, region- and subregion-based models, are introduced to develop the guidance scheme and to test its effect, respectively. Notably, these two models define state variables and route choice behavior at relatively different granularity levels. Therefore, the challenge here is to translate certain variables across the traffic models without a loss of significance and assure certain degree of consistency. In addition, Chapter 3 establishes equilibrium conditions in the subregion-based model and investigates the implications of equilibrium on the resulting region MFDs. Chapter 4 extracts and reconstructs aggregated route choice patterns through

an extensive GPS data set from taxis in a mega city. Observed GPS trajectories are first grouped together to provide a physical evidence for consistent route patterns. Second, in order to investigate the consistency of equilibrium assumptions considered in Chapter 2, observed trajectories are replaced with shortest path trajectories, and aggregated route choice patterns are reconstructed.

Part II introduces novel travel time prediction and variability models. Travel time is a crucial performance measure in assessing the efficiency of transportation systems, and it provides a common index for both practitioners and travelers. Chapter 5 develops a travel time prediction model that jointly exploits traffic flow fundamentals and advanced data mining techniques. The prediction method detects the congestion patterns through the identification of active bottlenecks, and clusters the days with similar traffic patterns. This approach basically allows the model to train its predictions with relevant historical data sets. In addition, the model does not aim for the prediction of instantaneous travel times in future time periods. Instead, it considers changing traffic conditions along the trip duration, and predicts the travel time to be experienced by the vehicles departing at the current interval. The method is applicable in oversaturated conditions and consistent with physics of traffic flow. Nevertheless, travelers not only consider travel time on average, but also value its variation. Day-to-day travel time variability, addressing the travel time variations of vehicles crossing the same route at the same period of time on different days, reveals interesting patterns. Departure time periods with similar mean travel times in the onset and offset of congestion exhibit quite different variance values. This phenomenon causes counter-clockwise hysteresis loops on the mean-variance curves. Chapter 6 investigates the empirical implications of hysteresis shape within the context of day-to-day travel time variability. The method approximates the evolution of travel time within the day with a piecewise linear function, which enables us to define the same problem with a number of key variables (e.g. start time of congestion, maximum travel time, etc.). A Monte Carlo simulation is later built in order to identify the partial effects of these variables on the existence of hysteresis loops.

Key words: *macroscopic fundamental diagram, dynamic traffic assignment, dynamic user equilibrium, dynamic stochastic user equilibrium, route guidance, system optimum, traffic hysteresis, probe vehicles, travel time prediction, bottleneck identification, traffic flow, data mining, clustering, stochastic congestion maps, travel time variability, Monte Carlo simulation*

# Résumé

La congestion routière est devenue une maladie de la vie moderne dans les grandes régions métropolitaines. Les symptômes incluent la perte de temps, le gaspillage de carburant, la pollution de l'air, les conducteurs stressés, etc. Une réponse 'rapide et simple' à ces problèmes est de construire de nouvelles infrastructures et d'augmenter la capacité. Néanmoins, l'expansion du système de transport se traduit souvent par un étalement urbain et une demande induite. Le développement récent des systèmes de transport intelligents (STI) ouvre la voie pour une alternative plus durable, faisant appel aux nouvelles méthodes de modélisation, de prévision et de guidage routier qui permettent de prédire les états de la circulation en temps réel et de mettre en œuvre des services de guidage adaptatifs.

Aujourd'hui, le développement et l'évaluation des stratégies de gestion du trafic reposent largement sur des modèles microscopiques de trafic. Ces simulations requièrent des données extrêmement détaillées (par exemple la demande entre les origines et destinations, les paramètres des signaux, etc.) et peuvent fournir des prévisions du trafic routier précieuses. Cependant, ces données sont très rarement disponibles à grande échelle et le comportement chaotique du trafic dans les réseaux saturés rend ces modèles microscopiques de trafic instables et difficiles à valider. Une alternative est un modèle de réseau découvert récemment ; le diagramme macroscopique fondamental (MFD). Néanmoins, les stratégies de gestion du trafic restent un défi à grande échelle du fait du caractère imprévisible du comportement des voyageurs (p. ex choix d'itinéraire et de l'heure de départ). La Partie I de la thèse tente de combler cette lacune. Les Chapitres 2, 3 et 4 intègrent de nouveaux aspects à la modélisation du trafic à grande échelle, ainsi que le choix d'itinéraire. Le Chapitre 2 propose un modèle d'affectation dynamique du trafic (DTA) pour établir des conditions d'équilibre dans les réseaux urbains composés de multiples régions où les dynamiques sont modélisées à l'aide du MFD. La méthode gère les composantes stochastiques qui existent dans le modèle agrégé grâce à une approche d'échantillonnage. En outre, le modèle d'affectation nous permet d'examiner la réponse des conducteurs à l'évolution de la circulation routière dans un cadre de modélisation agrégée. Le Chapitre 3 étend le modèle de DTA présenté dans le chapitre 2 à un modèle de guidage routier, où les conducteurs reçoivent une séquence de sous-régions à suivre. Deux modèles agrégés, incluant des régions et sous-régions, sont introduits pour respectivement développer le système de guidage et en tester les effets. En particulier, ces deux modèles utilisent des variables d'état et de choix d'itinéraire qui ont des niveaux de granularité différents. Par conséquent, le défi est de traduire certaines variables d'un modèle à l'autre sans perdre d'information et conservant des modèles cohérents. Par ailleurs, le Chapitre 3

établit les conditions d'équilibre dans le modèle utilisant des sous-régions et étudie les effets de l'équilibre sur les MFDs des régions. Le Chapitre 4 extrait et reconstruit des motifs agrégés de choix d'itinéraire grâce aux nombreuses données GPS de la part d'un ensemble de taxis dans une mégapole. Les trajectoires GPS observées sont d'abord regroupées pour fournir une preuve physique de l'existence de motifs clairs de choix d'itinéraire. Deuxièmement, pour étudier la cohérence des hypothèses d'équilibre considérées dans le Chapitre 2, les trajectoires observées sont remplacés par les plus courts chemins, et les motifs agrégées sont reconstruits. La Partie II développe de nouveaux modèles de prévision et de la variabilité pour le temps de parcours. Le temps de parcours est une mesure cruciale de performance pour l'évaluation de l'efficacité des systèmes de transport, et il fournit un indice commun pour les praticiens et les voyageurs. Le Chapitre 5 développe un modèle de prévision pour le temps de parcours. Ce modèle exploite conjointement les fondamentaux de la théorie des flux routiers et des techniques avancées d'exploration de données (data mining). La méthode de prévision détecte les schémas de congestion grâce à l'identification des goulets actifs d'étranglement, et regroupe les jours qui ont des schémas similaires de trafic. Cette approche permet essentiellement d'entraîner le modèle avec des données historiques pertinentes. En outre, le modèle ne vise pas à prédire des temps instantanés de parcours sur des périodes futures. Au lieu de cela, il considère l'évolution des conditions routières au long de la durée du parcours, et prédit le temps de parcours qui va être expérimenté par les véhicules. La méthode est applicable dans des conditions sursaturées et compatible avec la physique des flux de trafic. Néanmoins, les voyageurs ne considèrent pas seulement le temps de parcours en moyenne, mais apprécient aussi sa variation. L'étude des variations de temps de parcours expérimentés par les véhicules qui traversent la même route à la même période mais à des jours différents révèle des schémas intéressants. Des heures de départ ayant le même temps de parcours moyen peuvent être associées à des variabilités du temps de parcours très différentes. Ce phénomène provoque des boucles d'hystérésis dans le sens antihoraire sur les courbes moyenne-variance. Le Chapitre 6 étudie les implications empiriques de forme d'hystérésis dans le cadre de la variabilité du temps de parcours. La méthode se rapproche de l'évolution du temps de parcours pendant la journée avec une fonction linéaire par morceaux, ce qui nous permet de définir le même problème avec un certain nombre de variables clés (par exemple l'heure de début d'embouteillage, le temps de parcours maximal, etc.). Une simulation de Monte Carlo est construite plus tard afin d'identifier les effets partiels de ces variables sur l'existence des boucles d'hystérésis.

Mots clefs : *diagramme macroscopique fondamental, affectation dynamique du trafic, l'équilibre dynamique de l'utilisateur, l'équilibre dynamique stochastique de l'utilisateur, le guidage routier, le système optimal, l'hystérésis de la circulation, les véhicules de la sonde, la prévision du temps de parcours, l'identification des goulets d'étranglement, le flux de trafic, data mining, la partitionnement de données, cartes stochastiques de congestion, la variabilité des temps de parcours, la simulation de Monte Carlo*

# Contents

# Contents

# List of Figures

# List of Tables

# 1 Introduction

TRAFFIC congestion increasing everyday has a number of negative outcomes; wasted time of travelers, wasted fuel increasing air pollution, stressed drivers and increased chance of accidents which by itself imposes a significant financial burden. Recent studies show that vehicle kilometers traveled (VKT) increases proportionally with the available roadway lane kilometers (Duranton and Turner, 2011). Hence, building new infrastructure is not a sustainable solution to alleviate traffic congestion. To this end, novel traffic monitoring and management techniques are needed to improve traffic mobility in large-scale transportation networks. This dissertation revolves around the incorporation of traffic assignment / routing techniques into parsimonious traffic models and the travel time prediction / variability models.

In this introductory chapter, Section 1.1 presents the thesis motivation, and briefly introduces the existing literature, while a detailed literature review is provided in each chapter. Section 1.2 elaborates the objectives and the contributions that this dissertation aims to achieve. Section 1.3 outlines the structure of the dissertation.

## 1.1 Motivation and background

Traffic congestion occurs basically when the traffic demand is greater than the roadway supply, and it is characterized by low speeds, long vehicle queues and high travel times. Along with its negative impacts (e.g. wasted time, fuel, etc.), the economical aspects of traffic congestion are of great importance for modern societies. For instance, the annual cost of congestion in Europe is about 1% of its gross domestic product (GDP), and it is estimated to increase by 50% to nearly €200 billion by 2050 (European Commission, 2011).

The potential countermeasures to alleviate traffic congestion can be divided into several groups. (i) Modification in road infrastructure, e.g. junction improvement, bus lanes, high occupancy vehicle (HOV) lanes, etc.; (ii) demand management, e.g. road pricing, parking restrictions, number plate restrictions, etc.; (iii) traffic control, e.g. ramp metering, coordinated

traffic signals, route guidance, traveler information systems, etc. For a detailed overview of traffic control strategies, please refer to Papageorgiou et al. (2003).

However, large-scale traffic control strategies remain a big challenge due to complexity and unpredictability of transportation networks. In addition, current traffic control strategies are considered less efficient in over-saturated traffic conditions, because the modeling of traffic networks underlying these strategies is quite complex and inaccurate. Given accurate inputs, detailed traffic models (e.g. microscopic simulation models) can provide substantial traffic predictions at the link level in large scale networks. However, congestion management based on these predictions can be quite inefficient due to (i) exhaustive number of required inputs (e.g. detailed OD tables, signal settings, etc.), (ii) inaccuracy of behavioral models (e.g. car-following, route choice models, etc.), (iii) chaotic behavior of traffic in over-saturated networks. Hence, efficient management of transportation systems requires parsimonious modeling of traffic, understanding of travelers' choice (e.g. route choice) and accurate forecasting of traffic state. In this respect, first part of this dissertation work addresses the choice of travelers in large-scale networks, and develops a traffic assignment model that can be further exploited in parsimonious network models. On the other hand, second part of the dissertation provides a deep understanding of expected travel costs and their variation over days.

Chapter 2, 3 and 4 elaborate large scale traffic modeling and assignment. Traffic flow modeling is mainly derived in analogy with a model for river flow (Lighthill and Whitham, 1955a,b). The key difference between fluid and vehicular flow is that vehicles or drivers all behave differently unlike the particles that are subject to certain physical laws. Papageorgiou (1998) claims that macroscopic traffic flow models may never reach the descriptive accuracy level of other domains in physics and engineering. Hence, the challenge of traffic flow researchers is to look for practical models that have sufficient descriptive power, where sufficiency depends on the application purpose. The efforts on mathematical description of traffic flow have resulted in a broad range of models that describe different aspects of traffic streams; considering the time-space behavior of individual vehicles depending on the vehicles in proximity (*microscopic models*), keeping track of individual vehicles without explicitly defining their time-space behavior (*mesoscopic models*), using the properties of collective traffic flow (*macroscopic models*). All the aforementioned models are suitable to define the link traffic flow, while the literature on network flow modeling is limited.

The literature on parsimonious network traffic models is quite recent, yet effective. A unimodal, low-scatter, and demand-insensitive relationship between network vehicle density and space-mean flow has been observed from empirical data in downtown Yokohama (Geroliminis and Daganzo, 2008). By spatially aggregating the highly scattered plots of flow vs. density from individual detectors, they identify a well defined relationship where scatter almost disappears; Macroscopic Fundamental Diagram (MFD). Figure 1.1(a) depicts the scattered fundamental diagram from two loop detectors in the Yokohama urban area, while Figure 1.1(b) presents the low scatter MFD for Yokohama region. MFD clearly identifies free flow and congested network states, and depicts a critical average occupancy (or accumulation) threshold where

the network flow is maximized. The idea of such relationship with an optimum accumulation belongs to Godfrey (1969), and two-fluid models of urban traffic introduced by Herman and Prigogine (1979) and Mahmassani et al. (1984) are other approaches that investigate macroscopic traffic relations in urban networks. Nevertheless, the empirical verification of the network-level traffic model with dynamic features and indifference to demand patterns is conducted much later by Geroliminis and Daganzo (2008). Figure 1.1(c) introduces the ratio of network production (i.e. the product of average flow and network length) to the outflow through the region boundary. The almost stationary relation between them represents an almost time-invariant value for the network average trip length. This feature is quite important as the average network flow can be easily observed with available sensors (e.g. loop detectors), while the outflow calculation requires full surveillance of region boundary and detection of ending trips. In this thesis, we demonstrate that heterogeneous urban networks might as well experience non-steady average trip length as a result of drivers' response to changing traffic conditions.



(a)

(b)

(c)
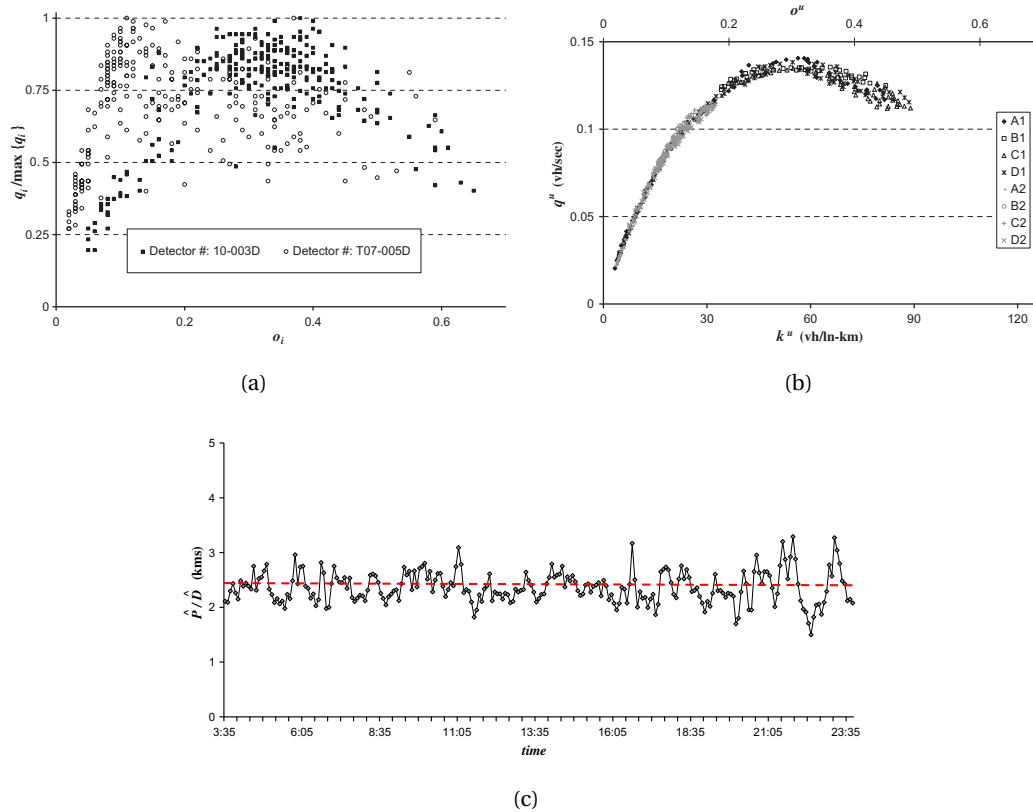
Figure 1.1: (a) Flow vs. occupancy for two detectors across a day, (b) average flow vs. average occupancy from all detectors across two days, (c) ratio of region production to outflow (Geroliminis and Daganzo, 2008)

Unfortunately, well-defined MFDs with low scatter are not a universal law. Heterogeneity in congestion distribution can affect the shape/scatter or even the existence of MFD (Buisson

and Ladier, 2009; Mazloumian et al., 2010). Mazloumian et al. (2010) define the variance of link density as a key variable and observe well-defined relations between average flow and average density when link density variance is approximately constant. Results from heterogeneity analysis are very critical, because they imply that MFD approach can be valid even in heterogeneously loaded cities if the network is partitioned into homogenous compact areas. Ji and Geroliminis (2012) develop a partitioning algorithm for large-scale heterogeneous networks, and produce subnetworks that are compact in shape and that have low link density variance. Figure 1.2 introduces an implementation of the partitioning algorithm on a microscopic simulation of San Francisco network. Figure 1.2(a) shows a snapshot of link densities (the darker the color is, the more congested the links are), Figure 1.2(b) illustrates the resulting 3 partitioned compact regions, and Figure 1.2(c) presents the MFDs of the aforementioned 3 regions. Note that the time each region reaches its critical accumulation is different, which implies that loading pattern of each subnetwork is different. Partitioning results can be exploited to develop large-scale traffic management strategies that act on resulting subnetworks. Such strategies e.g. perimeter control (Geroliminis et al., 2013; Haddad et al., 2013; Ramezani et al., 2015); gating (Keyvan-Ekbatani et al., 2012, 2015; Gayah et al., 2014; Haddad and Shraiber, 2014) that benefit from the parsimonious MFD model, provide promising results towards a new generation of smart hierarchical strategies. However, a significant part of aforementioned control strategies require prediction of traffic conditions. In a multi-region urban network where multiple regional path options exist for a particular OD pair, accurate prediction of traffic relies on the correct modeling of traveler behavior and the achievement of equilibrium conditions. This thesis develops a dynamic traffic assignment (DTA) model that could cope with the characteristics of MFD modeling and establish equilibrium state.



(a)                          (b)                          (c)
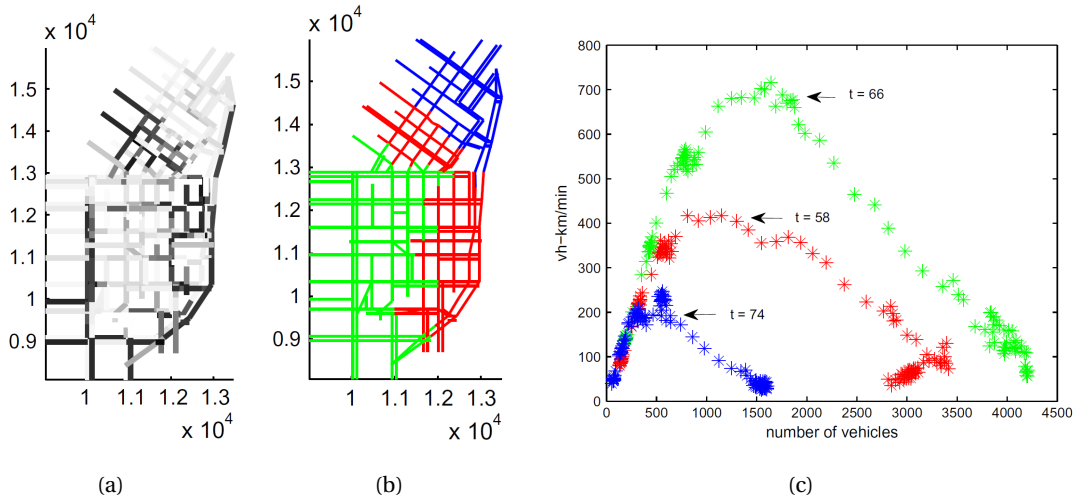
Figure 1.2: Partitioning results (a) a snapshot of link densities in San Francisco network, (b) partitioned network into 3 regions, (c) MFDs for 3 partitioned regions (Ji and Geroliminis, 2012).

An alternative large-scale traffic management strategy is a regional route guidance system, where a network is partitioned into homogeneous regions and drivers are given a regional path to follow (e.g. going through the city center or using the longer route at the periphery). Despite different OD configurations, chosen routes, travel times and driving behaviors, all travelers interact with each other and cause delays to the transportation system as a whole. Therefore, a route guidance system could force the travelers to cooperate with each other and succeed the minimization of total delay in the system. Modeling of the route choice behavior at an aggregated level (sequence of neighborhoods) unlike the link-based traffic assignment models (sequence of links) may overcome the computational difficulties and lead to a higher system performance in overall. Nevertheless, significant effort is needed to apply this concept in field implementations. Furthermore, the impact of driver adaptivity on the network performance, in particular MFDs, is not well explored. MFD estimations (from simulated networks) in the literature depend on one-shot assignment models where route choice behavior is incorporated through enroute decision mechanisms (e.g. re-routing to the shortest path based on instantaneous travel costs). Although this approach is able to distribute the congestion in the network, it hardly represents the real driver behavior. This dissertation contributes toward this direction by integrating equilibrium flows in MFD estimation, modeling and control.

Given the complexity and unpredictability of human behavior, route choice modeling is quite challenging. However, it is one of the core components of dynamic traffic assignment (DTA) models that are expected to accurately predict traffic conditions. For instance, dynamic user equilibrium (DUE) implies that travelers have the perfect knowledge of surrounding traffic, and choose the routes with minimum travel cost. In this case, a stochastic model that describes behavioral aspects of route choice phenomenon is not needed. On the other hand, dynamic stochastic user equilibrium (DSUE) implies that travelers have imperfect knowledge, and choose the routes with minimum perceived travel cost, where perception depends on the application and the chosen discrete choice model. The increased use of Global Positioning System (GPS) devices in motor vehicles and smart phones over the last decade allow for sensing and collecting information about human mobility patterns, which could be put to use in order to test the equilibrium assumptions well acknowledged in traffic area. Trajectories of GPS equipped taxis, which represent proxies for human mobility patterns, have been exploited here to discover hidden patterns and identify models that can lead to efficient management strategies. This dissertation contributes in this direction by aggregating the route patterns resulting from observed and shortest path trajectories and by testing the equilibrium assumptions at a regional scale.

Chapter 5 and 6 elaborate travel time prediction and variability models, respectively. A crucial impact of traffic congestion is high travel times on the roadways. Dissemination of accurate travel time information can help travelers make informed decisions such as route choice or departure time. An accurate travel time prediction method should (i) combine historical data analysis and real-time data processing, and (ii) refer to future conditions in the roadway. However, an important part of the literature either uses instantaneous travel time assumption,

sums the travel time of roadway segments at the departure time, and/or deploys statistical forecasting algorithms to predict the future travel time. Statistical methods can be broadly classified in two major categories; parametric methods (e.g. linear regression (Zhang and Rice, 2003), time series models (Yang, 2005; Min and Wynter, 2011), Kalman filtering (Okutani and Stephanedes, 1984; Van Lint, 2008)) and non- parametric methods (neural network models (Ledoux, 1997; Vlahogianni et al., 2005; Van Lint, 2006), support vector regression (Vanajakshi and Rilett, 2007), simulation models (Liu et al., 2006)). The statistical models where the previous travel time is needed as an input (e.g. time series models) suffer from the fact that trip time is usually greater than the prediction interval. In other words, a trip that has started in the previous time interval is usually not completed by the beginning of the next interval. Therefore, experienced travel time from the previous interval is often not available for prediction purposes, and data-driven approaches are usually fed with instantaneous travel time information.

The motivation of this dissertation with respect to travel time prediction is to propose a methodology that integrates identification of traffic patterns with traffic flow theory fundamentals (e.g. with shockwave analysis and bottleneck identification techniques), distinguishes between distinct historical traffic patterns with advanced data mining techniques, and, finally, predicts experienced travel time for the current departure time period.

Another research direction that this dissertation aims to investigate is travel time variability. Travel time is an important index in assessing traffic efficiency, and it provides a common perception to travelers and practitioners. However, travelers not only seek to minimize their travel time on average, but also value its variation. Travel time variability, which develops the formal basis for reliability measures, can be investigated from several angles: vehicle-to-vehicle variability which corresponds to different vehicles traveling the same route at the same time, period-to-period variability corresponding to vehicles traveling the same route at different periods within a day, and day-to-day variability addressing the travel time variations of vehicles crossing the same route at the same period of time on different days (Noland and Polak, 2002). The analysis on day-to-day travel time variability reveals interesting patterns; congestion onset and offset periods with similar mean travel time may exhibit significantly different variance values. The variance of travel time is consistently higher for the departure time periods in the congestion offset, which leads to a counter-clockwise hysteresis loop for the mean-variance curve.

Reduction in travel time variability is at least as important as reduction in mean travel time, as it decreases uncertainty of decision making. Regarding this aspect, travel time reliability, a crucial performance indicator for roadways, has an important effect on route and departure time choice, especially for time constrained trips (e.g. commute to work, trip to airport, etc.). This dissertation takes a further step in day-to-day travel time variability and investigates the empirical implications of aforementioned hysteresis loops.

## 1.2   Thesis objectives and contributions

Part I of the thesis (Chapter 2, 3 and 4) focuses on the interaction between route choice decisions and MFD modeling. The concept of MFD is initially designed as a monitoring tool rather than a prediction model. Nevertheless, in the first part of the dissertation, we explore prediction capability of such models within a multi-region urban network framework. As in any other traffic model, network conditions are largely affected by travelers' route choice in MFD modeling. Therefore, capturing the effects that result from route choice decisions is a key element to accurate prediction of traffic conditions in MFD-based models. Chapter 2 proposes a DTA model that could cope with characteristics of MFD modeling, Chapter 3 expands this model to a route guidance system, and Chapter 4 investigates aggregated route choice patterns from real vehicle trajectories in a mega-city of China.

- **Chapter 2: Approximating dynamic equilibrium conditions with macroscopic fundamental diagrams**

  The main objective of this work is to develop a DTA model that can be integrated with MFD modeling in order to establish dynamic equilibrium conditions in the network. The proposed method should address the route choice behavior in case of heterogeneous urban networks, where subnetworks of the city exhibit different traffic features. In addition, the method should handle the stochastic components that exist in the parsimonious network model (e.g. trip length uncertainty, variation of speed across the links, traveler perception), and reach approximate dynamic equilibrium conditions.

  The proposed DTA model establishes dynamic equilibrium conditions in multi-region urban networks where the traffic modeling is done through MFD dynamics. The model consists of two main components; stochastic network loading (SNL) and a fixed-point solution method. Stochastic elements addressed by SNL include varying average trip length across the regions, random distribution of link speeds inside the regions and perception of travelers which can be quite significant due to the aggregated model in hand. Unlike conventional stochastic user equilibrium (SUE) definition, stochastic components do not solely arise from the the perception error that can be represented by statistical models. Therefore, the method does not assign any explicit statistical distribution on the error components. Instead, the model follows a sampling approach where the location-specific route choice behavior can be taken into account. The second part of the model consists of a well-known fixed point solution heuristics; method of successive averages (MSA). The method averages auxiliary directions taken from the SNL procedure, and establishes dynamic equilibrium conditions in the network with an iterative approach.

- **Chapter 3: Equilibrium analysis and route guidance in large-scale networks with MFD dynamics**

  The main objective of Chapter 3 is to extend the DTA model presented in Chapter 2 to a route guidance model, where travelers are given a sequence of neighborhoods to

follow. In addition, this chapter investigates the effect of equilibrium conditions on the network-wide properties, in particular MFD functions.

The extension of DTA model to a route guidance scheme, first of all, requires two traffic models with different aggregation levels; an aggregated model for route guidance design (i.e. operation model) and a detailed model for testing (i.e. plant, reality). In this respect, the network is partitioned into few regions where each region includes several subregions. We assume that a well defined low scatter MFD holds for subregions. However, the aggregation of subregion MFDs does not guarantee a well defined region MFD as variability across the subregions can cause hysteresis. Heterogeneity modeling presented in (Ramezani et al., 2015) allows us to build such two models (i.e. subregion and region based) and to incorporate link density heterogeneity in the MFD analysis. In order to minimize the system cost, the method establishes system optimum (SO) conditions in the region-based model and produces route choice decisions in the subregion-based model that would satisfy the same conditions. This would require translation of certain variables across the traffic models and yields a certain degree of consistency between them. The procedure follows a rolling horizon approach. That is, at each time step the method establishes SO conditions in the operation model, i.e. the region-based model, over a rolling horizon, identifies a sequence of route choice decisions, but applies only the ones corresponding to the first time step in the plant, i.e. the subregion-based model. The procedure is repeated with a shifted horizon in the next time step. In addition, this chapter establishes equilibrium conditions in the lower-level model (i.e. subregion-based model in this case) and investigate its impact on the resulting region MFDs. On the other hand, existing studies depend on one-shot assignment models that do not necessarily result in equilibrium flows.

- **Chapter 4: Aggregated dynamic route choice patterns for large-scale mixed urban/freeway networks**

  The key objective of Chapter 4 is to observe and reconstruct aggregated route choice patterns in large-scale congested urban/freeway networks. The proposed method should be based on probe data to exploit the aggregated patterns in vehicular traffic dynamics. Aggregated routing patterns that result from this analysis could be integrated with MFD modeling in order to produce better traffic predictions in large-scale congested networks.

  A large-scale mixed network could require traffic models that represent different layers; homogeneous urban regions could be modeled with MFD, and freeway dynamics could be based on link traffic flow models. However, an accurate model at this scale would yield certain variables that result from the aggregated route choice decisions. In this chapter, we exploit a very detailed GPS data set of 20,000 taxis from Shenzhen, China, and we investigate if consistent route choice patterns could be observed. In addition, we investigate how well these patterns can be reconstructed with shortest path trajectories that result from a number of travel cost definitions. In other words, we replace each observed trajectory with the shortest path that connects the same origin and destination

points, and reconstruct the aggregated patterns. Notably, this assumption is in line with DUE conditions. Furthermore, high number of probe vehicles in the data set allows us to monitor traffic conditions in the network and incorporate time-dependent travel costs (due to congestion) into the procedure. On the other hand, due to data availability issues, state of the art in route choice modeling focuses on either survey data or uncongested networks where travel times can be assumed static (Ramming, 2001; Cantillo et al., 2006; Frejinger and Bierlaire, 2007). While there is strong understanding and vast literature of route choice modeling, this study is a first attempt, to the authors' knowledge, that investigates aggregated route choice patterns. This approach allows us to contemplate the validity of equilibrium conditions at the network scale.

Part II focuses on the analysis of travel time as an important performance measure of transportation systems. Travel time information can help travelers make informed decisions, and allow practitioners evaluate the level of service in the network. While prediction of travel time could enable both travelers and traffic engineers to make smarter decisions and avoid delays in real-time, reliability (or variability) measures of the roadways might influence structured recurring travel decisions (e.g. route choice, departure time) and signal chronic traffic problems. Chapter 5 proposes a novel travel time prediction method, and Chapter 6 investigates the hysteresis patterns in day-to-day travel time variability.

- **Chapter 5: Experienced travel time prediction for congested freeways**

  The main objective of Chapter 5 is to develop a method to predict travel times in congested freeway routes. The method should identify the congestion patterns along the route, combine historical and real-time traffic information and predict the experienced travel time for the departing trips.

  This chapter proposes a novel travel time prediction method that relies on traffic flow fundamentals and advanced data mining techniques rather than pure statistical models. First, the proposed method identifies the congestion patterns (i.e. active bottleneck locations and their spatial extent) through an existing algorithm that is based on speed difference at successive links. Second, we cluster the days with similar traffic patterns and build sub-libraries in the historical data set. Third, for each sub-library in the historical data set, we compute the probability of hitting congestion at a space-time domain. Finally, we determine the congestion pattern (among all in the sub-library) that best matches the real-time traffic information, and estimate the congestion evolution which leads to the prediction of experienced travel time. The proposed method is robust to measurement errors, is consistent with the physics of traffic flow and easy to implement.

- **Chapter 6: Investigating empirical implications of hysteresis in day-to-day travel time variability**

  The main objective of this work is to investigate empirical causes of counter-clockwise loops (i.e. hysteresis) observed in day-to-day travel time variability analysis. The pro-

posed method should decompose the problem into certain 'observable' components, and analyze their partial effects on the existence and the shape of hysteresis loops.

Substantially different variance levels in the network loading and unloading periods (where mean travel times are similar) can significantly diversify travel behavior (i.e. departure time, route choice) in the two phases. This chapter develops a within-day travel time evolution model or a piecewise linear function where the change in travel time can be described by a set of 'key' variables (e.g. start time of congestion, maximum travel time, etc.). Further, we build a Monte Carlo simulation model where we can generate key variables and consider the correlation structure associated with them. The proposed method finally argues on the partial effect of these variables by manipulating the correlation structure.

## 1.3   Thesis outline

This thesis consists of 7 chapters including the introduction and the conclusion. The main 5 chapters (Chapter 2-6) are presented in two parts. Part I includes Chapter 2, 3 and 4 that propose novel methods to integrate route choice effects into MFD modeling. Part II includes Chapter 5 and 6 that introduce novel travel time prediction and variability models, respectively. Note that each chapter is a stand-alone article with its own mathematical notation.

**Chapter 2** proposes a DTA model to incorporate equilibrium conditions into MFD modeling schemes. The method addresses the traffic loading problem in a heterogeneous urban network that is partitioned into multiple homogeneous regions. The preliminary results of this chapter are presented in:

- Yildirimoglu, M., and Geroliminis, N. "Dynamic traffic assignment with macroscopic fundamental diagrams", In *13th Swiss Transportation Reasearch Conference*, Ascona, Switzerland, 2013.

- Yildirimoglu, M., and Geroliminis, N. "Approximating dynamic equilibrium conditions with macroscopic fundamental diagrams", In *93rd Annual Meeting of the Transportation Research Board*, Washington, D.C., 2014.

- Yildirimoglu, M., and Geroliminis, N. "Approximating dynamic equilibrium conditions with macroscopic fundamental diagrams", In *14th Swiss Transportation Reasearch Conference*, Ascona, Switzerland, 2014.

Chapter 2 is a self stand-alone article published as:

- Yildirimoglu, M., and Geroliminis, N. "Approximating dynamic equilibrium conditions with macroscopic fundamental diagrams", *Transportation Research Part B: Methodological*, 70, pp. 186-200, 2014.

**Chapter 3** extends the DTA model presented in Chapter 2 to a route guidance system that minimizes the total delay in the system through SO conditions. This chapter also investigates the effect of equilibrium flows on the resulting MFDs. The preliminary results of this chapter are presented in:

- Yildirimoglu, M., Ramezani, M., and Geroliminis, N. "A rolling horizon approach for route guidance in large-scale networks", In *94th Annual Meeting of the Transportation Research Board*, Washington, D.C., 2015.

- Yildirimoglu, M., Ramezani, M., and Geroliminis, N. "Equilibrium analysis and route guidance in large-scale networks with MFD dynamics", *Transportation Research Procedia*, 9, pp. 185-204, 2015.

Chapter 3 is a self stand-alone article published as:

- Yildirimoglu, M., Ramezani, M., and Geroliminis, N. "Equilibrium analysis and route guidance in large-scale networks with MFD dynamics", *Transportation Research Part C: Emerging Technologies*, (available online).

**Chapter 4** observes and reconstructs the aggregated route choice patterns in a large-scale mixed network. The probe vehicles provide the trajectory information that is further exploited in this work to extract consistent route patterns. The method also reconstructs these patterns by replacing each observed trajectory with the shortest path that results from a number of travel cost definitions. This part of the thesis is still research in progress. This chapter is accepted to *95th Annual Meeting of the Transportation Research Board* for presentation.

**Chapter 5** proposes a novel travel time prediction model based on traffic flow fundamentals and advanced data mining techniques. Travel time information provides travelers with a crucial performance measure that helps them assess the roadway system and make better decisions regarding the route choice or departure time. The preliminary results of this chapter are presented in:

- Yildirimoglu, M., and Geroliminis, N., "Travel time estimation and prediction in freeway systems", In *12th Swiss Transportation Reasearch Conference*, Ascona, Switzerland, 2012.

- Yildirimoglu, M., and Geroliminis, N., "Experienced travel time prediction in freeway systems", In *Intelligent Transportation Systems (ITSC), 15th International IEEE Conference*, Anchorage, AK, 2012.

- Yildirimoglu, M., and Geroliminis, N. "Experienced travel time prediction in congested freeway routes", In *92nd Annual Meeting of the Transportation Research Board*, Washington, D.C., 2013.

Chapter 5 is a self stand-alone article published as:

- Yildirimoglu, M., and Geroliminis, N. "Experienced travel time prediction for congested freeways", *Transportation Research Part B: Methodological*, 53, pp. 45-63, 2013.

**Chapter 6** investigates the empirical implications of hysteresis shape in day-to-day travel time variability. The method approximates the evolution of travel time within the day with a piecewise linear function that enables us to extract certain key variables. This chapter identifies the partial effects of these variables on the existence of hysteresis through a Monte Carlo simulation. Chapter 6 is a self stand-alone article published as:

- Yildirimoglu, M., Limniati, Y., and Geroliminis, N. "Investigating empirical implications of hysteresis in day-to-day travel time variability", *Transportation Research Part C: Emerging Technologies*, 55, pp. 340-350, 2015.

# Part I

# 2 Approximating dynamic equilibrium conditions with macroscopic fundamental diagrams

R EAL-TIME coordinated traffic management strategies that benefit from parsimonious models with aggregated network dynamics, provide a new generation of smart hierarchical strategies to improve network capacity and performance. However, the question of route choice behavior is raised in case of heterogeneous urban networks, where different parts of the city are subject to different types of control. Traffic equilibrium phenomena have not been thoroughly investigated in these models. Approximate traffic equilibrium conditions can be integrated within the parsimonious traffic models to develop regional routing strategies, while detailed route choice strategies can be incorporated at a later stage in a hierarchical framework. In this study, we develop an aggregated and approximate dynamic traffic assignment (DTA) procedure to be incorporated in the macroscopic fundamental diagram (MFD) dynamics, and establish dynamic stochastic user equilibrium (DSUE) conditions. The methodology consists of two main components; stochastic network loading and a fixed-point solution method. Loading procedure is designed to handle stochastic components in the model such as trip length uncertainty, variation of speeds across the links, perception error of travelers. The results taken from this procedure are averaged through the well-known method of successive averages (MSA) to reach fixed-point solution for the system. Real-time route guidance strategies can be revisited towards a "system of systems" approach.

## 2.1 Introduction

[1] Traffic congestion management in large-scale systems is currently fragmented and uncoordinated with respect to optimizing the goals of travel efficiency and equity for multiple regions of a city. An alternative is a hierarchical control structure, where a network can be partitioned in homogeneous regions and optimal control methodologies can identify the inter-transfers among regions to maximize the system output. Understanding these interactions for het-

---

[1]The notation in this chapter has been altered from Yildirimoglu and Geroliminis (2014) in order to build consistency within the thesis.

erogeneously congested cities is a big challenge, which will allow revisiting, redesigning and integrating smarter traffic management techniques to alleviate congestion with a "system of systems" approach (Haddad et al., 2013).

Real-time coordinated traffic management strategies (e.g. perimeter control, gating) that benefit from parsimonious models with aggregated network dynamics, provide promising results towards a new generation of smart hierarchical strategies to improve network capacity and performance, see for example (Keyvan-Ekbatani et al., 2012; Geroliminis et al., 2013; Aboudolas and Geroliminis, 2013). However, the developed optimization framework will not be realistic if it does not consider that people might adapt their routes in response to the control strategies. For example, people might choose longer but more reliable routes when heavy congestion occurs. Additionally, traffic management that aims to establish equilibrium conditions by routing vehicles to the minimal travel time paths can significantly improve the performance. While active traffic management for large scale networks with route choice adaptation will be studied in the future, this work advances the MFD modeling with aggregated route choice dynamics and investigates the feasibility of such approach.

Dynamic traffic assignment (DTA) has been extensively studied to satisfy dynamic equilibrium conditions. Various traffic performance models spanning from link travel time functions to detailed microscopic simulation models have been incorporated in DTA. In this study, we incorporate the macroscopic fundamental diagram (MFD) framework into a DTA model, and establish dynamic equilibrium conditions. Given the heterogeneous distribution of congestion in a city and the multiple MFD regions, a region-based route choice model is developed, which can be exploited to test the real performance of dynamic traffic management systems or to develop regional routing strategies.

### 2.1.1 Background on traffic assignment

Static assignment procedures, where flows and travel costs are time-invariant, are classified in two groups according to target network conditions: user equilibrium (UE) and system optimum (SO) (Sheffi, 1985). On the other hand, DTA models consider time-varying flows to adequately capture traffic dynamics and the interaction between travel choices. DTA models aim for dynamic extensions of above target network conditions. Ran et al. (1996) defines dynamic user equilibrium (DUE) conditions: *for each OD pair, if the actual travel times experienced by travelers departing at the same time are equal and minimal, then the dynamic flow over the network is in a travel time based ideal dynamic user equilibrium (DUE) state.* On the other hand, SO assignment involves minimizing total travel cost (Ziliaskopoulos, 2000) by forcing users to cooperate with each other. This normative approach rather than descriptive is beyond the scope of this chapter. Note that DTA models can be further distinguished based on underlying travel choice component: the reactive dynamic user equilibrium (RDUE) assignment, in which travelers choose the route with minimum cost based on instantaneous travel times (Papageorgiou, 1990; Kuwahara and Akamatsu, 1997), and the predictive dynamic

user equilibrium assignment (PDUE), in which travelers choose the route that minimizes their actual travel cost (Mahmassani and Peeta, 1993; Ben-Akiva et al., 1997).

The relaxation of the presumptions considered in the deterministic utility maximization rule to reach UE conditions has been proposed by Daganzo and Sheffi (1977) as stochastic user equilibrium (SUE). SUE defines the equilibrium conditions where users can no longer improve their perceived utility. Traveler's perceived utility incorporates a random component which represents perception errors or randomness in system performance. SUE notion has also been extended to the dynamic case by De Palma et al. (1983): For each OD pair, if travel times perceived by travelers departing at the same time are equal and minimal, then the network is in dynamic stochastic user equilibrium (DSUE) state.

Traffic performance models used in DTA to describe traffic flow propagation on time-varying networks include cell transmission models (Lo and Szeto, 2002), outflow models (Merchant and Nemhauser, 1978), deterministic queuing models (Ben-Akiva et al., 1986), whole link models (Ran and Boyce, 1996), nonlinear time models (Jayakrishnan et al., 1995), and mesoscopic models where vehicles are modeled explicitly (Mahmassani and Peeta, 1993; Peeta and Mahmassani, 1995; Ben-Akiva et al., 1997).Traffic performance models listed above implement a discrete modeling approach, where the elements of traffic network (e.g. links, intersections) are modeled separately, and demand is distributed along the network through hypothetical zone centroids. An alternative to this approach is the continuum modeling where the focus is on the general trend of travel choices at the macroscopic level. The continuum approach models a network as a continuum where travelers are free to choose their paths in two dimensional continuous space. As these models are mainly used for initial planning of large regions, traffic assignments have been mostly limited to static cases. Hoogendoorn and Bovy (2004), Jiang et al. (2011) and Du et al. (2013) develop DTA models that establish DUE conditions via continuum modeling. Both MFD and continuum modeling are important attempts to model a dense urban network, and developed DTA models deal with common aspects of urban network modeling. While the DTA literature review is vast, a detailed description is beyond the scope of this chapter. The interested reader can refer to Peeta and Ziliaskopoulos (2001) and Chiu et al. (2011).

The integration of a DTA framework in real-time traffic management is mainly performed for toy networks with simplified dynamics that are far from real-sized networks or focuses on decentralized control, which is not appropriate for heterogeneously loaded networks. Integrating traffic assignment with traffic signal control has been studied for toy networks with iterative procedures that consider equilibrium flows and resulting signal parameters (Cantarella et al., 1991; Yang and Yagar, 1995). Scenario analysis through simulation models is another direction, which make the problem intractable in real time. In this work we study if an approximate region-to-region route choice can be integrated in a large-scale network which is modeled with multiple MFDs for different regions. Integrating the developed models in a control framework is a necessary future direction.

## 2.1.2 Background on MFD

A homogeneous urban region (with small spatial link density heterogeneity) can be modeled with MFD, which provides a unimodal, low-scatter, and demand-insensitive relationship between network vehicle density and space-mean flow (Geroliminis and Daganzo, 2008). However, urban transportation networks exhibit uneven distribution of congestion which leads to a scattered flow-density relationship. Heterogeneity in congestion distribution can affect the shape/scatter or even the existence of MFD (Buisson and Ladier, 2009; Geroliminis and Sun, 2011b). By using a grid network and considering variance of link density as independent variable, Mazloumian et al. (2010) shows that MFD remains well-defined in sub-regions of the urban network. These results are very critical, because MFD concept can be useful for heterogeneously loaded cities, if the network can be partitioned into a small number of homogenous regions. The effect of heterogeneity has been recently studied by many researchers with similar conclusions with empirical data and simulation, see for example (Knoop et al., 2013; Mahmassani et al., 2013b; Geroliminis and Sun, 2011b). In addition, Gayah and Daganzo (2011), Mahmassani et al. (2013b) and Leclercq and Geroliminis (2013) thoroughly investigate the effect of driver adaptation and route choice on the shape of MFD. To deal with heterogeneous urban networks, Ji and Geroliminis (2012) develop a partitioning mechanism to minimize the variance of link densities while maintaining a spatially compact shape. Resulting sub-regions can be used to develop macroscopic traffic control strategies; e.g. perimeter control.

When the network is heavily congested, the well-defined MFD tends to be less consistent and less reproducible due to scatter and hysteresis issues. Investigating the causes of instability or design of remedies that would have a stabilizing influence is a very important research direction, because well-defined MFDs (even in the congested regime) are very useful in the design and control of urban networks. Note that most of the previous research on MFD estimation is based on microscopic or mesoscopic simulation models where driver adaptivity is modeled with enroute decision mechanisms. Therefore, equilibrium flows and their effects on the MFD stability must be carefully investigated through real data or simulation models with realistic route choice mechanisms. However, despite these disadvantages, various traffic control strategies that employ parsimonious MFD model show promise in more efficient traffic flow at the regional level. As most traffic management schemes tend to keep the urban networks at the free flow or moderately congested regime, where MFD is still well defined, the outcome of the control strategies is expected to have a significant effect on the network. Some studies that report benefit from control schemes using MFD are Keyvan-Ekbatani et al. (2013), Geroliminis et al. (2013), Aboudolas and Geroliminis (2013). These studies either consider a single protected region operating close to the maximum MFD flow or time-independent multiple regions with static partitioning. Nevertheless, congestion propagation in both time and space might require the development of more advanced time-dependent partitioning algorithms to be integrated with dynamic perimeter control. This work is under development.

Haddad et al. (2013) develop cooperative control strategies for a large scale mixed trans-

portation network that consists of one freeway and two homogeneous urban regions. They implement model predictive control schemes to determine the optimal inter-transfer flows at the boundary of two regions and the metering rate at the freeway entrance. In order to account for the change in route choice decision in response to the new control policy, they develop a dynamic simple route choice model, in which travelers choose the route with minimum instantaneous travel time. However, this model does not satisfy the equilibrium conditions. The exemplary multi-region urban network presented in Figure 2.1 shows a case where there are multiple paths connecting the same OD regions. In this framework, assigned flows on the alternative paths must be known in advance to predict traffic conditions in the network and to take better control decisions accordingly. A traffic assignment procedure that establishes equilibrium conditions is the proper way to achieve the demand level on the alternative paths. Note that previous empirical studies observed time-invariant trip lengths for homogeneously congested networks (Geroliminis and Daganzo, 2008). In this work, we show that in case of heterogeneous networks with multiple MFD regions, regional trip lengths can vary over time, because drivers have the ability to choose a different sequence of regions to decrease their travel times.
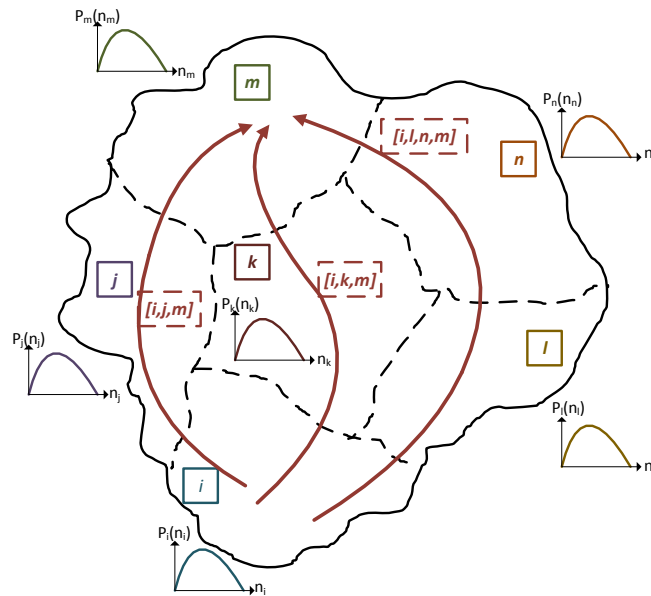


Figure 2.1: A multi-region urban network

Alternatively, one can develop routing strategies using the aggregate information from subnetworks and MFD relationship. Traffic control on the regional level is promising, because it can remove the computational burden of link-level modeling and overcome real-time deployment issues. Knoop et al. (2012) develop dynamic routing strategies considering the aggregated information from multiple grid sub-networks.

In this chapter, we incorporate MFD into a DTA model, and establish DSUE conditions regarding experienced travel costs. The approach presented in this study can be deployed as a next step to evaluate the real effect of dynamic traffic management strategies using aggregated network dynamics or to develop regional routing strategies aiming for equilibrium conditions. The outline of the chapter is as follows. The next section introduces the methodology that establishes DSUE conditions in a multi-region MFD framework. The following section presents the results of a case study, and the last section gives the discussion and the conclusion of the study.

## 2.2    Methodological framework

The problem is to identify equilibrium conditions at the regional level, as expressed by time dependent regional route choice estimation, in a city divided into N regions with a low-scatter MFD representing travel production (veh-km travelled per unit time) vs. accumulation. A time-aggregated regional OD (i.e. origin and destination regions) demand is also given. The objective of this work is to integrate later the effect of aggregated route choice in active traffic management schemes (e.g. route guidance, perimeter control).

As opposed to the existing DTA studies, we try to determine the sequence of regions between different origin-destination pairs instead of the exact sequence of links. While this approach might sound computationally lighter, it creates additional challenges in the modeling part (from an assignment point of view), as for example that trips which start later in a region can finish earlier, because of variant trip lengths. Additional challenges (from an MFD point of view) is that this approach requires aggregate modeling of traffic flow within urban regions, and it describes traffic flow propagation within or between the regions using average trip lengths. However, average trip length is not informative enough to build a route choice framework; instead, trip length distributions within each region and for each regional OD must be considered to ensure equilibrium conditions in the system. This study does not explicitly determine trip length distributions, but it deals with them in an iterative way within stochastic network loading (SNL) procedure. In addition, perception error of travelers in the aggregated network dynamics is addressed in SNL through a logit formula. In that respect, this approach should be considered to aim for DSUE rather than its deterministic ancestor, DUE. The results taken from SNL are processed through the well-known fixed-point solution algorithm called method of successive averages (MSA). It is an effective solution heuristic which is highly implemented in simulation-based DTA (Peeta and Mahmassani, 1995). Given that it does not require derivative information for the flow-cost mapping function, MSA can be easily adapted to simulation studies. MSA uses the traffic model (i.e. aggregated network or MFD dynamics in this case) in each iteration, to project future traffic information as part of the direction finding mechanism in searching for a solution. Figure 2.2 depicts a flowchart of the methodological framework, whose details will be presented in the following subsections.

As simulators take significant amount of time, simulation-based DTA strategies are not suitable
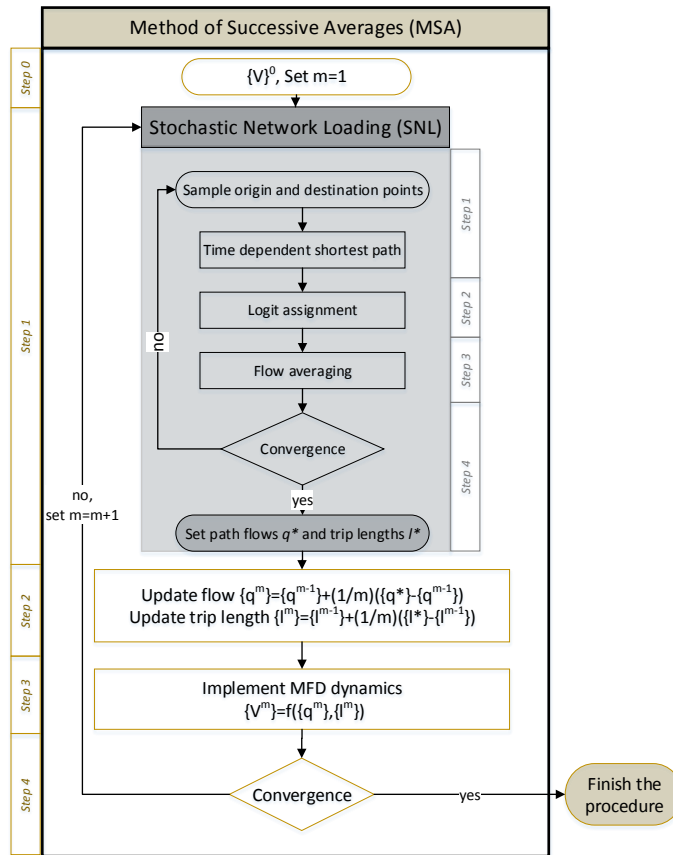
Figure 2.2: Methodological framework

for real-time deployment (Peeta and Ziliaskopoulos, 2001). On the other hand, MFD approach, even in case of large-network modeling, can get along with a relatively simple representation structure. However, this does not overcome real-time deployment issues. To observe the state of the urban network or MFD in real-time, (preferably midblock) loop detectors must be installed in major links of the network and the data from them must be available in real-time. The real world implementation of the proposed model would also require rolling horizon or feedback loops which would correct the error in the state variables. However, this would yield trajectory information from significant part of the demand along with their destination information. MFD observation is a rather easy and well defined problem, while obtaining trajectory data from GPS devices which still have very low penetration rate is still an open question. Therefore, the real-time deployment issues remain valid. Recent work towards this direction has been reported by Ortigosa et al. (2013a), Gayah and Dixit (2013a), Leclercq et al. (2014) and Ji et al. (2014).

Figure 2.3 presents two representations of the same network that we investigate in this chapter. Link-level representation is valid for discrete traffic performance models (e.g. outflow

models, mesoscopic models, etc.). However, detailed information (e.g. signal settings, lane configuration, etc.) is needed to accurately model link and node dynamics. On the other hand, regional representation, which can be exploited in the aggregated dynamics, consists of partitions with separate MFD functions and boundaries connecting them. The assignment model developed in this chapter makes use of both representations. The graph structure that builds the link-level representation (without any link-level traffic dynamics) is employed in SNL to produce aggregated traffic parameters, while regional representation is needed to run aggregated dynamics.



(a)

(b)
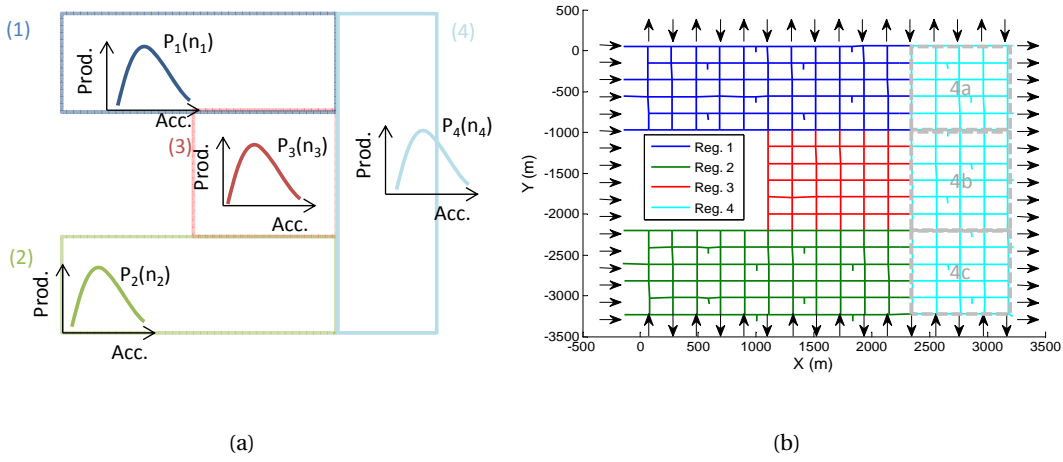
Figure 2.3: Case study network (a) regional representation, (b) link-level representation

Methodology section is structured as follows. The next subsection introduces the aggregated network dynamics defined for multiple homogeneous regions with the newly proposed dynamic average trip length consideration. The following subsection gives details of SNL which addresses the stochastic components existing in the model. The last subsection presents MSA, the fixed point solution algorithm deployed to reach equilibrium conditions.

### 2.2.1 Aggregated network dynamics in a multi-region urban framework

Consider an urban region $r$ with homogeneous distribution of congestion whose traffic performance is well described by MFD, $f_r(n_r(t))$ represents the production [veh.km/h] corresponding to the accumulation [veh] $n_r(t)$ at time $t$, $v_r(t) = (f_r(n_r(t)))/(n_r(t))$ represents the speed [km/h] in region $r$, and trip completion rate [veh/h] is calculated to be $m_r(t) = (f_r(n_r(t)))/(l_r)$, considering the average trip length [km] $l_r$. Note that time $t$ is a continuous variable.

Geroliminis and Daganzo (2008) investigated the relation between the production of an urban network and the outflow at its boundary, and observed that the proportion of the production to the outflow indicates a relatively constant value throughout the day. This constant represents the average vehicular trip length, $l_r$, in urban region $r$. If vehicular trajectory data was made

available, one could test the consistency between this constant (proportion of production to outflow) and the average distance crossed by actual vehicles (through trajectories). However, this analysis is beyond the scope of this chapter due to lack of real data. Nevertheless, in case of heterogeneous networks with multiple MFD regions, regional trip lengths can vary over time, because drivers might choose a different sequence of regions to decrease their travel times. For this analysis, we remove the assumption of stationarity (constant average trip length) to enable a higher level of flexibility in the system (e.g. travelers can choose a longer but more reliable route in case of congestion).

In this study, as there are multiple groups of travelers that have different paths and cross different boundaries of the same region, average distance travelled within the region to reach the boundaries might not be the same. In addition, network topology and orientation of the urban regions might play a significant role in the average trip length consideration. This study revisits the constant average trip length assumption in the MFD approach and replaces it with the dynamic average trip length defined specifically for each group. This allows us to rewrite aggregate MFD dynamics for a multi-region urban network.

Consider a city partitioned in $N$ regions (see Figure 2.1 and Figure 2.3). Denote by $r=1,2,...,N$ a region in the network, $n_r(t)$ its total accumulation at time $t$, and $n_{o,d}^{p,r}(t)$ the number of vehicles in region $r$ at time $t$ with origin $o$, destination $d$, and path $p$ (set containing the sequence of regions to reach $d$ starting from $o$; note that $r$ belongs to $p$). Trip completion rate $m_{o,d}^{p,r}(t)$ for the vehicles in region $r$ at time $t$ with origin $o$, destination $d$, and path $p$ is calculated by the following equation that can be derived by Little's formula[2].

$$m_{o,d}^{p,r}(t) = v_r(t) \cdot \frac{n_{o,d}^{p,r}(t)}{l_{o,d}^{p,r}(t)} = \frac{f_r(n_r(t))}{n_r(t)} \cdot \frac{n_{o,d}^{p,r}(t)}{l_{o,d}^{p,r}(t)} \tag{2.1}$$

Note that $l_{o,d}^{p,r}(t)$ represents the average trip length crossed in region $r$ by the people who travel between origin $o$ and destination $d$, using path $p$ at time $t$. Similarly, $l_{o,d}^{p}(t) = \sum_{r \in p} l_{o,d}^{p,r}(t)$ is the total average trip length across regions on path $p$. For instance, considering traffic conditions in the network, travelers who start their trips from different parts of the region can take different decisions causing a significant change in the average trip length. We will show later that such an addition is necessary for realistic representation of the dynamics.

Denote $q_{o,d}(t)$ the exogenous demand generated at time $t$, from origin $o$ to destination $d$. As there are multiple paths connecting the same OD pair, demand will be distributed among them in a way that satisfies DSUE conditions. The sum of path-specific flows $q_{o,d}^{p}(t)$, computed through DTA, should be equal to $q_{o,d}(t)$. Let $\hat{m}_{o,d}^{r \to p^+(r)}(t)$ be transferring flow from region $r$ to region $p^+(r)$, which is the next region in the sequence described by path $p$. Transferring flow from the previous region $p^-(r)$ to region $r$ in path $p$, is the minimum of two terms: (i) the sending flow from region $p^-(r)$, which depends on the accumulations of region $p^-(r)$

---

[2]Using Little's formula and omitting time $t$: (i) for the total accumulation of region r and (ii) only for vehicles in $(o, d, p, r)$ group, we have (i) $n_r = m_r * l_r / V_r$ and (ii) $n_{o,d}^{p,r} = m_{o,d}^{p,r} * l_{o,d}^{p,r} / V_r$

and (ii) the boundary capacity of region $r$, which is a function of the accumulation in $r$; $c_{p^-(r),r}(n_r(t))$. Note that boundary capacity decreases for accumulations much larger than the critical accumulation that maximizes production. Dynamic equations are listed below. Time $t$ is omitted from the equations for the sake of notational simplicity.

$$
\frac{\mathrm{d}n_{o,d}^{p,r}}{\mathrm{d}t} = \begin{cases} q_{o,d}^{p} - m_{o,d}^{p,r} & \text{(i) if } r = o \ \& \ r = d, \\ q_{o,d}^{p} - \hat{m}_{o,d}^{r \to p^+(r)} & \text{(ii) if } r = o \ \& \ r \neq d, \\ \hat{m}_{o,d}^{p^-(r) \to r} - m_{o,d}^{p,r} & \text{(iii) if } r \neq o \ \& \ r = d, \\ \hat{m}_{o,d}^{p^-(r) \to r} - \hat{m}_{o,d}^{r \to p^+(r)} & \text{(iv) otherwise.} \end{cases} \tag{2.2}
$$

where

$$
\hat{m}_{o,d}^{r \to p^+(r)} = \min[m_{o,d}^{p,r}, c_r^{p^+(r)}(n_{p^+(r)}) \cdot a_{o,d}^{r \to p^+(r)}]. \quad \forall \, r \neq d \tag{2.3}
$$

Eq. 2.2 defines the rate of change in accumulation $n_{o,d}^{p,r}$ such that in case of (i) internal demand within the same subregion, the rate is simply the exogenous demand minus the trip completion rate which is not bounded by any capacity function. Note that the model assumes that internal demand never leaves the region. In case (ii) current region $r$ is the origin and not the destination, then the rate is the exogenous demand minus the transfer flow to the next region in path $p$. In case (iii) current region $r$ is destination and not the origin, the rate is defined as the transfer flow from the previous region minus the trip completion rate which is again not bounded by any capacity function. In (iv) other cases, the rate is equal to the transfer flow from the previous region minus the transfer flow to the next region.

$a_{r \to p^+(r)}^{o,d}$ in Eq. 2.3 is the fraction of inflow capacity $c_{r,p^+(r)}(n_{p^+(r)})$ from region $r$ to $p^+(r)$, for vehicles with next region $p^+(r)$, and can be computed with Eq. 2.4. Conceptually speaking, Eq. 2.4 states that the ratio of the quartet $(o, d, p, r)$ depends on its relative accumulation and regional trip length among all possible traveler groups that will cross the same boundary between $r$ and $p^+(r)$.

$$
a_{o,d}^{r \to p^+(r)} = \frac{n_{o,d}^{p,r}/l_{o,d}^{p,r}}{\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N_{ij}} \mathbf{1}_{(p^+(r))}(k^+(r)) \cdot n_{i,j}^{t,r}/l_{i,j}^{k,r}} \tag{2.4}
$$

where $N_{ij}$ is the number of paths between origin $i$ and destination $j$, and $\mathbf{1}_{(p^+(r))}(t^+(r))$ is an indicator function with value equal to 1 if the next reservoir in the paths $p$ and $t$ is the same. This equation can be derived considering the outflow from different traveler groups[3].

---

[3] Following Eq. 2.1 and omitting time $t$, total outflow from region $r$ is $m_r = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N_{ij}} m_{i,j}^{k,r} = v_r \cdot \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N_{ij}} \frac{n_{i,j}^{k,r}}{l_{i,j}^{k,r}}$

and proportion of $(o, d, p, r)$ specific flow to total outflow is $\dfrac{m_{o,d}^{p,r}}{m_r} = \dfrac{n_{o,d}^{p,r}/l_{o,d}^{p,r}}{\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N_{ij}} n_{i,j}^{k,r}/l_{i,j}^{k,r}}$. Eq. 2.4 defines the

the ratio of $(o, d, p, r)$ specific flow to the total outflow through a particular boundary assuming the proportion remains the same when boundary capacity is reached.

Time $t$ is defined as a continuous variable from traffic modeling perspective and Eq. 2.1-2.4 are computed through a differential equation solver in order to satisfy (almost) continuous accumulation updates. However, assignment decisions (i.e. $q_{o,d}^{p}(t)$ and $l_{o,d}^{p,r}(t)$ ) have to be taken at discrete time intervals. Therefore, between any two discrete time points, these variables are assumed constant. In the remainder of Chapter 2, time $t$ refers to discrete time units.

### 2.2.2 Stochastic network loading (SNL)

Stochastic network loading procedure has been deployed to address the stochastic components in the proposed approach. These components include variant trip lengths, randomness of the system due to aggregate traffic modeling and perception error of travelers. Variant trip lengths, in particular, play a very important role in the proposed structure. Although the original MFD modeling considers constant trip length to determine the outflow relationship, dynamic nature of trip length parameter has to be investigated in a route choice framework. Although various stochastic loading procedures (e.g. (Dial, 1971; Ran and Boyce, 1996), etc.) have been proposed according to the behavioral assumption governing route choice, the aggregate modeling in hand creates additional challenges which cannot be addressed by the existing methodologies.

The approach presented in this study, initially, does not assign any explicit travel time distribution on the paths. Instead, by assuming origins and destinations to be randomly distributed in the corresponding regions, the model generates start and end points for trips within the same region or across regions. Distribution of origin and destination points reflects both the volume and the spatial structure of the traffic demand. In this work, we assume they are uniformly distributed within the regions. The procedure, then, deploys the time dependent shortest path algorithm developed by Chabini (1998) on each alternative regional path for the generated start and end points. This algorithm provides the optimal run solution for all-to-one dynamic shortest path problems. Note that regional path defines the sequence of regions to be followed to reach the destination point, not the detailed sequence of links. However, shortest path calculation is done on the graph that represents the connectivity between all the links (see Figure 2.4b). Speed on the links is based on the accumulation of the vehicles in the corresponding region; speed information is determined from the MFD relationship and accumulation level. In other words, the average speed in the region extracted from the MFD relationship is considered as the representative for the speed of all links in the region. Finally, the procedure takes a regional path decision based on logit discrete choice model and shortest travel times on alternative regional paths. Note that this procedure does not account for capacity, it simply assigns the demand given the traffic conditions. The interaction between route choice decisions and traffic conditions will be taken into account in Section 2.2.3.

The loading algorithm can be summarized as follows:

## Chapter 2. Approximating dynamic equilibrium conditions with macroscopic fundamental diagrams

**Step 0:** *Initialization.* Set iteration number $n=1$.

**Step 1:** *Sampling and travel time computation.*

- For each $(o, d, t)$ triplet, sample origin and destination points within the first and last regions of the trip (see the sampled points -O and D- in Figure 2.4a).

- For each regional path $p$ between origin $o$ and destination $d$, identify time dependent shortest path at departure time $t$. As path $p$ represents only the sequence of regions, there are many possible link sequences that comply with this definition. Therefore, Chabini's algorithm (Chabini, 1998) is applied to determine the lowest experienced travel time for the specific $(o, d, p, t)$ quartet, denoted by $\overline{T}_{o,d}^{p}(t)$. The algorithm also returns the sequence of links with $\overline{T}_{o,d}^{p}(t)$ among all possible sequences that comply with regional path $p$. Figure 2.4b indicates shortest paths on the alternative regional paths. Note that small circles at the boundary nodes divide the paths into smaller sections, which define the distance travelled in different regions, denoted by $\overline{l}_{o,d}^{p,r}(t)$. $\alpha$ is the scale factor of the logit formula, and it should be carefully estimated in case real choice data sets are available.

**Step 2:** *Logit assignment.* Based on the lowest travel times $\overline{T}_{o,d}^{p}(t)$ calculated in the previous step, compute probability of using path $p$ for the demand between $o$ and $d$ with the following logit formula; $\overline{P}_{o,d}^{p}(t) = e^{-\alpha \cdot \overline{T}_{o,d}^{p}(t)} / \sum_k e^{-\alpha \cdot \overline{T}_{o,d}^{k}(t)}$, where $\alpha$ is the scale factor that should be properly estimated with maximum likelihood methods in case real choice data sets are available.

**Step 3:** *Flow and trip length averaging.*

- $[q_{o,d}^{p}(t)]^{n} = \left( (n-1) \cdot [q_{o,d}^{p}(t)]^{n-1} + \overline{P}_{o,d}^{p}(t) \cdot q_{o,d}(t) \right) / n$ for each $(o, d, p, t)$ quartet.

- $[l_{o,d}^{p,r}(t)]^{n} = \dfrac{(n-1) \cdot [q_{o,d}^{p}(t)]^{n-1} \cdot [l_{o,d}^{p,r}(t)]^{n-1} + \overline{P}_{o,d}^{p}(t) \cdot q_{o,d}(t) \cdot \overline{l}_{o,d}^{p,r}(t)}{(n-1) \cdot [q_{o,d}^{p}(t)]^{n-1} + \overline{P}_{o,d}^{p}(t) \cdot q_{o,d}(t)}$

  for each $(o, d, p, r, t)$. Note that assigned flows from the previous iteration are weighted with the corresponding iteration number $n-1$ to account for all preceding samples.

**Step 4:** *Stopping test.*

- For each $(o, d, p, t)$ quartet, evaluate $\left[ [q_{o,d}^{p}(t)]^{n} - [q_{o,d}^{p}(t)]^{n-1} \right] \geq \epsilon$.

- Calculate $N(\epsilon)$ the number of cases for which the above criterion is violated.

- If $N(\epsilon) \geq \psi$, set $n = n + 1$ and go to Step 1.

- Otherwise, finish the procedure, assign $[l_{o,d}^{p,r}(t)]^{*} = [l_{o,d}^{p,r}(t)]^{n}$ for each $(o, d, p, r, t)$, and assign $[q_{o,d}^{p}(t)]^{*} = [q_{o,d}^{p}(t)]^{n}$ for each $(o, d, p, t)$ quartet.

- Alternatively, one can deploy a predetermined number of steps regardless of the convergence state. This might eventually increase the speed of the whole assignment procedure.
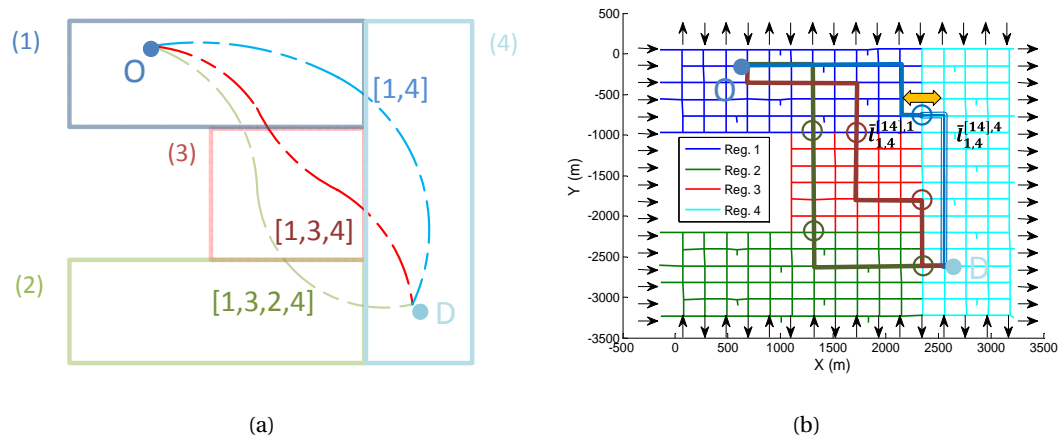
Figure 2.4: Sampling procedure on (a) regional representation, (b) link-level representation

Note that the loading procedure incorporates two critical steps; random sampling of origin and destination points and logit assignment. The former is implemented to account for variant trip lengths within and between regions. On the other hand, logit assignment has been incorporated in the procedure to address travelers' perception of travel time and randomness in the system. MFD approach requires aggregated network modelling in the spatial aspect, and variation of speeds across the links in the same region is not considered in the shortest path calculation. This additional variation along with perception errors are expected to be satisfied by the logit assignment procedure. Logit model could also be replaced with a cross nested logit or C-logit model to address the correlation between alternative paths. However, in our approach, correlation between alternatives arises from both (i) physical path overlapping and (ii) sharing the same region and contributing to its accumulation. Therefore, path overlapping measures could not solely extract the correlation structure between alternative regional paths. A rigorous analysis that investigates correlation issues is beyond the scope of this chapter.

### 2.2.3 Method of successive averages (MSA)

Traffic equilibrium can be formulated as a fixed-point problem, where an additional cycle of assignment and loading steps yield the same traffic conditions. The well-known solution heuristics, MSA, is a suitable method in our study considering the characteristics of the problem in hand. MSA has been used in both static and dynamic network equilibrium problems as an incremental assignment type heuristic (Daganzo and Sheffi, 1977; Mahmassani and Peeta, 1993). The method is based on predetermined step sizes along the descent direction. In other words, step size is not determined with respect to the characteristics of the current solution, which requires derivative information. Instead, it is determined a priori. Therefore, it stands as one of the effective solution heuristics in case the derivative information is difficult to be acquired. Nevertheless, traffic assignment literature on MSA relies on small 'toy' networks, because it is known to suffer from slow convergence problems in case of large-scale congested

transportation networks. On the other hand, this study, using the MFD approach in the modeling of urban networks, is not expected to suffer from these limitations. In addition, other averaging techniques such as adaptive averaging (Magnanti and Perakis, 1997), method of successive weighted averages (Liu et al., 2009), generalization of secant methods (Bierlaire and Crittin, 2006) could be employed here to reach equilibrium flows.

Denote $v = \{v_r(t)|\forall\, r, t\}$, $\rho = \{q_{o,d}^p(t)|\forall\, o, d, p, t\}$ and $\lambda = \{l_{o,d}^{p,r}(t)|\forall\, o, d, p, r, t\}$. The MSA algorithm can be summarized as follows:

**Step 0:** *Initialization.* Initialize region speeds $v^0$, $\rho^0$ and $\lambda^0$. The initialization of $v^0$ can be based on free flow speeds; $v^0$ is the set of region speeds as if the regions are completely empty. $\rho^0$ and $\lambda^0$ are two sets where each element is zero, as they will be overwritten in the first iteration. Set $m$=1.

**Step 1:** *Direction finding.* Perform stochastic network loading based on region speeds $v^{m-1}$. This generates an auxiliary path flow $\rho^*$ and trip length set $\lambda^*$.

**Step 2:** *Calculating new assigned flow and trip length set.*

- For each $(o, d, p, t)$ quartet, set
$$[q_{o,d}^p(t)]^m = [q_{o,d}^p(t)]^{m-1} + (1/m)\left([q_{o,d}^p(t)]^* - [q_{o,d}^p(t)]^{m-1}\right).$$
- For each $(o, d, p, r, t)$ quintet, set
$$[l_{o,d}^{p,r}(t)]^m = [l_{o,d}^{p,r}(t)]^{m-1} + (1/m)\left([l_{o,d}^{p,r}(t)]^* - [l_{o,d}^{p,r}(t)]^{m-1}\right).$$

**Step 3:** *Update.* Implement MFD dynamics. $v^m = f\left(\rho^m, \lambda^m\right)$.

**Step 4:** *Stopping test.*

- Evaluate the following for each $(r, t)$ pair $\left[[v_r(t)]^m - [v_r(t)]^{m-1}\right] \geq \epsilon$.
- Calculate $M(\epsilon)$ the number of cases for which the above criterion is violated.
- If $M(\epsilon) \geq \phi$, set $m = m + 1$ and go to Step 1. Otherwise, terminate the procedure.

Note that the convergence of MSA is not monotonic. This is because of random search direction (auxiliary values produced by stochastic network loading may sometimes point in a direction where objective function increases) and the fixed move size (predetermined step size, $\alpha_m = 1/m$, may overshoot the reduction in the objective function, as it incorporates no information related to the optimal solution neighbourhood). In addition, one can claim that convergence criterion used in MSA is forced to converge due to the nature of step size sequence $\{\alpha_m\}$. However, practical experience indicates reasonable convergence speed and existence of stable solution, before it is forced by the sequence of step size. Regarding size, the number of variables (i.e. speed in each region at each time period) that have to converge is significantly low in this model compared to traditional traffic assignment models.

## 2.3 Model implementation

The network presented in Figure 2.3 is modeled in two distinct ways; as an aggregate model using multi-region MFD dynamics presented in Section 2.2.1 and as a detailed model in AIMSUN (Barceló and Casas, 2005). Regarding the aggregated traffic dynamics, the system is partitioned into four regions with different MFD functions, see Figure 2.3a. The model consists of two regional OD pairs, and assumes uniform distribution of origin and destination points across the regions. The network modeled in the micro-simulator (see Figure 2.3b) consists of one-way streets with one lane per link, 200 m block length, free flow speed of 50 km/h and 240 OD pairs. All intersections are signalized with a constant cycle length of 60s and pre-timed two phase sequence. The direction of travel is successively altered as Figure 2.3b shows. Aimsun utilizes a current best strategy with recalculation, i.e. based on current conditions drivers estimate shortest paths in the network based on their current position and the final destination every predefined time interval (e.g. 1min). In this study, the demand profile is trapezoidal, starting from low values and increasing gradually to the peak level, the demand is, then, gradually reduced in order to empty the network. The origins are distributed within the regions 1 and 2, while the destinations are distributed within region 4. We assume that there is no demand that originates from region 3, but many trips originated from other regions cross region 3.

The purpose of this section is three fold: to test the accuracy of the modified aggregated network dynamics, to analyze equilibrium state that results from the developed assignment model and to highlight the difference between equilibrium type and non-equilibrium type of assignment models. This section consists of three subsections. Section 2.3.1 investigates how the aggregated dynamic model developed in Section 2.2.1 can approximate well the dynamics of a detailed link based model with link description, route choice of individual vehicles for each origin-destination pair and intersection traffic control. The aggregated dynamic model is not aware of the above information at the link level, but requires as input only the MFD shape, the boundary capacity between regions and regional demand profile. The aggregated network dynamics have been modified in this study to consider multiple traveler groups that appear in the same urban region. Therefore, the accuracy test of the aggregated network dynamics is a necessary step before the testing of the developed DTA model. Section 2.3.2 presents the results from the DTA model developed in Section 2.2.2 and Section 2.2.3. Section 2.3.3 compares the proposed model with a simple one-shot assignment strategy, which employs instantaneous travel times through SNL procedure and which does not yield equilibrium conditions in the network. Note that Sections 2.3.2 and 2.3.3 do not incorporate micro-simulation results, they compare two different assignment approaches using the aggregated network dynamics.

Note that accumulation updates in the aggregated traffic model are performed continuously, while traffic assignment decisions (i.e. $q_{o,d}^{p}(t)$ and $l_{o,d}^{p,r}(t)$ ) are renewed every 1 min. These variables are assumed constant between any two renewal points.

### 2.3.1   Comparison of the aggregated dynamics with micro-simulation

This section introduces the comparison of the aggregated network dynamics with micro-simulation. Note that we do not employ the proposed assignment model to validate the traffic dynamics, we only test the accuracy of traffic model introduced in Section 2.2.1. The purpose of the analysis is to test the performance of aggregated network dynamics with and without the newly proposed dynamic average trip length, and its ability to predict traffic states in the network.

To conduct the comparison analysis, MFD functions, boundary capacity values and time-dependent regional route choice information (i.e. $q_{o,d}^{p}(t)$, $l_{o,d}^{p,r}(t)$) are extracted from the microscopic simulation environment using vehicle trajectories and virtual loop detectors. It is expected that recent advances in estimation methods with large data, can reproduce with decent accuracy these variables in a real environment (compared with the difficulty of link estimations, which is practically very difficult for large scale networks).

Figure 2.5a compares the total transfer flow (between region 2 and 4) that is extracted from the micro-simulator and that is calculated by the outflow (or trip completion) formula defined in the aggregated dynamics, see Section 2.2.1. The outflow formula is applied in two different ways; with the constant average trip length, $l_r$, $m_{o,d}^{p,r}(t) = (n_{o,d}^{p,r}(t))/(n_r(t) * (f_r(n_r(t)))/(l_r))$ and with dynamic average trip length defined for each group, $l_{o,d}^{p,r}(t)$ (please see Eq. 2.1). Note that aggregated network dynamic equations are not constantly employed in this step to avoid accumulating modeling error. Instead, time-dependent partial accumulation values, $n_{o,d}^{p,r}(t)$, taken from the micro-simulator are incorporated in trip completion formula at each time step. In other words, the state variables (i.e. accumulation) are corrected in each time step through a feedback mechanism that connects micro-simulation and the aggregated dynamics. Dynamic average trip length, $l_{o,d}^{p,r}(t)$, is extracted from the vehicle trajectories in the micro-simulator, while the constant average trip length, $l_r$, is the time-independent proportion of the production [veh.km/h] of the region to the outflow [veh/h] at the whole boundary. Note that there are two groups of travelers that use the boundary between region 2 and 4; $n_{2,4}^{[24],2}$ and $n_{1,4}^{[1324],2}$. Due to the network topology, the first group crosses a long distance in region 2 compared with the second group (see Figure 2.4). Figure 2.5a clearly shows that constant average trip length assumption is not accurate in case of groups of travelers with significantly different average trip lengths. On the other hand, the outflow calculation with the dynamic average trip length consideration is able to follow the trend of the detailed micro-simulation model. Note that dynamic average trip length, $l_{o,d}^{p,r}(t)$, is extracted from only 10% of vehicles in the network, which is very promising considering the applicability of the developed methodology. Figure 2.5b provides the partial accumulations and the average trip lengths for the two traveler groups that use the boundary between regions 2 and 4. Note that trip lengths for two groups of travelers are very different and the way the accumulations evolve is not quite alike. This implies that a simpler version of Eq. 2.4 which ignores the changes in average trip lengths or partial accumulations would not be able to accurately represent traffic dynamics. Further analysis with real data is needed to investigate average trip length

phenomenon. Nevertheless, this analysis provides a preliminary but convincing observation about simulation data.
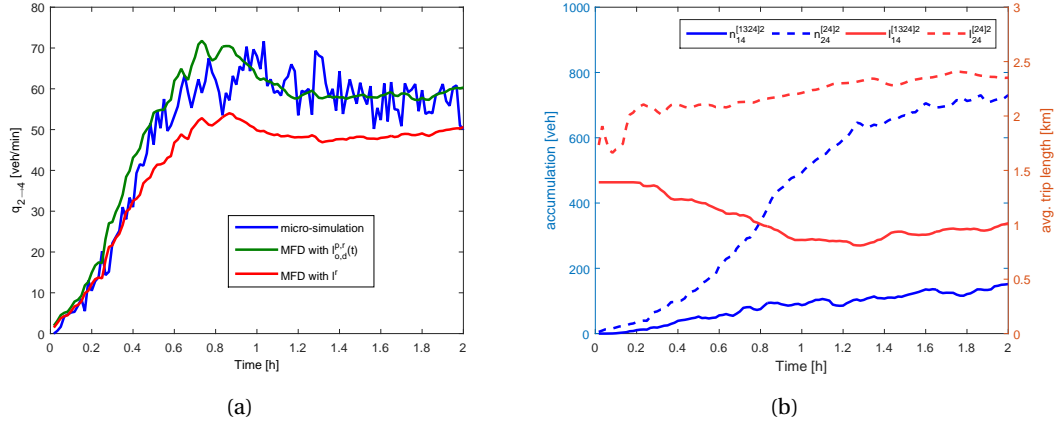


Figure 2.5: Simulation results (a) Transfer flow with different models, (b) partial accumulation and average trip length for different traveler groups

Another important question is how reliable the developed model is in estimating network traffic states. Every minute, partial accumulation ($n_{o,d}^{p,r}(t)$) and average trip length ($l_{o,d}^{p,r}(t)$) values are extracted from the micro-simulator and plugged in the aggregated dynamics (i.e. Eq 2.1-2.4). These dynamics are run along the length of the prediction horizon, and the accumulation values at the end of the prediction horizon are calculated. Figure 2.6 presents the region accumulations resulting from the micro-simulation (dotted curves) and from the MFD equations (solid curves) along with critical accumulations of MFD (dashed horizontal lines). The plots in Figure 2.6 depicts the prediction results for different horizon lengths. Note that all MFD-based accumulation values are estimated one prediction horizon ahead of time. Although the error expectedly increases with the increasing length of the prediction horizon, aggregated dynamics are able to follow the traffic evolution in the network in all the cases. Despite the different time and space resolution in two models, resulting traffic conditions in MFD and micro-simulation models are quite comparable with each other. Note that these results do not incorporate DTA model presented in this study, they reflect the results of the assignment strategy deployed in the micro-simulation. This analysis concludes that aggregated network dynamics in hand are able to reproduce network traffic state. It also shows that the developed modeling framework is a promising tool that can be integrated in an aggregated control framework to identify optimal regional route choices with a model predictive approach (see for example Haddad et al. (2013)). As micro-simulation contains a number of assumptions that might not represent reality, a field test will shed more light in the validity of such a modeling approach and the integration in large scale traffic control. This work is under development.
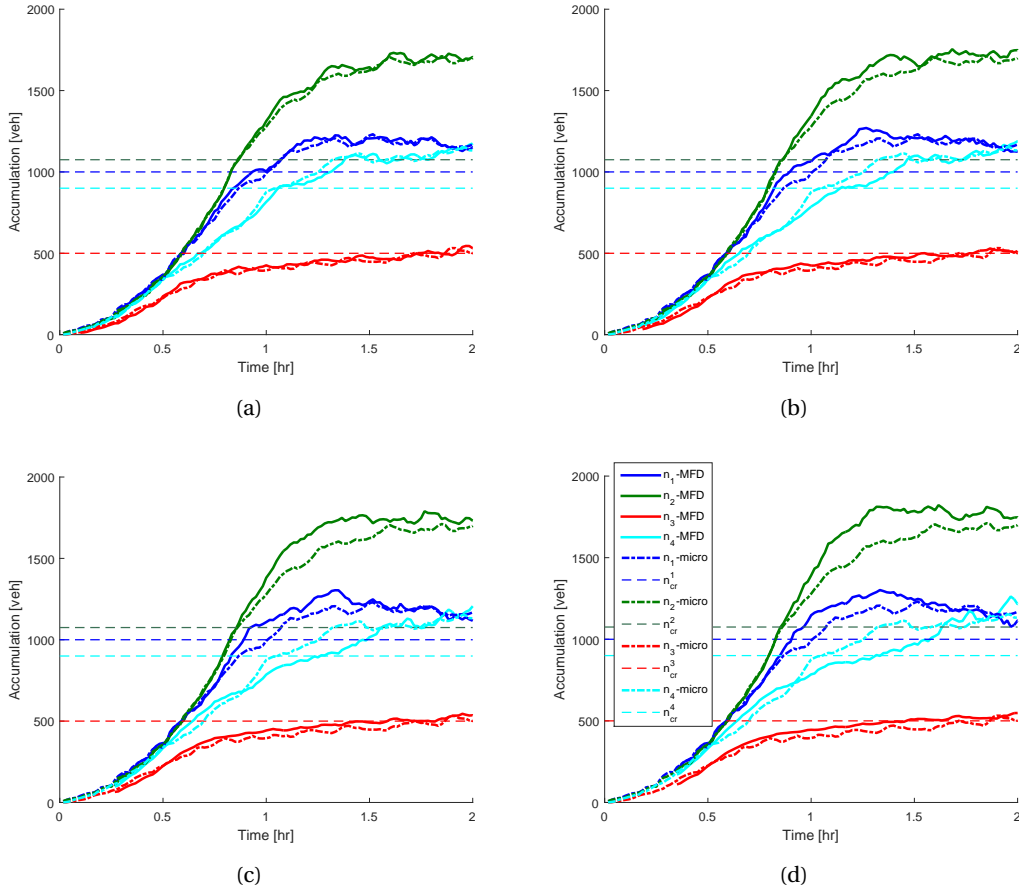
Figure 2.6: Accumulation of regions with prediction horizon of (a) 5 min, (b) 10 min, (c) 15 min, (d) 20 min

### 2.3.2 Assignment model with approximate equilibrium conditions

This section introduces the results of the DTA model developed in Sections 2.2.2 and 2.2.3. Note that, the traffic output (e.g. region accumulations and speeds) presented in the remainder of the chapter is produced by the aggregated network dynamics, not by the micro-simulator. However, MFD functions and boundary capacity values extracted from the micro-simulation stay in use, as they are valid network properties.

Figure 2.7 presents the evaluation of inconsistency (i.e. deviation between iterations) along the number of MSA evaluations. Figure 2.7(a) introduces L2-norm of the regional speed difference between current and previous iteration. Note that a predetermined number of SNL iterations (in this case 20) has been deployed in this analysis. A stationary value for the inconsistency has been reached in 15 iterations, and there is actually no need to further iterate the solution mechanism. However, the difference between current and previous iteration might go to zero due to the nature of MSA formula. To investigate this phenomenon, we have calculated the L1-norm of the difference between auxiliary (i.e. $\rho^*$ and $\lambda^*$) and preceding (i.e.

$\rho^{m-1}$ and $\lambda^{m-1}$) decision variables. As auxiliary variables are not subject to MSA formula, this analysis can indicate whether the convergence is forced by MSA or not. Figure 2.7(b) depicts the relative inconsistency between auxiliary and preceding values, i.e. $\frac{||\rho^*-\rho^{m-1}||_1+||\lambda^*-\lambda^{m-1}||_1}{||\rho^{m-1}||_1+||\lambda^{m-1}||_1}$. Similarly, inconsistency values converge approximately to 0.05 at around 15 iterations, which indicates 5% difference (in terms of L1-norm) between auxiliary and preceding directions. Consequently, the convergence in Figure 2.7(a) is strongly related to the one in Figure 2.7(b), and the algorithm does not stop due to decreasing effect of the step size (i.e. $1/m$). All tests have been performed with MATLAB R2013a on a standard Intel Core2 Quad 3.0 GHz personal computer, and each iteration (i.e. SNL and run of the 3-hour traffic simulation) takes about 45-50s. Note that a faster coding environment and a more powerful computer can easily decrease the computation time. In addition, the problem has been repeated with different initial points. We have observed that the number of iterations needed for the algorithm to converge is not very sensitive to the initial point, and the solution is unique in each case.
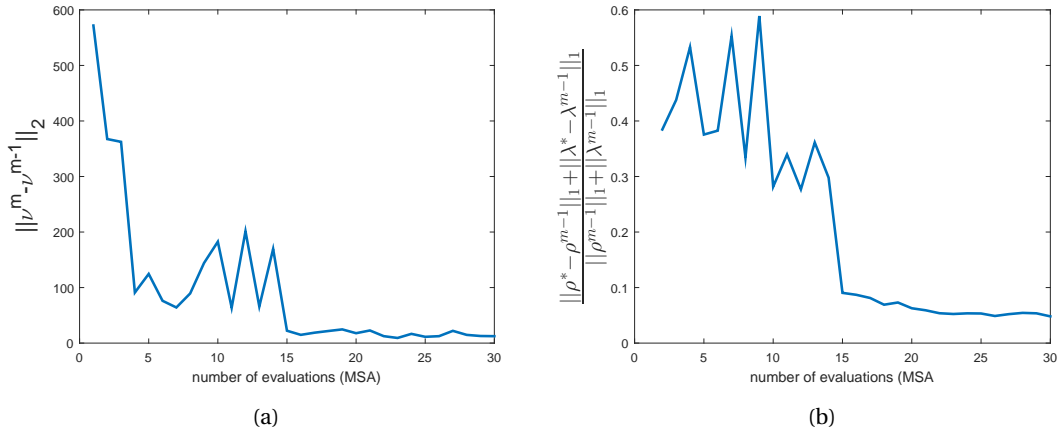


Figure 2.7: Convergence of (a) regional speeds, (b) assigned flow and average trip lengths.

Figure 2.8a and b display the evolution of speed and accumulation in the regions that results from the developed traffic assignment model. Total travel time in the network is $6.75 * 10^3$ veh.h (12.4 min/veh). Results show that assignment model presented in this chapter is able to distribute the congestion in an equal manner without details at individual link level. Although the accumulation level differs across the regions, speed values are very close to each other (see Figure 2.8a and b). Although DSUE conditions do not always yield similar traffic conditions throughout the network, in this specific example, we expect to observe similar conditions in the equilibrium state. Unless they by-pass relatively congested links considering the full speed information, vehicles in a grid network are intended to choose a path with the physical length equal to the manhattan distance between origin and destination points. Therefore, in the equilibrium state, where the actual travel times experienced by travelers of the same OD pair departing at the same time are equal and minimal, regions are supposed to have similar average speeds, as there are certain paths in use that cross them. Note that scale factor $\alpha$ presented in the logit formula at Section 2.2.2 has been arbitrarily chosen here in order to

satisfy a certain degree of homogeneity in the network.
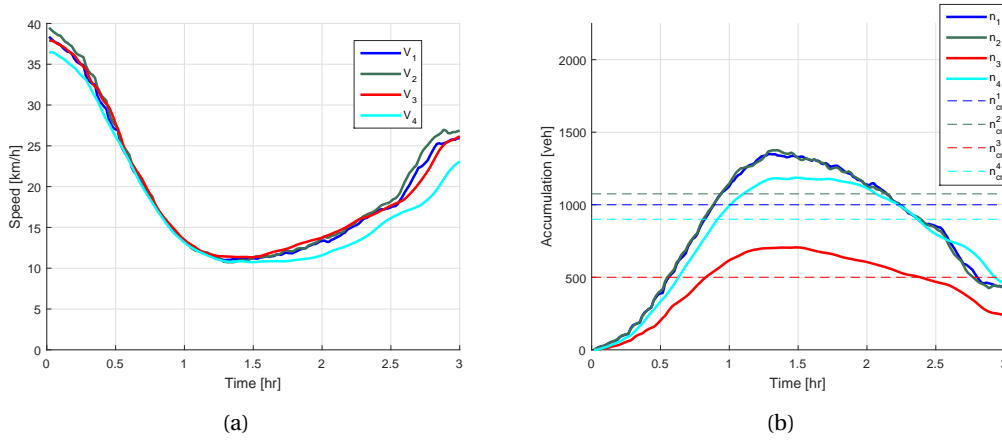


Figure 2.8: Results of the developed DTA model (a) speed evolution, (b) accumulation evolution

Figure 2.9 provides detailed results for the regional OD pair (24), see Figure 2.4a for regions. Figure 2.9a displays the route choice parameters; portion of the demand that chooses particular regional paths. Note that $\theta_{od}^p$ represents the portion of the demand that travels between region $o$ and $d$, using path $p$. In this context, path enumeration is as follows: path 1 is 2→4, path 2 is 2→3→4, and path 3 is 2→3→1→4 (see Figure 2.4a for regional path examples). Note that the demand is always assigned to one of these three paths, and the sum of route choice parameters is always equal to 1. Route choice parameters are not subject to oscillatory behavior throughout the simulation time, which indicates consistency with the neighboring time periods.

Figure 2.9b depicts the route choice parameters for the subset of the OD demand that has multiple options. Note that region 4 is divided into three sub-regions (see Figure 2.3b). As described in the previous sections, this study deploys a time-dependent shortest path algorithm on a graph where the link speeds are represented by the average speed of the region. Therefore, it is very likely that the shortest path between an origin and destination point has a physical length which is equal to the manhattan distance between the same points. This behavior also affects the route choice decision; travelers have different number of alternative regional paths depending on the exact location of their destination point in region 4. For example, people who travel between region 2 and 4c have no alternative. The only possible regional path that gives manhattan distance solution is 2→4. Travelers between region 2 and 4b have two alternatives; 2→4 and 2→3→4. Likewise, travelers between region 2 and 4a have three alternatives; 2→4, 2→3→4 and 2→3→1→4. Figure 2.9b presents the portion of the demand that chooses a particular path among a set of alternatives. In other words, $\theta_{od}^{p*}$ represents the portion which is not forced to use path $p$, but prefers path $p$ over other alternatives. Recalculation of route choice parameters allows us to see the actual dynamic

change in path decisions throughout the simulation time, especially for the demand assigned to the third path.
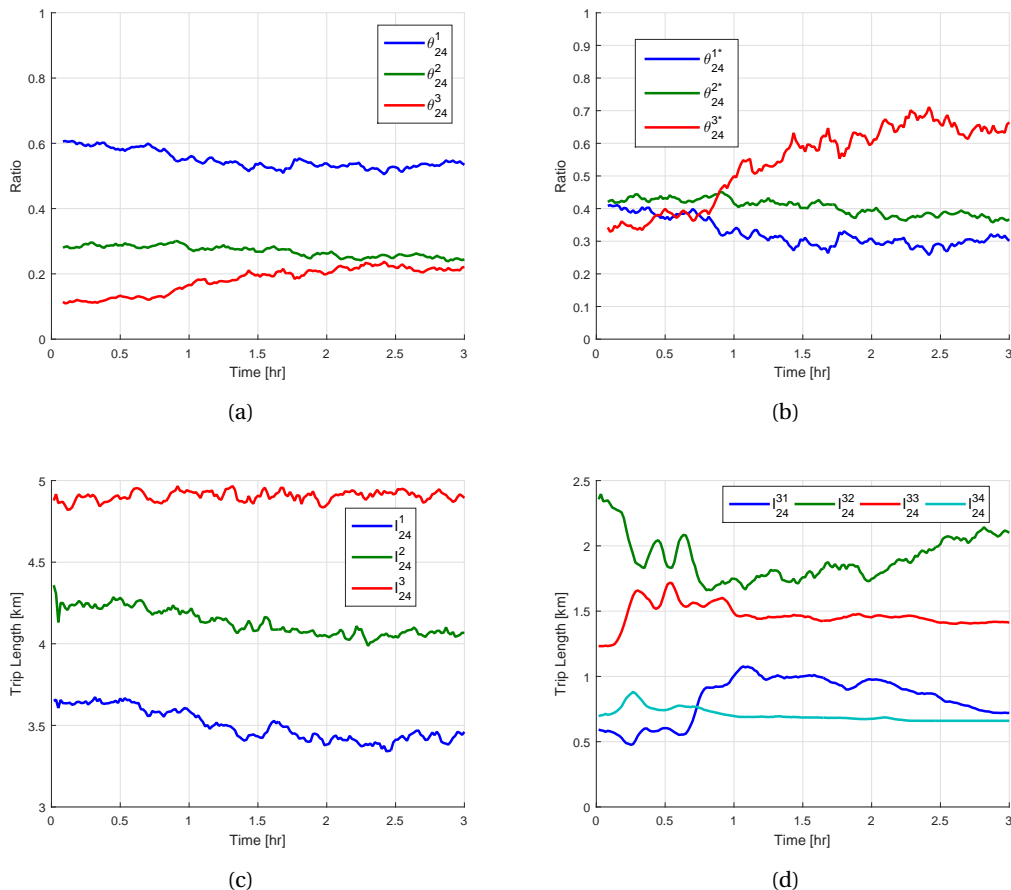


Figure 2.9: Results for OD pair (24) (a) route choice parameters, (b) route choice parameters for part of the demand with multiple path options, (c) average trip lengths throughout the path, (d) partial trip lengths on a particular path

Figure 2.9c displays total average trip length on different paths. Note that $l_{od}^p$ represents the total average trip length for vehicles that travel between region $o$ and $d$, using path $p$. Because of the underlying route choice phenomenon explained in the previous paragraph, the order of magnitude of the trip lengths is different; third path has always the longest average trip length, and first path has always the smallest. The total average trip length for the first and second path slightly goes down with time because of the shift of demand to third path. In other words, an increasing number of travelers prefer to use third path to reach their destination in 4a, this leads to a decreasing portion of travelers that have relatively longer trip length among the vehicles that choose first and second path. Therefore, the total average trip length decreases within this group. Figure 2.9d displays partial average trip length across different regions on the third path; 2→3→1→4. Note that $l_{od}^{pr}$ represents the total average trip length crossed in

region $r$ by the vehicles that travel between region $o$ and $d$, using path $p$. The evolution of partial trip lengths shows a quite dynamic behavior, which means travelers change the trip length they travel in different regions according to the traffic conditions in the network. For instance, Figure 2.8a shows that region 2 has slightly higher speed values in the beginning and end of the simulation compared to other regions. The difference in speeds leads to an increase in the trip length travelled in region 2 in the start and end of the simulation compared to other time periods (see Figure 2.9d). In addition, region 4 has always the lowest speed values throughout the simulation time, this results in trip length in region 4 almost always at its lower bound (see Figure 2.9d).

### 2.3.3 One-shot traffic assignment with the aggregate dynamics

This section presents a one-shot assignment model (as defined in Chiu et al. (2011)) in which departing vehicles are given a route periodically updated based on instantaneous travel times (measured at the time vehicles are generated). This assignment model which does not result in equilibrium highlights the added value (in terms of delays and system performance) of DTA models. This approach periodically updates the instantaneous travel times in the network and assigns paths to the newly generated vehicles according to them. As the choices are based on myopic traffic conditions, the network congestion may be overstated.

In the framework of aggregated dynamics, one-shot assignment model can be implemented through SNL. As the equilibrium conditions are not satisfied, MSA is removed from the procedure and SNL is implemented in each time step with the updated traffic conditions. In other words, the aggregate traffic parameters (i.e. $q_{o,d}^{p}$ and $l_{o,d}^{p,r}$) are computed for the following time step through SNL, which employs prevailing traffic conditions (instantaneous travel times) to determine the paths of the newly generated vehicles through the logit model. Then, aggregate dynamics are employed for one time step to generate traffic conditions in the network, which are again fed to SNL to do the assignment in the next time step. This procedure is repeated until the end of the period. Figure 2.10 presents the resulting traffic conditions from the one-shot traffic assignment model. As the equilibrium state is missing, the network is clearly more congested; total travel time in the network is $1.39 * 10^4$ veh.h (21.9 min/veh) which is more than double the travel time with equilibrium conditions. Myopic route choice decisions lead the network to a congested state (close-to-gridlock) at the end of the analysis. Integrating real-time information to provide regional route guidance to users can provide significant improvement in active traffic management schemes. This should be a research priority.

## 2.4 Summary

In this study, we develop a DTA model that is integrated with MFD dynamics in order to establish DSUE conditions. The approach presented in this chapter has two main components; SNL and MSA. The first part addresses time-dependent trip lengths within and between regions, randomness of the system and perception errors which may become significant in case of
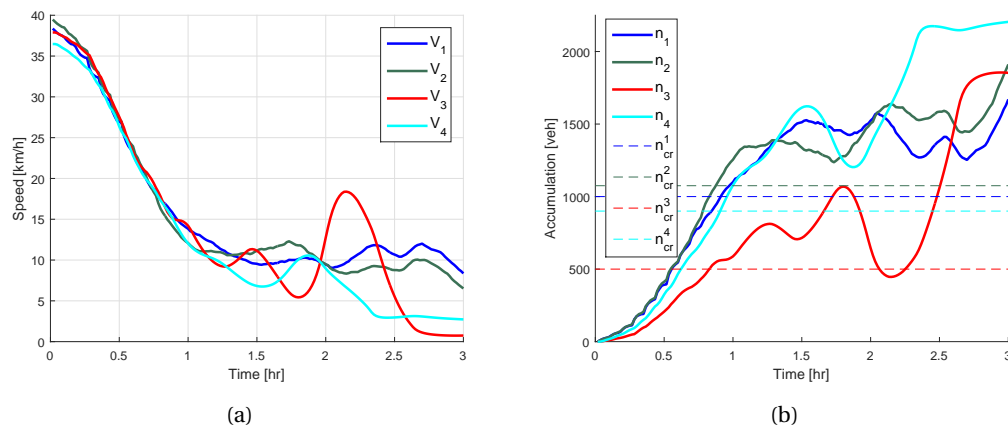
Figure 2.10: Results of the one-shot assignment model (a) speed evolution, (b) accumulation evolution

the parsimonious model in hand. The second part has been deployed to average auxiliary directions taken from SNL and to establish equilibrium conditions in an iterative manner. Although MSA applications suffer from real-time feasibility issues in case of detailed traffic modeling, this approach, relying on the aggregated dynamics defined by MFD relationship, has exhibited reasonable convergence speeds.

Considering both DSUE conditions and system performance, DTA model developed in this study outperforms the other assignment models which do not result in equilibrium. Network congestion is likely to be overstated in non-equilibrium type of assignment models, while the developed model leads the network to a more equitable state. Aggregated network dynamics do not provide a detailed modeling opportunity, but it allows us to consider the evolution of network traffic state. Since it is very fast, aggregated dynamics can be iteratively deployed to compute experienced travel time and identify the interaction of route decisions in the following time periods. A validation of the developed modeling with data from field experiments is a necessary next step. Estimation procedures for aggregated variables utilized in the model with different quality and size of data (type of sensors, penetration rates, etc.) is also required.

MFD-based DTA model that we have developed in this study could also be further exploited as a novel technique for transportation planning. State-of-the-art transportation planning tools make use of link-level static traffic flow models. At the large-scale, these models produce excessive amount of (link-level) traffic information, which is neither highly useful nor accurate regarding the planning purposes. However, the model we have proposed can address within-day traffic dynamics, and produce average traffic conditions expected in certain neighborhoods. While critical links (e.g. freeways, major arterials) could be represented with link-based models and jointly simulated with MFD-based urban areas, the proposed model might bring the advantages of a parsimonious approach, low computation effort and incorporation of within-day traffic dynamics into the area of transportation planning.

## Chapter 2. Approximating dynamic equilibrium conditions with macroscopic fundamental diagrams

Although a medium scale network has been employed in this study, the proposed methodology could cope with larger networks and more regions. As MFD allows us to define traffic flow in the regional context with few differential equations, implementation of traffic dynamics will still be very fast in case of more regions and multi-directional flow. On the other hand, the experienced shortest path algorithm deployed in this study is under the direct influence of network size (e.g. number of links, number of time periods). In addition, as it gives all-to-one shortest path solutions, the algorithm has to be repeated with respect to the number of destination points to be considered. The part which seems to be the computational burden of the proposed methodology is, therefore, the sampling procedure, especially sampling of destination points. One can identify the most frequent destination points using historical data or other sources, and incorporate them in the methodology. Despite the additional error that it would bring, this might significantly improve the performance of the overall algorithm.

Chapter 3 extends this study to a route guidance strategy, where advantages of MFD dynamics and conclusions from this study can be utilized in conjunction. The route guidance system can impact the travelers' route decision by providing them with useful information regarding the traffic states of the urban regions. Therefore, drivers can follow a series of regions that has lower cost (in terms of time, fuel consumption, etc.) for individuals, which might outcome to a better overall system performance.

# 3 Equilibrium analysis and route guidance in large-scale networks with MFD dynamics

RECENT studies have demonstrated that Macroscopic Fundamental Diagram (MFD), which provides an aggregated model of urban traffic dynamics linking network production and density, offers a new generation of real-time traffic management strategies to improve the network performance. An alternative management technique that is not explored yet is a route guidance scheme based on MFD modeling. The chapter advances in this direction by firstly extending two MFD-based traffic models with different granularity of vehicle accumulation state and route choice behavior aggregation. This configuration enables us to address limited traffic state observability and scrutinize implications of drivers' route choice in MFD modeling. We consider a city that is partitioned in a small number of large-size regions (aggregated model) where each region consists of medium-size subregions (more detailed model) exhibiting a well-defined MFD. This chapter proposes a route guidance advisory control system based on the aggregated model as a large-scale traffic management strategy, while subregional information is partially known. In addition, we investigate the effect of equilibrium conditions (i.e. user equilibrium and system optimum) on the overall network performance, in particular MFD functions.

## 3.1 Introduction

Research on congestion propagation in large urban networks has been mainly based on micro-simulations where traffic dynamics are defined at the link level. However, because of unpredictability of travelers' behavior and high complexity of traffic physical modeling, microscopic simulation models are time consuming and case dependent. In addition, traffic congestion management in large-scale urban systems is currently fragmented and uncoordinated with respect to optimizing the goals of travel efficiency and equity for multiple regions of a city. An alternative, which is investigated in this chapter, is a regional route guidance strategy, where a network is partitioned into homogeneous regions and drivers are given a regional path to follow (e.g. going through the city center or using the longer route at the periphery). On the other hand, the effect of driver adaptation models on the aggregate network performance is not fully explored. Microscopic simulation models depend on enroute decision

mechanisms, where drivers update their routes based on instantaneous travel times at the network. This behavioral assumption may cause uneven distribution of congestion in the network and reduce the network flow especially in the unloading phase where certain parts of the network are already more congested than others. However, under equilibrium conditions drivers are expected to make decisions based on experienced travel costs, which may lead to a more homogenous network in both loading and unloading phases. Understanding these interactions for heterogeneously congested cities is a big challenge, which will allow revisiting, redesigning, and integrating smarter traffic management approaches to alleviate congestion with a "system of systems" approach.

The traffic modeling in this chapter is based on the network macroscopic fundamental diagram (MFD) that provides a unimodal, low-scatter, and demand-insensitive relationship between network vehicle density and space-mean flow in homogeneous urban areas (with small spatial link density heterogeneity) Geroliminis and Daganzo (2008). The idea of an MFD with an optimum accumulation belongs to Godfrey (1969) and similar approaches were introduced later by Herman and Prigogine (1979), Mahmassani et al. (1984) and Daganzo (2007). However, urban transportation networks exhibit uneven distribution of congestion which leads to a scattered flow-density relationship. Heterogeneity in congestion distribution can affect the shape/scatter or even the existence of MFD Buisson and Ladier (2009); Geroliminis and Sun (2011b). By using a grid network and considering variance of link density (over space) as an independent variable, Mazloumian et al. (2010) shows that MFD remains well-defined in subregions of the urban network. These results are very critical, because MFD concept can be useful for heterogeneously loaded cities, if the network can be partitioned into smaller homogenous areas. The effect of heterogeneity has been recently studied by many researchers with similar conclusions, see for example Daganzo et al. (2011); Mahmassani et al. (2013c); Geroliminis and Sun (2011b); Doig et al. (2013) and others. Ji and Geroliminis (2012) develops a static partitioning mechanism to minimize the variance of link densities while maintaining a spatially compact shape. Resulting subregions can be used to develop macroscopic traffic control strategies. In case of hierarchical networks with respect to topology, e.g. mixed freeway/arterial networks, hybrid models might be utilized. For example, Haddad et al. (2013) models the urban part of a city with multi-region MFDs, while the freeway network is represented by the Cell Transmission Model (Daganzo, 1994). Transfer flows between the two models guarantee consistency and conservation of flows. More complicated network structures with strong directional flows, limited connectivity, and high variability in trip lengths and connection with MFD modeling and clustering should be further investigated. This is work in progress.

Mahmassani et al. (2013c) and Gayah and Daganzo (2011) investigate the effect of driver adaptation on the shape of MFD through microscopic simulation for different network sizes. They both identify that an increase in driver adaptivity through enroute decision models can influence and shrink the size of hysteresis loops in the network MFD. In addition, Knoop et al. (2012) has developed myopic local re-routing strategies considering the aggregated information from multiple sub-networks. This strategy does not consider the interaction

of the vehicles in the following time periods. Leclercq and Geroliminis (2013) investigates approximations of user equilibrium (UE) and system optimum (SO) conditions of a simple network with MFD dynamics and two parallel routes with instantaneous travel times. Their findings reveal that SO improves system performance compared to UE mainly when initial network conditions are not in the congested regime of the MFDs. Nevertheless, the effect of equilibrium flows in real-sized networks has not been tested with MFD dynamics. In addition, stability analysis for MFD dynamics and control has been investigated in Haddad and Geroliminis (2012) and Haddad and Shraiber (2014).

Real-time large-scale traffic management strategies, e.g. multi-region perimeter control (Geroliminis et al., 2013; Haddad et al., 2013; Aboudolas and Geroliminis, 2013), gating (Daganzo, 2007; Keyvan-Ekbatani et al., 2012, 2015)) that benefit from parsimonious models with aggregated network dynamics, provide promising results towards a new generation of smart hierarchical strategies. On the other hand, the estimation of network traffic states for MFD analysis with different types of sensors identifies the applicability of MFD in large scale networks even if limited data exist, see for example Ortigosa et al. (2013b); Gayah and Dixit (2013b); Nagle and Gayah (2013); Leclercq et al. (2014); Ji et al. (2014). Furthermore, a connection of travel time reliability with network heterogeneity based on MFD concepts have been investigated by Gayah et al. (2013a); Mahmassani et al. (2013a); Yildirimoglu et al. (2015a).

The primary motivation of the chapter is to develop a network-level traffic management scheme to mitigate congestion in urban areas by considering the effect of route choice at an aggregated level. The management scheme is developed based on MFD and consists of a route guidance system that advises drivers a sequence of subregions to assist them in reaching their destination. This study extends the work in Yildirimoglu and Geroliminis (2014) to a route guidance system based on SO conditions. It is worth mentioning that we aim at network-level management strategies, thus the route guidance operates on subregional basis as opposed to conventional link-based route guidance, see for example Papageorgiou (1990), Ben-Akiva et al. (1997), Zhou et al. (2008), Ben-Elia et al. (2013) and Zhu and Ukkusuri (2013). The route guidance system can impact the travelers' route decision by providing them with useful information regarding the traffic states of the urban regions. Therefore, drivers can follow a series of subregions that has lower cost (in terms of travel time, fuel consumption, etc.), which might lead to a better overall system performance. The second motivation is to investigate the impact of driver adaptivity on the overall network performance, in particular MFDs. Most of the previous MFD estimations in the literature are based on one-shot simulations where driver adaptivity is incorporated by en-route decision mechanisms (e.g. current best or myopic local re-routing). This study tests the effect of dynamic UE (or DUE) and dynamic DSO (or DSO) flows in the network on the observed MFD functions and the existence of hysteresis loops. This chapter is the first attempt to integrate equilibrium flow conditions in the network MFD analysis, while different approaches have been utilized in Mahmassani et al. (2013c) (micro-simulation and instantaneous traffic conditions as opposed to equilibrium conditions), Gayah and Daganzo (2011) (a toy network with two rings), and Leclercq and Geroliminis (2013)

(n-parallel routes).

We investigate the problem where a heterogeneous city, in terms of link density, consists of several smaller and more homogenous subregions, see Figure 3.1, where each subregion (1–19) can be represented by a well-defined low scatter MFD. In result, the MFD of (larger) regions might exhibit hysteresis (due to heterogeneity) and have significant level of scatter in the congested regime. The partitioning of the city network to homogenous subregions enables us to model and study the effect of link density heterogeneity on the MFD characteristics, and also results in a configuration that obliges us to integrate a route choice model into the MFD traffic flow modeling. (Larger) regions should have a comparable size with the average trip length of the vehicles to allow alternative (subregional) route options. The modeling contribution of the chapter is to address the dynamic UE and SO route choice within the MFD framework, which consists of a detailed modeling of traffic flow for each subregion. We introduce traffic dynamics at the subregional (detailed) and regional (parsimonious) level. While the detailed representation provides more information on network traffic state, some variables of this model (e.g. route choice decisions) might not be fully observable. Thus, the parsimonious model is essential for management (i.e. route guidance) purposes. Compared with previous works, the introduced regional dynamics is the generalized multi-region model (instead of 2-region model), incorporated with heterogeneity modeling (instead of assuming homogenous conditions), and with integrated route choice modeling (instead of no regional route choice). Additionally, this integration of 2 different scales of modeling unveils interesting regularities with respect to the trip length of drivers under different strategies. Note that the intension of this work is not to develop traffic management strategies for networks partitioned in too many regions. In this case, dynamic origin-destination matrices and other detailed information would be very difficult to be estimated with small errors. The subregional model can be considered as a more detailed (but not well-observed with data) representation (in control theory this is called the "plant"), while the regional model (3 regions in Figure 3.1) is analyzed for management purposes ("model"). We apply the current analytical plant instead of a detailed micro-simulation, as this chapter chooses a more methodological path, which allows us to create further insights of the dynamics of heterogeneity and route guidance strategy.

The remainder of the chapter is organized as follows; in Section 3.2, we introduce the dynamics of the region-based model (optimization model) and the subregion-based model (plant). In Section 3.3, the methodological frameworks for dynamic traffic assignment (DTA) and route guidance (RG) system are elaborated. Section 3.4 presents and discusses results of a case study with different demand levels. Finally, Section 3.5 concludes the chapter with future work directions.
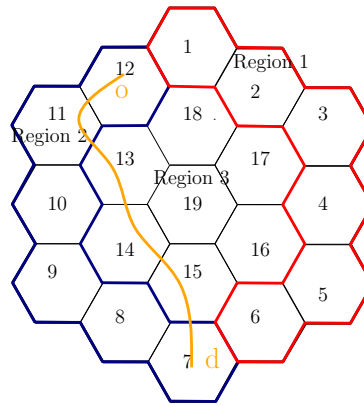
Figure 3.1: The schematic of a multi-region urban network and a path example. The network consists of 19 subregions that are aggregated in 3 larger regions (Region 1 contains subregions 1 to 6, Region 2 contains 7 to 12, and Region 3 contains 13 to 19).

## 3.2 Traffic models

In this study, we introduce two macroscopic traffic models: (i) a region-based model considering that the urban network is partitioned into a small number of regions, and (ii) a subregion-based model deploying a more detailed approach where the above regions are divided into smaller subregions. Two levels of modeling enables us to account for different layers of traffic state measurement and control and to incorporate heterogeneity effect in the urban network dynamics. The region-based model serves as the operation model that the management strategy (e.g. route guidance) is based on, while subregion-based model serves as the plant (reality) dynamics, which represents the ground truth traffic conditions. Region-based model integrates heterogeneity dynamics in two ways: (a) a time-varying average regional trip length and (b) an MFD depending on regional accumulation and the heterogeneity of the spatial distribution of congestion. On the other hand, the subregion-based model describes the evolution of traffic based on the MFD considering constant subregional average trip length, subregional paths, and the boundary flow capacity between subregions. We show in the chapter that the two models are consistent and that the less detailed model can integrate variable trip lengths, hysteresis loops and spatial heterogeneity in the distribution of congestion.

This configuration with two levels of traffic aggregation, i.e. regional and subregional, allows us to develop more realistic control schemes such that the partial observability of traffic states and control tractability can be investigated. While time-dependent trip lengths have been considered in Ramezani et al. (2015) and Yildirimoglu and Geroliminis (2014), there are two significant modeling extensions in this work: (i) vehicles are allowed to cross the boundary between the regions more than one time (e.g. a trip with the sequence 11-13-10-14-8-7 is allowed) and (ii) equilibrium conditions have been satisfied in both models.

### 3.2.1 Region-based model

Let us assume that the urban network is partitioned into several subregions with well-defined MFDs. We consider a region as a set of multiple adjacent subregions, see Figure 3.1, forming $R$ regions, $\mathcal{R} = \{1, 2, \cdots, R\}$. Let $Q_{IJ}(t)$ [veh/s] denote the exogenous traffic flow demand generated in region $I$ with *final* destination region $J$, $N_{IJ}(t)$ [veh] be the vehicle accumulation in region $I$ with final destination region $J$, and $N_I(t)$ [veh] be the total accumulation in region $I$ at continuous time $t$; $I, J \in \mathcal{R}$; $N_I(t) = \sum_{J \in \mathcal{R}} N_{IJ}(t)$. Note that time $t$ is defined as a continuous variable.

The traffic flow conservation equations of an $R$-region MFDs system are as follows:

$$\frac{\mathrm{d}N_{II}(t)}{\mathrm{d}t} = Q_{II}(t) - M_{II}^I(t) - \sum_{H \in \mathcal{V}_I} \hat{M}_{II}^H(t) + \sum_{H \in \mathcal{V}_I} \hat{M}_{HI}^I(t) \tag{3.1}$$

$$\frac{\mathrm{d}N_{IJ}(t)}{\mathrm{d}t} = Q_{IJ}(t) - \sum_{H \in \mathcal{V}_I} \hat{M}_{IJ}^H(t) + \sum_{H \in \mathcal{V}_I} \hat{M}_{HJ}^I(t) \quad I \neq J \tag{3.2}$$

where $\mathcal{V}_I$ is the set of regions that are directly reachable from region $I$, i.e. adjacent regions to region $I$. $M_{II}^I(t)$ [veh/s] denotes the *internal trip completion rate* for accumulation in $I$ with destination $I$ (without going through another region), while the *transfer flow* for accumulation in $I$ with final destination $J$ through the next immediate region $H$ is denoted by $\hat{M}_{IJ}^H(t)$ [veh/s] that considers the receiving flow capacity on the transfer flow from region $I$ to region $H$. Note that, Eq. 3.1 and 3.2 permit the vehicle paths to include more than one crossing over the boundaries between the regions. For example a trip through subregions 11-13-19-15-7 (see Figure 3.1), starts and ends in region 2 while crossing region 3.

Internal trip completion rates and transfer flows are estimated corresponding to the ratio between accumulations as:

$$M_{II}^I(t) = \theta_{II}^I(t) \cdot \frac{N_{II}(t)}{N_I(t)} \cdot \frac{F_I\big(N_I(t), \sigma(N_I(t))\big)}{L_{II}(t)}, \tag{3.3}$$

$$M_{IJ}^H(t) = \theta_{IJ}^H(t) \cdot \frac{N_{IJ}(t)}{N_I(t)} \cdot \frac{F_I\big(N_I(t), \sigma(N_I(t))\big)}{L_{IH}(t)}, \tag{3.4}$$

$$\hat{M}_{IJ}^H(t) = \min\left[ M_{IJ}^H(t), C_{IH}\big(N_H(t)\big) \cdot \frac{N_{IJ}(t) \cdot \theta_{IJ}^H(t)}{\sum_{K \in \mathcal{R}} \big(N_{IK}(t) \cdot \theta_{IK}^H(t)\big)} \right], \tag{3.5}$$

where $F_I(\cdot)$ [veh.m/s] is the production MFD of region $I$ (i.e. the total distance travelled per unit time in the region) that is a function of the region accumulation, $N_I(t)$, and the link density *heterogeneity* across all region $I$ links, $\sigma(N_I(t))$ [1]. In this work, heterogeneity term, i.e. $\sigma(N_I(t))$, represents the accumulation variance across all subregions in region $I$. Moreover, $L_{II}(t)$ [m] is the average trip length corresponding to internal trips in region $I$, $L_{IH}(t)$ [m] is the average trip length corresponding to transfer trips from region $I$ to its neighbor region $H \in \mathcal{V}_I$,

---

[1] Heterogeneity modeling presented in Ramezani et al. (2015) allows us to consider the effect of density variability on the region production. Please refer to the paper for detailed information.

and $\theta_{IJ}^H(t)$ is the percentage of outflow from region $I$ to final destination $J$ through the next immediate region $H$; hence $\sum_{H \in \mathcal{V}_I} \theta_{IJ}^H(t) = 1$. [2] Modeling the region link density heterogeneity, i.e. $\sigma(N_I(t))$, is investigated in Ramezani et al. (2015); while estimation of traffic states $L_{IJ}^H(t)$ and $\theta_{IJ}^H(t)$ is discussed in the next subsection. [3]

Eq. 3.5 estimates the transfer flow from region $I$ to the next region $H$ that is the minimum of two terms: (i) the sending flow from region $I$ that solely depends on the region $I$ accumulation and (ii) the product of the boundary capacity between regions $I$ and $H$ that is a function of region $H$ accumulation, i.e. $C_{IH}(N_H(t))$, and the proportion of vehicles in $I$ with final destination $J$ through the next region $H$, i.e. $N_{IJ}(t) \cdot \theta_{IJ}^H(t)$, among all the vehicles that will cross the same boundary, i.e. $\sum_{K \in \mathcal{R}} (N_{IK}(t) \cdot \theta_{IK}^H(t))$.

Eq. 3.1-3.5 represent the traffic dynamics in an $R$-region urban network considering the heterogeneity effect and integrating an aggregated route choice model, which is further exploited in the route guidance system. The model assumes that drivers can choose any arbitrary sequence of regions as their path and cross region boundaries without any limitation. While in previous publications related to urban traffic control (e.g. Geroliminis et al. (2013); Haddad et al. (2013)) region receiving capacity was omitted as perimeter control was operating in the boundary, in this modelling formulation the region receiving capacity is present in Eq. 3.1 and 3.2. The reason is that no control situations might significantly increase the accumulation of receiving regions and in this way the inter-transfer flows from sending regions are constrained. Note that the perimeter control variables are not included in the model as the control decisions here are $L_{IJ}^H(t)$ and $\theta_{IJ}^H(t)$, $I, J, H \in \mathcal{R}$, computed by the advisory route guidance system.

### 3.2.2 Subregion-based model

[4] Consider subregion $r$ with homogeneous distribution of congestion whose traffic performance is well described by MFD, $f_r(n_r(t))$, representing the subregion production [veh.m/s] corresponding to the accumulation $n_r(t)$ [veh] at continuous time $t$. The average subregion $r$ speed is $v_r(t) = f_r(n_r(t))/n_r(t)$ [m/s], and trip completion rate is $m_r(t) = f_r(n_r(t))/l_r$ [veh/s], considering a constant subregional average trip length $l_r$ [m] independent of destination or next region.

---

[2]Note that $L_{IH}(t)$ is not the full trip length of vehicles, but the part of the trip in region $I$. Thus, trip 11-13-19-5-7 in Figure 1, will first contribute to $L_{23}(t)$ and then to the $L_{32}(t)$ and finally to $L_{22}(t)$. In a similar manner this trip will be part of $N_{23}(t)$ accumulation while in region 2, then part of $N_{32}(t)$ accumulation while in region 3 and finally part of $N_{22}(t)$ accumulation while returning to region 2 and until it finishes its trip.

[3]If vehicles are only allowed to cross the boundary between the regions at most one time (in the region-based model) then Eq. 3.1 and 3.2 are simplified to $\mathrm{d}N_{II}(t)/\mathrm{d}t = Q_{II}(t) - M_{II}(t) + \sum_{H \in \mathcal{V}_I} \hat{M}_{HI}^I(t)$ and $\mathrm{d}N_{IJ}(t)/\mathrm{d}t = Q_{IJ}(t) - \sum_{H \in \mathcal{V}_I} \hat{M}_{IJ}^H(t), I \neq J, J = H$ as described in Ramezani et al. (2015), while $\theta_{II}^I(t), \forall I \in \mathcal{R}$ and $\theta_{IJ}^H(t), J = H$ in Eqs. 3.3 and 3.4 are equal to 1.

[4]The model presented here is very similar to the one in Section 2.2.1. The main difference is the assumption of constant average trip length in the new model, which has implications in Eq. 3.6 and 3.9. While the reader can skip the rest, Section 3.2.2 is preserved in order to satisfy completeness of the chapter.

## Chapter 3. Equilibrium analysis and route guidance in large-scale networks with MFD dynamics

Let $n_{o,d}^{p,r}(t)$ denote the number of vehicles in subregion $r$ at continuous time $t$ with origin subregion $o$, destination subregion $d$, and path $p$, i.e. the set containing the sequence of subregions to reach $d$ starting from $o$; note that $r$ belongs to $p$. Consequently $\sum_o \sum_d \sum_p n_{o,d}^{p,r}(t) = n_r(t)$.[5] Trip completion rate $m_{o,d}^{p,r}(t)$ for the vehicles in region $r$ at time $t$ with origin $o$, destination $d$, and path $p$ reads:

$$m_{o,d}^{p,r}(t) = \frac{n_{o,d}^{p,r}(t)}{n_r(t)} \cdot m_r(t) = \frac{f_r(n_r(t))}{n_r(t)} \cdot \frac{n_{o,d}^{p,r}(t)}{l_r} = v_r(t) \cdot \frac{n_{o,d}^{p,r}(t)}{l_r}. \tag{3.6}$$

Let $q_{o,d}(t)$ denote the exogenous demand generated at continuous time $t$, from origin $o$ to destination $d$, and $q_{o,d}^p(t)$ be the assigned demand to path $p$; $\sum_p q_{o,d}^p(t) = q_{o,d}(t)$. Let $\hat{m}_{o,d}^{r \to p^+(r)}(t)$ be the transferring flow from subregion $r$ to subregion $p^+(r)$, which is the next subregion in the sequence described by path $p$. Similarly, $p^-(r)$ is the previous subregion before $r$ in path $p$. The subregion traffic dynamics are as follows: (Note time $t$ is omitted from the equations for the sake of notation simplicity.)

$$\frac{dn_{o,d}^{p,r}}{dt} = \begin{cases} q_{o,d}^p - m_{o,d}^{p,r} & \text{(i) if } r = o \ \& \ r = d, \\ q_{o,d}^p - \hat{m}_{o,d}^{r \to p^+(r)} & \text{(ii) if } r = o \ \& \ r \neq d, \\ \hat{m}_{o,d}^{p^-(r) \to r} - m_{o,d}^{p,r} & \text{(iii) if } r \neq o \ \& \ r = d, \\ \hat{m}_{o,d}^{p^-(r) \to r} - \hat{m}_{o,d}^{r \to p^+(r)} & \text{(iv) otherwise.} \end{cases} \tag{3.7}$$

where

$$\hat{m}_{o,d}^{r \to p^+(r)} = \min[m_{o,d}^{p,r}, c_r^{p^+(r)}(n_{p^+(r)}) \cdot a_{o,d}^{r \to p^+(r)}]. \quad \forall r \neq d \tag{3.8}$$

Eq. 3.7 defines the rate of change in accumulation $n_{o,d}^{p,r}$ such that in case of (i) internal demand within the same subregion, the rate is simply the exogenous demand minus the trip completion rate which is not bounded by any capacity function. Note that the subregion-based model assumes that internal subregional demand never leaves the subregion; therefore, in this case the subregional path $p$ consists of only one subregion. In case (ii) current subregion $r$ is the origin and not the destination, then the rate is the exogenous demand minus the transfer flow to the next subregion in path $p$. In case (iii) current subregion $r$ is destination and not the origin, the rate is defined as the transfer flow from the previous subregion minus the trip completion rate which is again not bounded by any capacity function. In (iv) other cases, the rate is equal to the transfer flow from the previous subregion minus the transfer flow to the next subregion.

Eq. 3.8 estimates transfer flow from subregion $r$ to the next subregion $p^+(r)$ in path $p$ for all subregions except destination subregion $d$. It is the minimum of two terms: (i) the sending

---

[5]Note that $n_r(t)$ can be estimated with good precision from different type of sensors in real time, while $n_{o,d}^{p,r}$ would require more data and is expected to exhibit larger estimation errors.

flow from subregion $r$, which solely depends on the accumulation of subregion $r$, and (ii) the receiving capacity fraction of subregion $p^+(r)$ that is a function of two terms; $c_r^{p^+(r)}(n_{p^+(r)}(t))$ and $a_{o,d}^{r \to p^+(r)}$. Capacity at boundary between $r$ and $p^+(r)$, i.e. $c_r^{p^+(r)}(n_{p^+(r)}(t))$, is a function of accumulation in the receiving region $n_{p^+(r)}$, while $a_{o,d}^{r \to p^+(r)}$ is the fraction of boundary capacity that is assigned to $n_{o,d}^{p,r}$, which can be estimated by Eq. 3.9. Conceptually speaking, Eq. 3.9 states that $a_{o,d}^{r \to p^+(r)}$ corresponding to $(o, d, p, r)$ quartet depends on its relative accumulation among all traveler groups that cross the same boundary between subregion $r$ and $p^+(r)$. This equation can be derived by Little's formula Little (1961).

$$a_{o,d}^{r \to p^+(r)} = \frac{n_{o,d}^{p,r}}{\sum_i \sum_j \sum_w \mathbf{1}_{p^+(r)}(w^+(r)) \cdot n_{i,j}^{w,r}} \tag{3.9}$$

where $\mathbf{1}_{p^+(r)}(w^+(r))$ is an indicator function with value equal to 1 if the next subregions in the paths $p$ and $w$ are the same, otherwise zero. Note that traffic modeling presented through Eq. 3.6-3.9 is in compliance with the traffic model introduced in Yildirimoglu and Geroliminis (2014) except the constant average trip length assumption that is preserved here.

### 3.2.3 Transfer of variables from subregion-based to region-based model

This study aims at providing route guidance information in the subregion-based model (i.e. $q_{o,d}^p$; assigned flows on path $p$ between $o$ and $d$) by calculating equilibrium flows in the region-based model where route choice is taken into account by $\theta_{IJ}^H$ and $L_{IH}$. This procedure requires the transfer and translation of model variables in certain steps; aggregated route choice variables in the region-based model (i.e. $\theta_{IJ}^H$ and $L_{IH}$) must be matched with equivalent path flows (i.e. $q_{o,d}^p$;) in the subregion-based model. In the beginning of rolling horizon, traffic states from subregion-based model should be converted to region-based model equivalents (i.e. $N_{IJ}$, $\theta_{IJ}^H$, $L_{IH}$) and transferred to the route guidance model in order to correct the modeling error. In fact, in a real application, this step corresponds to traffic data acquisition from the network through loop detectors and probe vehicles. Similarly, equilibrium state in the region-based model has to be converted to $q_{o,d}^p$ values in order to apply them in the subregion-based model (plant). The following paragraph describes the estimation of region-based model states from subregion-based model states, while the transfer of equilibrium state from the region-based model to the subregion-based model will be further discussed in Section 3.3.2.

Let us consider region $I \in \mathcal{R}$ that consists of several subregions and that is heterogeneous in space with respect to link density. We use capital letters for variables related to regions and lower case letters for variables related to subregions. We denote $\mathcal{SR}$ the set of all subregions in the urban network, while $\mathcal{SR}_I$ is the set of subregions that belongs to region $I$. Let $\mathcal{P}^{od}$ be the set of all available paths connecting subregion $o$ and $d$.

In order to define a regional path equivalent, $P$, of a subregional path, $p$, we define function $\mathbb{P}(.)$. For example, if $p^*$ is a sequence of subregions $\{12, 11, 13, 14, 15, 7\}$, then $\mathbb{P}(p^*) = P^* = \{2, 3, 2, 2\}$

(see Figure 3.1, where subregions $\{12, 11\} \in \mathcal{SR}_2$, $\{13, 14, 15\} \in \mathcal{SR}_3$, and $\{7\} \in \mathcal{SR}_2$.) Note that the destination region along the path is always added to the end of the sequence $P$ in order to signal the destination region. Let $p_r^{|p|}$ denote the subregion $r$ and all the following subregions in path $p$. Eq. 3.10 represents function $\mathbb{N}(.)$ that helps define the next region along the path and regroup the variables accordingly. This function checks the next region in path $p$ given the current subregion $r \in \mathcal{SR}_I$ while ignoring the remaining part of the path.

$$\mathbb{N}\begin{pmatrix} I,H \\ p,r \end{pmatrix} = \begin{cases} 1 & \text{if } \mathbb{P}\left(p_r^{|p|}\right) = \{I, H, ...\}, \\ 0 & \text{otherwise.} \end{cases} \tag{3.10}$$

Eqs. 3.11-3.13 present the estimation of variables $N_{IJ}$, $L_{II}$ and $L_{IH}$, respectively. Note time $t$ is omitted for the sake of simplicity.

$$N_{IJ} = \sum_{o \in \mathcal{SR}} \sum_{d \in \mathcal{SR}_J} \sum_{p \in \mathcal{P}^{od}} \sum_{\substack{r \in \\ (p \cap \mathcal{SR}_I)}} n_{o,d}^{p,r} \tag{3.11}$$

$$L_{II} = \frac{\sum_{o \in \mathcal{SR}} \sum_{d \in \mathcal{SR}_I} \sum_{p \in \mathcal{P}^{od}} \sum_{r \in p} \left( \mathbb{N}\begin{pmatrix} I,I \\ p,r \end{pmatrix} \cdot n_{o,d}^{p,r} \right)}{\sum_{r \in \mathcal{SR}_I} n_r} \cdot \frac{\sum_{r \in \mathcal{SR}_I} f_r(n_r)}{\sum_{o \in \mathcal{SR}} \sum_{d \in \mathcal{SR}_I} \sum_{p \in \mathcal{P}^{od}} m_{o,d}^{p,d}} \tag{3.12}$$

$$L_{IH} = \frac{\sum_{o \in \mathcal{SR}} \sum_{d \in \mathcal{SR}} \sum_{p \in \mathcal{P}^{od}} \sum_{r \in p} \left( \mathbb{N}\begin{pmatrix} I,H \\ p,r \end{pmatrix} \cdot n_{o,d}^{p,r} \right)}{\sum_{r \in \mathcal{SR}_I} n_r} \cdot \frac{\sum_{r \in \mathcal{SR}_I} f_r(n_r)}{\sum_{o \in \mathcal{SR}} \sum_{d \in \mathcal{SR}} \sum_{p \in \mathcal{P}^{od}} \sum_{\substack{r \in \\ (p \cap \mathcal{SR}_I)}} \left( \mathbf{1}_{\mathcal{SR}_H}(p^+(r)) \cdot \hat{m}_{o,d}^{r \to p^+(r)} \right)} \tag{3.13}$$

where $\mathbf{1}_{\mathcal{SR}_H}(p^+(r))$ is the indicator function with value equal to 1 if $p^+(r) \in \mathcal{SR}_H$. In fact, Eqs. 3.12-3.13 are rewritten versions of Eqs. 3.3-3.4 with the variables at the subregion-based model. For example Eq. 3.12 could also be formulated as $L_{II} = \left( \theta_{II}^I(t) \cdot N_{II} \right) \ / \ N_I \cdot F_I\left(N_I, \sigma(N_I)\right) \ / \ M_{II}^I$.

Eq. 3.14 estimates the aggregated split ratio $\theta_{IJ}^H$, which represents the proportion of vehicles with next region $H$ among the vehicles currently in region $I$ and with destination region $J$. The denominator in Eq. 3.14 represents the accumulation in region $I$ and with destination region $J$, while the numerator depicts part of the same accumulation with next region $H$.

$$\theta_{IJ}^H = \frac{\sum_{o \in \mathcal{SR}} \sum_{d \in \mathcal{SR}_J} \sum_{p \in \mathcal{P}^{od}} \sum_{r \in p} \left( \mathbb{N}\begin{pmatrix} I,H \\ p,r \end{pmatrix} \cdot n_{o,d}^{p,r} \right)}{\sum_{o \in \mathcal{SR}} \sum_{d \in \mathcal{SR}_J} \sum_{p \in \mathcal{P}^{od}} \sum_{\substack{r \in \\ (p \cap \mathcal{SR}_I)}} n_{o,d}^{p,r}} \tag{3.14}$$

Following the same example in Figure 3.1 $o^* = 12$, $d^* = 7$ and $p^* = \{12, 11, 13, 14, 15, 7\}$, let us elaborate on Eqs. 3.11-3.14. Accumulation values $n_{12,7}^{p^*,12}$ and $n_{12,7}^{p^*,11}$ contribute to $N_{23}$, $L_{23}$ and $\theta_{22}^3$, as the next region along the path is 3. Note that $\mathbb{N}\binom{2,3}{p^*,12}$ and $\mathbb{N}\binom{2,3}{p^*,11}$ have the value of 1. Similarly, $n_{12,7}^{p^*,13}$, $n_{12,7}^{p^*,14}$ and $n_{12,7}^{p^*,15}$ contribute to $N_{32}$, $L_{32}$ and $\theta_{32}^2$. Finally, $n_{12,7}^{p^*,7}$ contributes to $N_{22}$, $L_{22}$ and $\theta_{22}^2$. In addition, the flow values at the boundary/destination $\hat{m}_{12,7}^{11 \to 13}$, $\hat{m}_{12,7}^{15 \to 7}$ and $m_{12,7}^{p^*,7}$ are incorporated in the calculation of $L_{23}$, $L_{32}$ and $L_{22}$, respectively.

Through Eqs. 3.11-3.14, we assume full observability of the state at the subregion-based model. Unfortunately, these estimations are subject to measurement errors in a real implementation. However, a decent loop detector infrastructure and sufficient amount of probe data would make these estimations possible within a tolerable error level. Note that underlying behavioral assumption about route choice does not have to be known in order to compute aggregated route choice variables (i.e. $\theta_{IJ}^H$ and $L_{IH}$); Eqs. 3.12-3.14 exploit accumulation and outflow values that can be directly observed in the plant. In that sense, these variables are accumulation-based variables, and they translate the traffic conditions in the plant into aggregated route choice variables in the model.

## 3.3 Methodological framework

The purpose of this section is three-fold: (i) to establish DUE conditions in the subregion-based model, which is considered as the base scenario, (ii) to establish DSO conditions in the subregion-based model, which enforces the users to fully cooperate at the subregion level that is the upper envelope of network performance improvement, and (iii) to provide travelers with route guidance information based on DSO conditions in the region-based model. In this study, we do not use a detailed microscopic or mesoscopic simulation model to reach equilibrium conditions. However, the framework that we employ is capable of reproducing variable trip lengths, hysteresis loops, and spatial heterogeneity. On the other hand, the aim of route guidance system is to minimize total network delay in the region-based model. However, given the heterogeneity effect in the network and different traffic characteristics in the region-based and subregion-based models, it becomes challenging to extract route guidance information, i.e. assigning vehicles to subregional paths based on the conditions at the region-based model without detailed information about the subregions. This study also establishes DSO conditions in the subregion-based model to emphasize the cost of using a more aggregated model to develop route guidance information. Note that the model where we estimate operation decisions (e.g. route guidance) and the plant where we test the decisions should be different. Therefore, a route guidance system based on microscopic or mesoscopic simulation models, as well, is exposed to the same modeling error. Yildirimoglu and Geroliminis (2014) showed through micro-simulation that subregion dynamics can represent quite accurately the evolution of congestion even if the link dynamics of the urban network are not known.

Note that the current chapter extends the work of Yildirimoglu and Geroliminis (2014) to a

route guidance scheme that is based on DSO state. This reference makes the path-level route decisions on a conventional link-based graph (i.e. actual roads and intersections), produces aggregated route choice variables, and establishes DUE conditions in the aggregated traffic model. However, it does not consider the interaction between the operation model (i.e. the region-based model) and the plant (i.e. the subregion-based model), and does not employ a rolling horizon procedure for the route guidance. In addition, the traffic model in Yildirimoglu and Geroliminis (2014) accounts for the route choice behavior by using explicit regional paths, while the region-based model in this study employs $\theta_{IJ}^H$ values.

Time $t$ is defined as a continuous variable in Section 3.2 from traffic modeling perspective, and Eq. 3.1-3.9 are computed through a differential equation solver in order to (almost) continuously update accumulation values. However, route choice decisions in the region-based and subregion-based model (i.e. $q_{o,d}^p(t)$, $L_{IH}(t)$ and $\theta_{IJ}^H(t)$) have to be taken at discrete time intervals. Therefore, between any two discrete time points, these variables are assumed constant. In the remainder of Chapter 3, time $t$ refers to discrete time points.

This section consists of two subsections. The next subsection presents the methodology to establish DUE and DSO conditions in the subregion-based model, while the following introduces the route guidance strategy based on DSO state in the region-based model.

### 3.3.1 Dynamic traffic assignment (DTA) in the subregion-based model

This subsection tackles the problem of establishing DUE or DSO conditions at the subregional level, as expressed by time dependent subregional route choice estimation. The urban network is divided into subregions with low-scatter MFDs and constant average trip lengths, i.e. $l_r$. We assume that the subregional O-D demand is given.

Traffic equilibrium (DUE or DSO) can be formulated as a fixed-point problem, where an additional cycle of assignment and network loading steps yield the same traffic conditions. A well-known heuristic solution, method of successive averages (MSA), is a suitable method in our study considering the characteristics of the problem in hand. MSA has been used in both static and dynamic network equilibrium problems as an incremental assignment type heuristic Daganzo and Sheffi (1977); Mahmassani and Peeta (1993). The method is based on predetermined step sizes along the descent direction. In other words, step size is not determined with respect to the characteristics of the current solution, which requires derivative information. Instead, it is determined a priori. Therefore, MSA stands as one of the most effective solution heuristics in case the derivative information is difficult to be acquired.

As DUE implementation requires equal and minimal experienced travel times on alternative paths at the same departure time, $T_{od}^p(t)$, the iterative traffic assignment procedure is performed based on time-dependent fastest path Chabini (1998). On the other hand, DSO formulation yields equal and minimal *marginal* travel times, $\tau_{od}^p(t)$, on the alternative paths. Time-dependent path marginal travel time represents the effect of one additional vehicle

using the same path on the system travel time. Hence, DSO conditions can be established by assigning vehicles to the paths with minimum time-dependent marginal costs (Peeta and Mahmassani, 1995; Peeta and Ziliaskopoulos, 2001). To identify these paths, subregion marginal travel times, $\tau_r(t)$, are computed and a time-dependent minimum cost algorithm Chabini (1998) is deployed. Marginal travel time in subregion $r$ is estimated as follows (time $t$ is omitted for the sake of simplicity):

$$\begin{aligned}
\tau_r &= T_r + n_r \cdot \frac{\mathrm{d}T_r}{\mathrm{d}n_r} = \frac{l_r \cdot n_r}{f_r(n_r)} + n_r \cdot \frac{\mathrm{d}}{\mathrm{d}n_r} \frac{l_r \cdot n_r}{f_r(n_r)} \\
&= l_r \cdot n_r \left( \frac{2}{f_r(n_r)} - n_r \cdot \frac{\mathrm{d}f_r(n_r)/\mathrm{d}n_r}{[f_r(n_r)]^2} \right)
\end{aligned} \tag{3.15}$$

where $T_r$ is the actual travel time of subregion $r$.

Let $\mathcal{H} = \{n_r(t)|\forall r, t\}$ and $\mathcal{S} = \{\rho_{odp}(t)|\forall o, d, p, t\}$, where $\rho_{odp}(t)$ denotes the ratio of the demand from origin $o$ to destination $d$ that chooses path $p$ at time $t$ (i.e. $q_{o,d}^p(t)/q_{o,d}(t)$). The flowchart to establish DUE/DSO state is presented in Figure 3.2a, while the corresponding algorithm is summarized as follows:

**Step 0:** *Initialization*

- Set $m = 1$.
- Set $\mathcal{P}^{od} = \{\}$   $\forall o, d$.
- Initialize subregion accumulations $\mathcal{H}^1$. The initialization step can reflect an empty network.

**Step 1:** *Direction finding*

- For each destination subregion $d$, apply all-to-one time-dependent fastest path or time-dependent minimum cost path algorithm Chabini (1998) depending on DUE or DSO implementation, respectively.
- To find the auxiliary paths in DUE and DSO formulations, employ actual travel times $\{T_r(t)|\forall r, t\}$ and marginal travel times $\{\tau_r(t)|\forall r, t\}$, respectively. Note that they both result from the accumulation set in the current iteration; i.e. $\mathcal{H}^m$.
- Perform an all-or-nothing assignment for each $(o, d, t)$ triplet, and assign $(\rho_{odp}(t))^* = 1$ for the chosen auxiliary path $p$.
- If the set $\mathcal{P}^{od}$ does not include the path $p$, add it to the set.

**Step 2:** *Calculating path assignment ratios*

- Set $(\rho_{odp}(t))^{m+1} = (\rho_{odp}(t))^m + (1/m) \cdot \left( (\rho_{odp}(t))^* - (\rho_{odp}(t))^m \right)$   $\forall o, d, p, t$.

**Step 3:** *Update*

- Set $m = m + 1$.

- Create the set of path assignment ratios $\mathcal{S}^m = \{(\rho_{odp}(t))^m | \forall o, d, p, t\}$.

- Implement the subregion-based model, i.e. Eq. 3.6-3.9, and get the new accumulations: $\mathcal{H}^m = g(\mathcal{S}^m)$. Note that $g(.)$ represents the subregion-based model dynamics.

**Step 4:** *Stopping criteria*

- Evaluate: $M = \sum_r \sum_t \left[(n_r(t))^m - (n_r(t))^{m-1}\right]^2$.

- If $M \geq \epsilon$, where $\epsilon$ is a pre-defined threshold, go to Step 1. Otherwise, terminate the procedure.
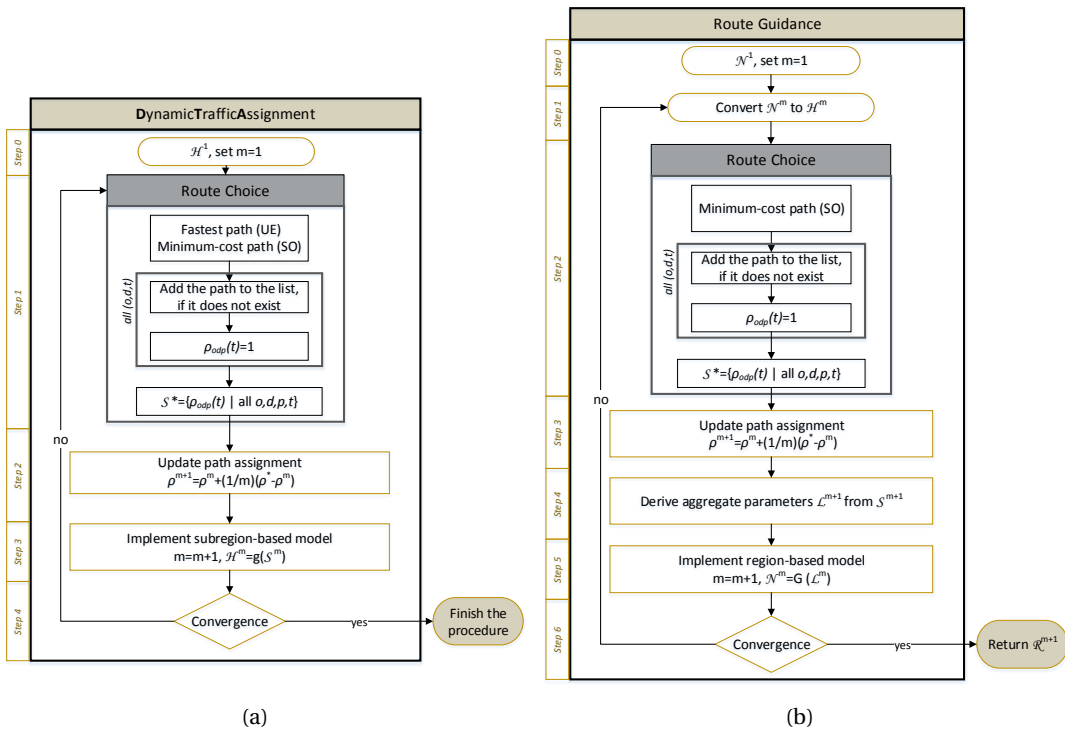


Figure 3.2: Methodological framework (a) DTA in the subregion-based model, (b) route guidance system in the region-based model.

### 3.3.2 Route guidance

The problem in this subsection is to establish DSO state in the region-based model and to provide travelers with route guidance information that satisfies the same conditions at the subregion-based model. This study produces the route guidance information through a rolling horizon framework, meaning that at each time step the method establishes DSO state in the operation model, i.e. the region-based model, over a rolling horizon and identifies a sequence of route choice decisions. However, only the first step of route choice decisions is applied

in the plant, i.e. the subregion-based model. The procedure is then repeated with a shifted horizon in the next time step. This closed-loop system allows the route guidance strategy, through a feedback loop, to correct disturbances due to variations in the predicted demand and the errors between the plant and the operation model.

The region-based model employs time-dependent average trip lengths, i.e. $L_{II}(t)$ and $L_{IH}(t)$, aggregated split ratios, i.e. $\theta_{IJ}^H(t)$, and an MFD affected by the heterogeneity level in the network, while the subregion-based model utilizes constant average trip lengths, i.e. $l_r$, and path assignment ratios, i.e. $\rho_{odp}(t)$. Therefore, a route guidance strategy based on the region-based model requires the conversion of route choice variables to the subregion-based model. This constitutes a significant part of the route guidance system.

Let $\mathcal{N} = \{N_I(t) | \forall I, t\}$ and $\mathcal{L} = \{L_{IH}(t), \theta_{IJ}^H(t) | \forall I, J, H, t\}$. Note that, in the rolling horizon framework, range of $t$ is limited by the length of the horizon. The route guidance flowchart is presented in Figure 3.2b, while the corresponding algorithm is summarized as follows:

**Step 0:** *Initialization*

- Set $m$=1.
- Apply Eqs. 3.11-3.14 to provide the operation model with required variables (i.e. $N_{IJ}(t_0)$, $L_{IH}(t_0)$, $\theta_{IJ}^H(t_0)$). Note that Eqs. 3.11-3.14 return the variables that reflect the state at the plant only in the beginning of rolling horizon; i.e. $t_0$. In a real implementation, this step corresponds to traffic data acquisition from the network.
- Assume $N_{IJ}(t) = N_{IJ}(t_0)$, $L_{IH}(t) = L_{IH}(t_0)$ and $\theta_{IJ}^H(t) = \theta_{IJ}^H(t_0)$ for $t > t_0$ and $\forall I, J, H$.
- At $t = t_0$, observe subregion accumulations in the plant, and calculate the relative accumulations for the subregions that belong to the same region, $\phi_r^I = n_r(t_0)/\sum_{i \in \mathcal{SR}_I} n_i(t_0); \forall r \in \mathcal{SR}_I$.

**Step 1:** *Conversion of region accumulations*

- Assume relative accumulation set, $\phi_r^I$, is constant over the rolling horizon, and convert regional accumulation set, $\mathcal{N}^m$, to subregional accumulation set, $\mathcal{H}^m$, i.e. $n_r(t) = N_I(t) \cdot \phi_r^I$ for $\forall t > t_0$ & $r \in \mathcal{SR}_I$. The region-based model in hand is able to evaluate regional accumulations. However, routing in the subregion-based model requires also the information on how the congestion is distributed within the region. This assumption allows us to account for the accumulation heterogeneity within the urban regions.

**Step 2:** *Direction finding*

- Apply Step 1 from DTA algorithm. Note that subregion accumulations, $\mathcal{H}^m$, deployed in this step are not exact as in DTA algorithm which is based on subregion-based model output. They are subject to the modeling error due to the granularity difference between the region-based and the subregion-based models and to the constant ratio assumption described above.

**Step 3:** *Calculating path assignment ratios*

- Apply Step 2 from DTA algorithm.

**Step 4:** *Producing aggregate route choice variables*

- In order to update regional accumulations, one needs to implement region-based model dynamics with the updated route choice variables (i.e. $\mathcal{L}^m$). However, the previous step here provides the route choice decisions at the subregional level. In fact, there are two ways to estimate these aggregate route choice variables. The first one is to compute *accumulation-based* variables through observing and regrouping the state at the plant, as in Eq. 3.12-3.14. This approach is exploited in Step 0 to complete the feedback loop and to correct the disturbances. However, subregion-based model is not available while we establish the equilibrium flows in the region-based model or while we produce route choice decisions based on predicted traffic state at the region-based model. The second approach, which we employ here, is a *trajectory-based* estimation of route choice variables, and it allows us to account for the impact of new path decisions. We assume properties of static traffic models (e.g. departing flow immediately reaches its destination, and it exists simultaneously in all the links along its assigned path) apply here.

- Denote $\tilde{L}_{II}, \tilde{L}_{IH}, \tilde{\theta}_{IJ}^H$ the aggregate route choice variables only for the departing flow in the corresponding time period. In fact, they represent marginal effects of the newly assigned demand on $L_{II}$, $L_{IH}$, and $\theta_{IJ}^H$. Therefore, a weighted average of them is later taken to compute the new aggregated route choice variables. In compliance with regional path $P$, let us denote $P_D$ the trip distance to cross in each region. Following the path example presented in Figure 3.1, $p^* = \{12, 11, 13, 14, 15, 7\}$, regional path $P^* = \{2, 3, 2, 2\}$ and assuming same constant average trip length (i.e. $l_r$) in each subregion, $P_D^*$ is $\{2 \cdot l_r, \ 3 \cdot l_r, \ l_r\}$. Eq. 3.16-3.18 provide trajectory-based estimation of aggregate route choice variables only for the departing flow. Note time $t$ is omitted, $P^{\text{end}}$ represents the last/destination region in sequence $P$, and $P^+(R)$ is the next region after $R$ in path sequence $P$.

$$\tilde{L}_{II} = \frac{\displaystyle\sum_{o\in\mathcal{SR}}\sum_{d\in\mathcal{SR}_I}\sum_{p\in\mathcal{P}^{od}}\sum_{\substack{R\in(I\cap P\backslash P^{\text{end}}),\\ P=\mathbb{P}(p)}}\left(\mathbf{1}_I(P^+(R))\cdot q_{o,d}\cdot\rho_{odp}\cdot P_D(R)\right)}{\displaystyle\sum_{o\in\mathcal{SR}}\sum_{d\in\mathcal{SR}_I}\sum_{p\in\mathcal{P}^{od}}\sum_{\substack{R\in(I\cap P\backslash P^{\text{end}}),\\ P=\mathbb{P}(p)}}\left(\mathbf{1}_I(P^+(R))\cdot q_{o,d}\cdot\rho_{odp}\right)} \tag{3.16}$$

$$\tilde{L}_{IH} = \frac{\displaystyle\sum_{o\in\mathcal{SR}}\sum_{d\in\mathcal{SR}}\sum_{p\in\mathcal{P}^{od}}\sum_{\substack{R\in(I\cap P\backslash P^{\text{end}}),\\ P=\mathbb{P}(p)}}\left(\mathbf{1}_H(P^+(R))\cdot q_{o,d}\cdot\rho_{odp}\cdot P_D(R)\right)}{\displaystyle\sum_{o\in\mathcal{SR}}\sum_{d\in\mathcal{SR}}\sum_{p\in\mathcal{P}^{od}}\sum_{\substack{R\in(I\cap P\backslash P^{\text{end}}),\\ P=\mathbb{P}(p)}}\left(\mathbf{1}_H(P^+(R))\cdot q_{o,d}\cdot\rho_{odp}\right)} \tag{3.17}$$

$$\tilde{\theta}_{IJ}^H = \frac{\displaystyle\sum_{o\in\mathcal{SR}}\sum_{d\in\mathcal{SR}_J}\sum_{p\in\mathcal{P}^{od}}\sum_{\substack{R\in(I\cap P\setminus P^{\mathrm{end}}),\\P=\mathbb{P}(p)}}\left(\mathbf{1}_H(P^+(R))\cdot q_{o,d}\cdot\rho_{odp}\right)}{\displaystyle\sum_{o\in\mathcal{SR}}\sum_{d\in\mathcal{SR}_J}\sum_{p\in\mathcal{P}^{od}}\sum_{\substack{R\in(I\cap P),\\P=\mathbb{P}(p)}}\left(q_{o,d}\cdot\rho_{odp}\right)} \tag{3.18}$$

- As opposed to Eq. 3.12-3.14, Eq. 3.16-3.18 make use of converted regional path $P$ in order to produce corresponding aggregated route choice variables. In addition, while Eq. 3.12-3.14 employ accumulation and outflow values observed through the plant, Eq. 3.16-3.18 incorporates only the departing flow on various paths (i.e. $q_{o,d}\cdot\rho_{odp}$) and the distance on the assigned trajectories (i.e. $P_{\mathrm{D}}(R)$). Again using the same example $p^* = \{12, 11, 13, 14, 15, 7\}$, assigned path flow at a departure time $q_{12,7}\cdot\rho_{12,7,p^*}$ simultaneously (i.e. at the very same departure time) contributes to the calculation of $\{\tilde{L}_{23}, \tilde{\theta}_{22}^3\}$, $\{\tilde{L}_{32}, \tilde{\theta}_{32}^2\}$, and $\{\tilde{L}_{22}, \tilde{\theta}_{22}^2\}$ considering the regions $R \in \{2, 3, 2\}$, respectively, on path $P^* \setminus P^{*end}$. Note that the contribution to the trip length in each region is weighted with respect to the distance crossed inside the same region (i.e. $P_{\mathrm{D}}^*$)

- Denote $\tilde{N}_{IJ}$ the *instantaneous* accumulation in region $I$ with destination region $J$ only due to the departing flow in the corresponding time period. Note that it also appears in the denominator of Eq. 3.18.

$$\tilde{N}_{IJ} = \sum_{o\in\mathcal{SR}}\sum_{d\in\mathcal{SR}_J}\sum_{p\in\mathcal{P}^{od}}\sum_{\substack{R\in(I\cap P),\\P=\mathbb{P}(p)}}\left(q_{o,d}\cdot\rho_{odp}\right) \tag{3.19}$$

- As $\tilde{L}_{II}, \tilde{L}_{IH}$, and $\tilde{\theta}_{IJ}^H$ represent the route choice variables only for the departing flow, compute the weighted average of accumulation in the network and departing flow. This allows us to combine the effect of departing flow (i.e. $\tilde{N}_{IJ}$) and existing accumulation (i.e. $N_{IJ}$) in the network.

$$(L_{II})^{m+1} = \frac{(N_{II})^m\cdot(L_{II})^m + \tilde{N}_{II}\cdot\tilde{L}_{II}}{(N_{II})^m + \tilde{N}_{II}} \tag{3.20}$$

$$(L_{IH})^{m+1} = \frac{\sum_J\left((N_{IJ})^m\cdot(\theta_{IJ}^H)^m\cdot(L_{IH})^m\right) + \sum_J\left(\tilde{N}_{IJ}\cdot\tilde{\theta}_{IJ}^H\cdot\tilde{L}_{IH}\right)}{\sum_J\left((N_{IJ})^m\cdot(\theta_{IJ}^H)^m\right) + \sum_J\left(\tilde{N}_{IJ}\cdot\tilde{\theta}_{IJ}^H\right)} \tag{3.21}$$

$$(\theta_{IJ}^H)^{m+1} = \frac{(N_{IJ})^m\cdot(\theta_{IJ}^H)^m + \tilde{N}_{IJ}\cdot\tilde{\theta}_{IJ}^H}{(N_{IJ})^m + \tilde{N}_{IJ}} \tag{3.22}$$

**Step 5:** *Update*

- Set $m = m + 1$.

- Create $\mathcal{N}^m$.

- Implement region-based model, Eq. 3.1-3.5, and get the new accumulations: $\mathcal{N}^m = \mathcal{G}(\mathcal{L}^m)$. Note that $\mathcal{G}(.)$ represents the region-based model dynamics.

**Step 6:** *Stopping test*

- Evaluate: $M = \sum_I \sum_t \left[(N_I(t))^m - (N_I(t))^{m-1}\right]^2$.

- If $M \geq \epsilon$, where $\epsilon$ is a pre-defined threshold, go to Step 1. Otherwise, terminate the procedure.

## 3.4   Results

This section presents several numerical studies with various demand levels spanning from uncongested scenarios to gridlock conditions. We test the proposed route guidance strategy under these different traffic conditions. In addition, we investigate the consistency between two traffic models and discuss the impact of DUE and DSO flows on the network MFD. Note that accumulation updates in region- and subregion-based model are performed continuously, while traffic assignment decisions (i.e. $q_{o,d}^p(t)$, $L_{IH}(t)$ and $\theta_{IJ}^H(t)$) are renewed every 1 min.

The case study network consists of three regions, where region 1 and 2 designate the suburb of the urban network, each comprising of 6 subregions, and region 3 designates the city center of the urban network comprising of 7 subregions, as schematically shown in Figure 3.1. For the sake of simplicity, we assume every subregion has the same production MFD and subregional average trip length. Note that the regional average trip lengths are time-varying, see Eq. 3.12 and 3.13. The exogenous time-varying demand simulates one hour of morning peak followed by another hour of very low demand to fully clear the network. Region 1 and 2 generate most of the demand towards region 3 that, as the central business district, attracts most of the trips.

One of the main contributions of the chapter is introducing two models with different levels of aggregation both in terms of vehicle accumulation state and route choice behavior. It is important that, despite their different granularity, the models can reproduce consistent description of aggregated traffic variables, that is necessary for real time traffic management implementation. To scrutinize the consistency of regional and subregional models, we test the models with similar exogenous demands and initial accumulations, i.e. $Q_{IJ}(t)$ and $q_{o,d}(t)$, and $N_{IJ}(0)$ and $n_{o,d}^{p,r}(0)$ are consistent. In addition, the only traffic state measurements needed for the region-based model are the regional average trip lengths, i.e. $L_{II}$ and $L_{IH}$, and the aggregated split ratios, $\theta_{IJ}^H$, that are available using Eq. 3.12–3.14. Figure 3.3(a)-3.3(b) illustrate the evolution of $N_{IJ}(t)$ and $N_I(t) = \sum_J N_{IJ}(t)$ as estimated by the region-based (Eq. 3.1-3.5) and subregion-based models (Eq. 3.6-3.8), which indicate that the two models are consistent. In addition, Figure 3.3(c) illustrates the results of the same test on a region-based model (as described in Ramezani et al. (2015) and Footnote 2 of the current chapter) where vehicle routes include at most one boundary crossing. The inconsistency here exposes the significant effect

of such modeling consideration when vehicles are allowed to cross more than once a boundary between two regions (in the aforementioned publication route choice was constrained to be consistent with this modeling assumption).

Figure 3.4 demonstrates the existence of DUE and DSO traffic conditions through the heuristic method described in Section 3.3.1. Figure 3.4(a) shows experienced travel times of two paths (4-16-19 and 4-17-19), along with the associated path assignment ratios. This specific case presents the traffic conditions corresponding to the DUE traffic assignment procedure (at the subregional level). Under DUE conditions, travelers tend to use the paths that provide lowest travel time among the alternatives. This results in minimal and equal travel times on the used alternative paths, apparent in Figure 3.4(a). On the other hand, Figure 3.4(b) depicts marginal travel times of the two paths and the associated path assignment ratios under DSO conditions. Similarly, the marginal travel times among the alternative paths are minimal and equal. Notably, the path assignment ratios in case of DUE and DSO are different. Note that Figure 3.4 spans only first one hour of the simulation where the exogenous demand exists in the scenario, as the route choice decision is being made at the departure of the trip.

Table 3.1 lists the total network delay in three assignment scenarios for various demand levels. In all demand levels, as expected, DSO performs better than DUE. The benefit from this cooperative assignment strategy ranges from 10.4% to 17.5%. RG strategy produces results fairly close to the DSO conditions except 110% demand scenario. This is an important finding as the developed strategy proves its applicability in field by giving information to drivers what parts of the city to avoid. This strategy can also be integrated in a hierarchical traffic management framework as the first layer, where a more detailed second layer module could assign link-based routes to the drivers. Given the fast computation of RG strategy in this chapter, such an approach could also ease the (sometimes) infeasible computational effort of detailed RG algorithms. The performance difference between DSO and RG scenarios can be represented as the cost of using a more aggregated model (i.e. the region-based model) to produce routing information for a relatively detailed model (i.e. the subregion-based model). The reason for the failure in 110% demand level might be the inconsistency between subregion-



(a)    (b)    (c)
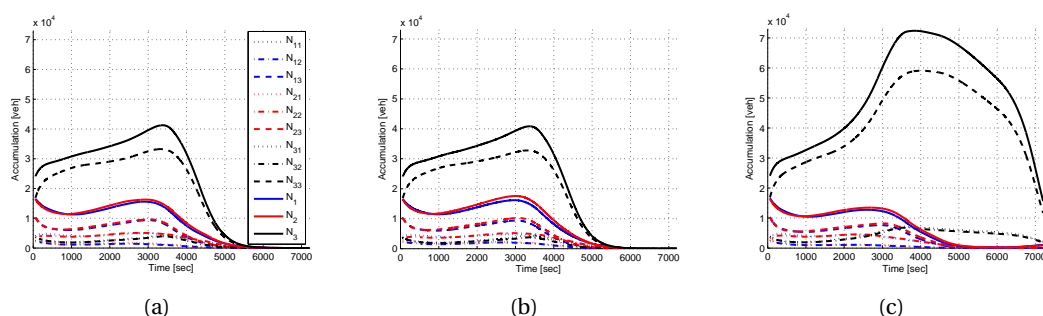
Figure 3.3: Validation of model consistency: (a) Region-based model (operation model), (b) Subregion-based model (plant), and (c) Region-based model with assumption that there is at most one boundary crossing.

(a)



(b)
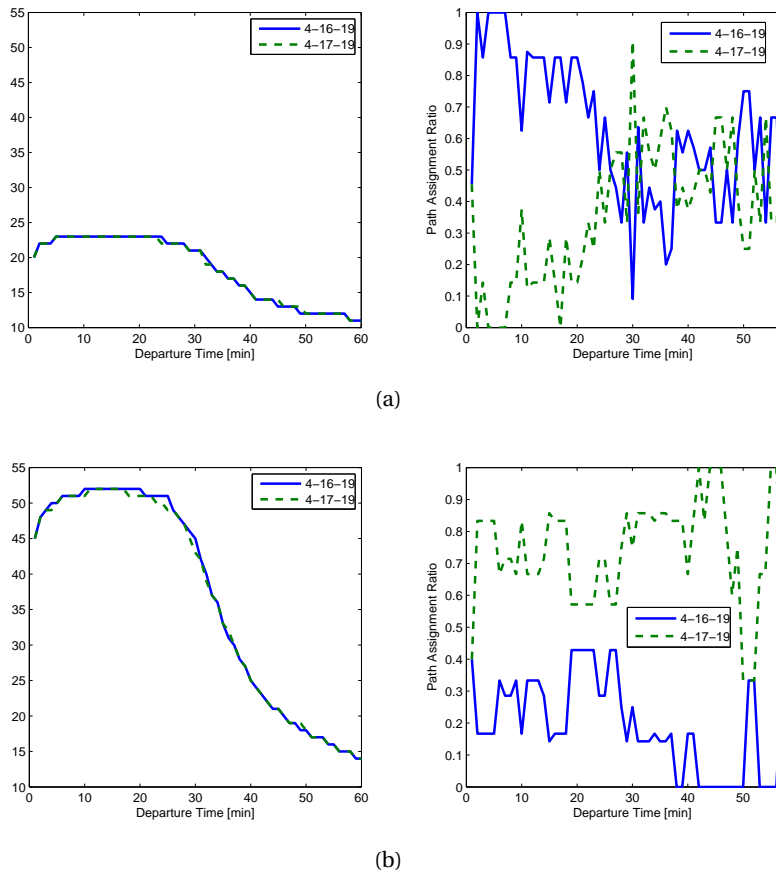
Figure 3.4:   Travel times and path assignment ratios for path 4-16-19 and path 4-17-19: (a) Actual (experienced) travel time in case of DUE, (b) Marginal (experienced) travel time in case of DSO.

based and region-based model for the accumulation values close to the jam accumulation. Although subregion-based model does not allow values beyond jam accumulation, such a constraint does not exist in region-based model. Over the iterations in the route guidance model, inaccurate route choice estimations are done probably due to this inconsistency. In addition, the inconsistency between accumulation-based (i.e. Eqs. 3.12-3.14) and trajectory-based (i.e. Eqs. 3.16-3.18) route choice variables probably becomes quite critical in high accumulation states, and the model fails to keep track of traffic in the rolling horizon. In cases of very high demand (many subregions reach gridlock), a combination of route guidance with perimeter control is expected to produce better results.

Figure 3.5 illustrates the MFD functions (production vs. accumulation) for different demand levels under three assignment scenarios. For all demand levels and DUE scenarios, region 3 experiences congestion (with states in the decreasing part of the MFD). 95%, 100% and 105% demand levels produce no significant hysteresis in any region with any of the assignment scenarios. On the other hand, 110% demand level creates extremely congested and almost
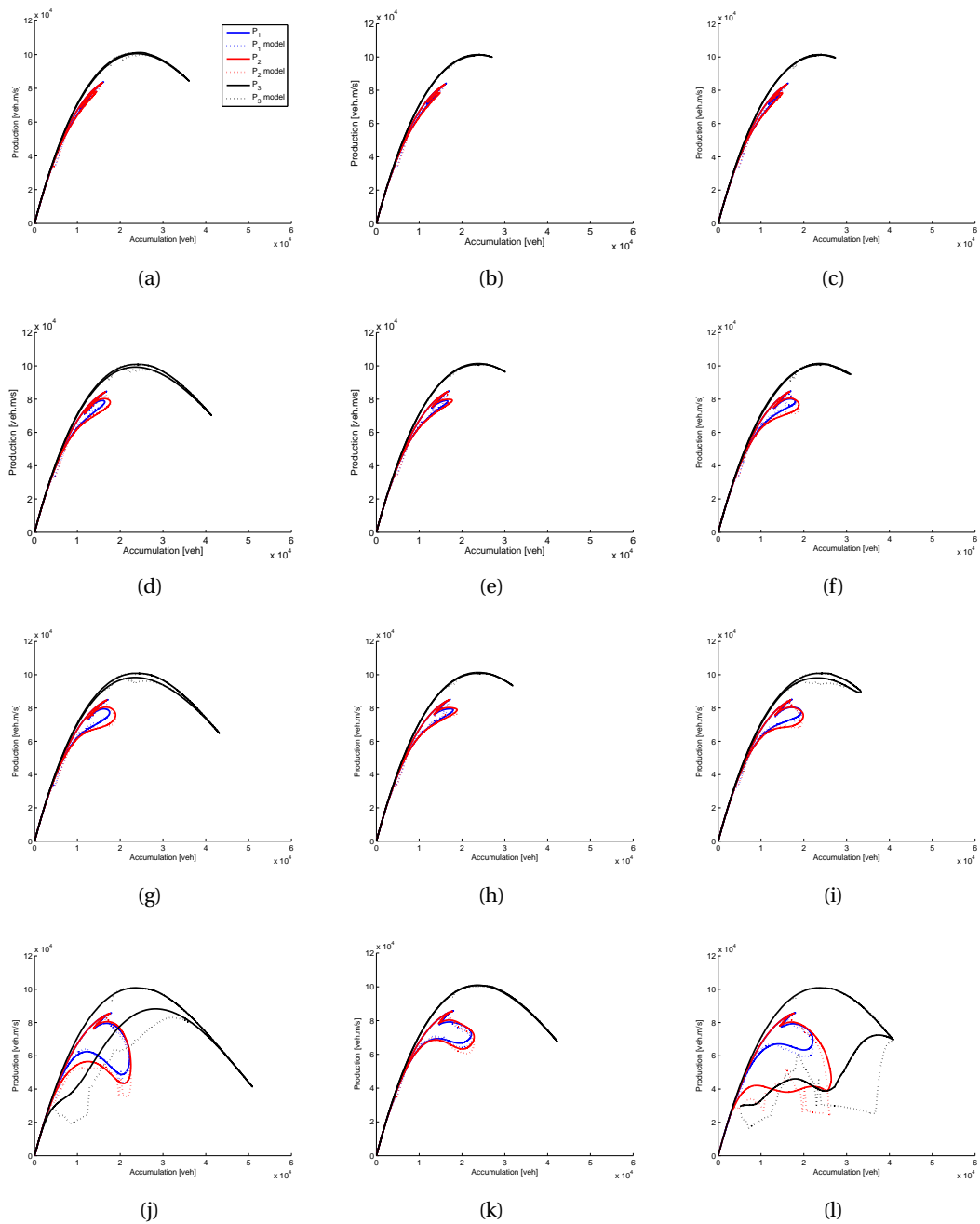
Figure 3.5: MFDs for the 3 regions of Figure 3.1, estimated with the regional and subregional models for different demand levels (rows) and different routing strategies (columns). (a) DUE 95%, (b) DSO 95%, (c) RG 95%, (d) DUE 100%, (e) DSO 100%, (f) RG 100%, (g) DUE 105%, (h) DSO 105%, (i) RG 105%, (j) DUE 110%, (k) DSO 110%, (l) RG 110%.

Table 3.1: Total network delay ($10^8$ [veh.s])

| | Demand Level | | | |
|---|---|---|---|---|
| | 95% | 100% | 105% | 110% |
| Dynamic user equilibrium (DUE) | 2.31 | 2.77 | 2.95 | 4.01 |
| Dynamic system optimum (DSO) | 2.07 (10.4%) | 2.39 (13.7%) | 2.55 (13.6%) | 3.31(17.5%) |
| Route guidance (RG) | 2.07 (10.4%) | 2.42 (12.6%) | 2.65 (10.2%) | 4.28(-6.7%) |

gridlock conditions in many of the subregions in DUE and RG scenarios, which destroys the homogeneity of the regions, and causes hysteresis in the network unloading phase (see Figure 3.5(j) and 3.5(l)). Even with 110% demand level, DSO is able to prevent gridlock in subregions and avoid hysteresis in region 3. Despite the fact that regions 1 and 2 are much less congested than region 3, they exhibit a significant hysteresis in this scenario (see Figure 3.5(k)). This issue will be further investigated in the chapter. Note that gridlock conditions are expected to be observed only during rare events (e.g. evacuation), and the region-based model is able to reproduce accurate MFD patterns and follow hysteresis loops (when exist) despite the lack of information for detailed route choices of drivers. Further research is required to investigate with real data the route choices of drivers at the subregion level that will shed more light in proper assignment tools for simulation and modeling.

Figure 3.6 illustrates subregional and regional accumulations for the three traffic assignment scenarios, i.e. DUE, DSO and RG strategy for 100% demand level. Notably, with DUE and DSO conditions, subregion accumulations are distributed in a harmonious way; they follow the same trend in the network loading and unloading phases. Although RG strategy is able to follow the same trend in region accumulations (see Figure 3.6(e)-3.6(f)), the evolution of subregion accumulations shows differences due to limitations between region-based and subregion-based models. Despite this limitation, RG strategy produces fairly close results to DSO (see Table 3.1). Note that odd-numbered subregions at the periphery (see Figure 3.1) are connected to one single subregion from the inner core, while even-numbered subregions are connected with two of them. This largely affects the inflow and outflow characteristics of subregions, and causes inhomogeneity within the regions. The mild hysteresis in regions 1 and 2 presented in Figure 3.5(d)-3.5(i) is due to this hexagonal network structure and difference in accumulations between odd- and even-numbered regions (see Figure 3.6(a)-3.6(c)). This result indicates that equilibrium type assignments cannot establish homogeneity within the regions by itself; physical network structure and connectivity between different parts of the network play a significant role as well. This should be an interesting research priority.

Figure 3.7 shows the subregion accumulations to visualize the spatial correlation structure and propagation of congestion in the network for demand level 100% and the DUE, DSO, and RG assignments. In this case study, traffic congestion mainly starts from the central subregion 19 and propagates in the remaining subregions of region 3, before it partially captures the subregions in the periphery (with less magnitude). In DUE conditions during the peak period, $t = 60$ (min), most of the subregions in region 3 are highly congested while regions 1 and 2
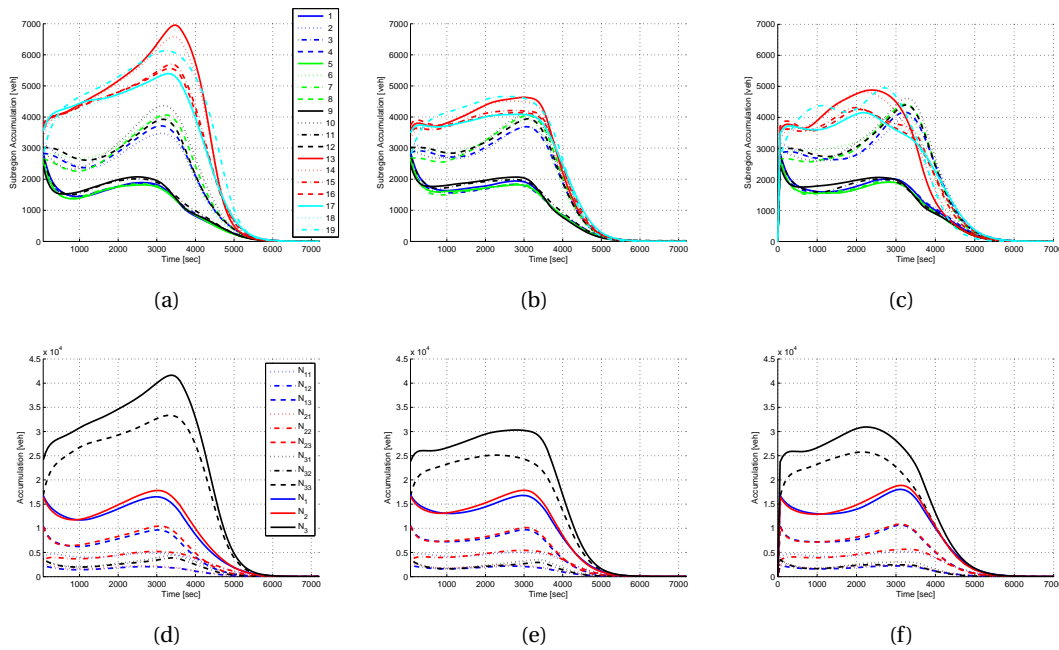
Figure 3.6: Subregional accumulations with 100% demand: (a) DUE, (b) DSO, and (c) RG. Regional accumulations: (d) DUE, (e) DSO, and (f) RG.

are uncongested, whereas in DSO conditions some drivers choose longer routes avoiding the center of congestion and thus, accumulations spread more homogeneously in all the 3 regions. Note that small changes in regional route choices can drastically alter the level of congestion for the whole city. Such an implementation of route guidance could be applied without great effort and driver compliance might not be a significant issue as the deviation from DUE conditions is small.
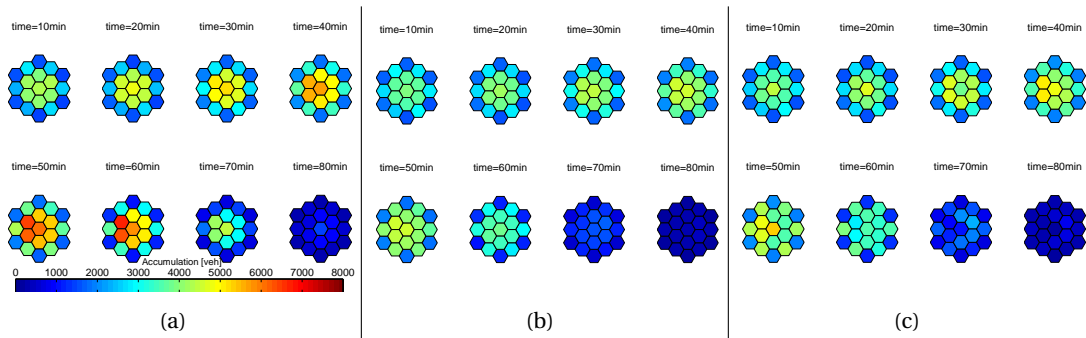


Figure 3.7: Evolution of accumulation over time with 100% demand (a) DUE, (b) DSO, (c) RG.

Figure 3.8 depicts the results from the "current best" or "en-route" assignment strategy with 100% demand level. In this strategy, drivers update their routes based on instantaneous travel times in every time interval. We notice that hysteresis phenomenon in MFD is quite significant here with respect to equilibrium scenarios (e.g. DUE, DSO or RG) investigated before (see

Figure 3.5(d)-3.5(f)). Although there is no subregion that suffers from gridlock conditions (see Figure 3.8(b)), especially the accumulation in region 3 is not properly distributed among subregions. This increases the inhomogeneity within region 3 and causes a significant hysteresis in the MFD shape (see Figure 3.8(a)). MFDs for region 1 and 2 exhibit hysteresis too, but this is mainly due to the physical network structure discussed above. Note that all previous simulation-based MFD estimations rely on enroute assignment mechanisms and instantaneous travel times. Although en-route decision mechanisms are able to properly distribute the congestion in the network, it hardly represents the real driver behavior. Drivers might switch to the alternative routes in response to unexpected traffic conditions. However, in a "normal" day they are expected to rely on their past experiences. These results clearly indicate that the way route choice behavior is modeled has significant impacts on network-wide properties such as MFD functions. This is in accordance with micro-simulation findings in Mahmassani et al. (2013c) that more informed and adaptive drivers can improve the network performance.



(a)    (b)    (c)
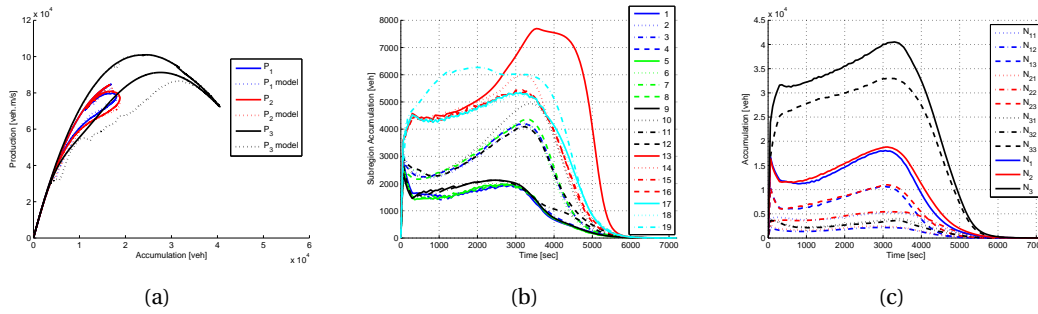
Figure 3.8: En-route assignment with 100% demand: (a) MFD, (b) subregion accumulations, and (c) region accumulations.

DSO conditions, in principle, are difficult to be put into practice, because some drivers might experience higher travel costs than they do under DUE conditions. Nevertheless, as we show now, the developed regional route guidance strategy does not penalize a significant portion of travelers and has some strong potential for a field implementation. Figure 3.9 depicts the distribution of travel time benefit across users for DSO and RG scenarios with respect to DUE conditions. Note that ~89% and ~87% of users benefit from DSO and RG, respectively. In addition, only ~0.6% of users suffer from more than 5 min delay in both DSO and RG scenarios. These results indicate that a RG scenario based on DSO conditions can bring an overall benefit to the system in the cost of causing little delay to a low percentage of drivers. In fact, Jahn et al. (2005) reaches near-DSO state by integrating explicit user constraints that guarantee certain level of fairness across users. A similar approach could be used here to prevent delays more than a certain amount. This should be a future research priority.
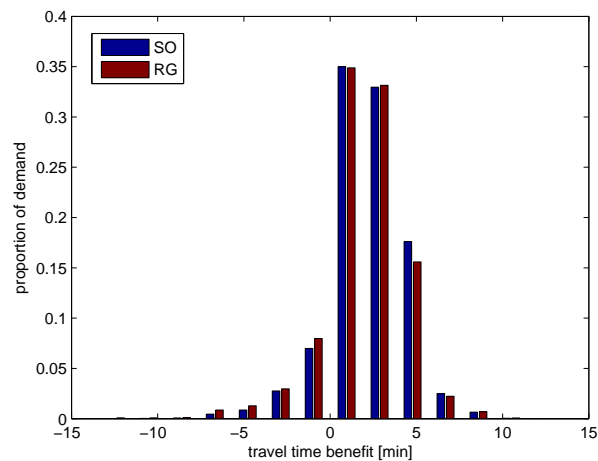
Figure 3.9: Travel time benefit of users in DSO and RG scenarios with respect to DUE conditions.

## 3.5 Summary

This chapter has presented two MFD-based traffic models with different levels of vehicle accumulation state and route choice behavior aggregation. This configuration enables us to incorporate the effect of spatial congestion heterogeneity and route choice dynamics in MFD modeling and to investigate the consequence of limited traffic state measurements measurements in performance of RG strategy. We have integrated DUE and DSO flows in the analysis. The results clearly demonstrate that equilibrium state has a significant impact on the existence of hysteresis in MFD. Previous MFD estimations based on microscopic and mesoscopic simulation models where route choice behavior is modeled with en-route decision mechanisms suffer from hysteresis even in mildly congested scenarios. However, in this study, we observe that equilibrium flows create a non-hysteresis MFD except in very high demand scenarios where gridlock is not avoidable. In addition, we have developed a route guidance control system based on the parsimonious region-based model, and applied the resulting commands in the subregion-based model. The results demonstrate that the advisory route guidance system produces conditions fairly close to DSO state, except very high demand scenarios where the consistency between the model and the plant should be further investigated.

# 4 Aggregated dynamic route choice patterns for large-scale mixed urban/freeway networks

THIS chapter observes aggregated route choice patterns (i.e. vehicle distance traveled in different systems, urban/freeway share among the trips, regional split ratios and average trip lengths) in a large scale mixed urban/freeway system through an extensive data set of 20,000 taxis in Shenzhen, China. The main objective of this work is to reconstruct the aggregated patterns through shortest path algorithm, and test the equilibrium assumptions widely recognized in traffic community at the network scale. In other words, we replace each observed trajectory with a shortest path that connects the same origin and destination points, and reproduce aggregated variables. We propose several different travel cost definitions to be considered in shortest path algorithm. These results can enhance parsimonious network models and lead to better traffic predictions for large-scale congested networks.

## 4.1 Introduction

Large-scale traffic modeling and macro-scale management strategies remain a big challenge partly due to unpredictability of choices of travelers (e.g. route, departure time and mode choice). While there is strong understanding and vast literature of route choice modeling, there is no rigorous efforts, to the authors' knowledge, that investigates aggregated patterns (e.g. traffic load on freeway and urban subnetworks) that result from such modeling. Most of the analysis at the network level is based on simplistic models or simulations, (i) which are difficult to calibrate and (ii) require a large number of input variables/parameters, that might not be observable with the current available data. In addition, there is not enough explanation on if and how people adapt their choices with respect to dynamic traffic conditions in the network, and how all this changes network traffic properties.

Route choice models are essential to forecast travelers' behavior under hypothetical scenarios, and most importantly to predict traffic conditions in transportation networks. Although modeling route choice is quite challenging, given the complexity of human behavior and uncertainty about travelers' perceptions, they are essential part of dynamic traffic assignment

## Chapter 4. Aggregated dynamic route choice patterns for large-scale mixed urban/freeway networks

(DTA) models that are expected to accurately predict traffic conditions in transportation networks. For instance, dynamic user equilibrium (DUE) assumes that travelers have the perfect knowledge of travel costs along the network, and choose the routes that minimize their travel costs. DUE state can be reached through the repeated implementation of shortest path algorithm over the iterations particularly in simulation-based DTA models (Chiu et al., 2011). On the other hand, in dynamic stochastic user equilibrium (DSUE), travelers are assumed to have imperfect knowledge of travel costs, and choose the routes that minimize their perceived travel costs. While shortest path algorithm would be the straightforward way to establish DUE conditions, there is a vast literature of discrete choice models that could be exploited to reach DSUE conditions. The most appreciated models are Multinomial Logit, C-Logit (Cascetta et al., 1996), Path Size Logit (Ben-Akiva and Bierlaire, 1999), Link-Nested Logit (Vovsha and Bekhor, 1998), Multinomial Probit (Daganzo et al., 1977) and Error-Component model (Bekhor et al., 2002). Calibration of the above models is a challenging task because of computational complexity, lack of detailed data and occasionally overfitting of the parameters. Nevertheless, nowadays massive real-time data from multiple sources allow us to revisit and reformulate previous models to describe more accurately realistic congested traffic conditions.

The objective of this work is dual. On the one side, we utilize a unique dataset from a large number of probe vehicles that provide GPS location every few seconds in a megacity in China. We investigate if consistent empirical observations for dynamic route choice patterns can be made. Secondly we investigate how well the network level aggregated patterns can be estimated through shortest path or DUE assumptions. Obviously, not all travelers choose the shortest path to go from the origin to the destination point due to either lack of information or uncertainty of perception. However, the question is; what is the cost of using shortest path assumption when we aggregate the results from many origin-destination pairs? Transportation networks, by design, consist of urban motorways, expressways, large arterials, local streets, etc. Normally, long trips are expected to be bound to higher category roads, while short trips may use the local, finer-meshed network that can be continuously approximated. In this work, we exploit a very detailed GPS dataset of 20,000 taxis from Shenzhen, China, and we aim at exploring traffic load in different components of the transportation network. In addition, aggregated route choice results could be employed in parsimonious network traffic models.

The literature on parsimonious network traffic models is quite recent due to lack of detailed data (simulation efforts or empirical data at the static level have been assessed in the past). It was observed from empirical data in downtown Yokohama (Geroliminis and Daganzo, 2008) that by spatially aggregating the highly scattered plots of flow vs. density from individual detectors, the scatter almost disappeared and a well-defined Macroscopic Fundamental Diagram (MFD) exists between space-mean flow and density. The idea of an MFD with an optimum accumulation belongs to Godfrey (1969) and similar approaches were introduced later by Herman and Prigogine (1979) and Mahmassani et al. (1984), while the empirical verification of its existence under dynamic and congested conditions is conducted by Geroliminis and Daganzo (2008). Despite these recent findings for the existence of MFDs with low scatter, these curves should not be a universal law. Heterogeneity in congestion distribution can affect

the shape/scatter or even the existence of MFD (Buisson and Ladier, 2009; Geroliminis and Sun, 2011b). Mazloumian et al. (2010) investigate the effect of heterogeneity on the MFD considering variance of link density as an independent variable. They observe well-defined relations between flow and density when link density variance is about constant.

These results are very critical, because they prove the applicability of MFD concept to heterogeneously loaded cities in case the network is partitioned into smaller homogenous areas. Ji and Geroliminis (2012) and Saeedmanesh and Geroliminis (2015) develop clustering algorithms for heterogeneous transportation networks. The proposed mechanism produces compact subnetworks that have small link density variance. Resulting subnetworks can be used to develop macroscopic traffic management strategies. Real-time large-scale traffic management strategies, e.g. perimeter control (Geroliminis et al., 2013; Haddad et al., 2013; Aboudolas and Geroliminis, 2013; Gayah et al., 2014; Haddad and Shraiber, 2014; Ramezani et al., 2015); gating (Keyvan-Ekbatani et al., 2012, 2015) that benefit from parsimonious models with aggregated network dynamics, provide promising results towards a new generation of smart hierarchical strategies. However, control strategies that require prediction of future traffic conditions face certain limitations regarding the route choice behavior in the multi-region urban network; travelers' reaction and adaptation to new management strategies is not considered in the control design. Yildirimoglu and Geroliminis (2014) tackle this problem and establish equilibrium conditions in a multi-region urban network with MFD dynamics. They assume travelers have the knowledge of average traffic conditions in the subnetworks (e.g. city center, periphery roads, etc.), and choose the routes that minimize their approximate travel cost. In that respect, they establish DSUE conditions in the traffic network. Yildirimoglu et al. (2015b) extends this work to a route guidance strategy where travelers are forced to cooperate with each other in order to reach system optimum conditions. This chapter aims for providing physical evidence from real data and exploiting aggregated route choice patterns in a large scale mixed urban/freeway network.

The remainder of the chapter is organized as follows; in the next section, we introduce the GPS dataset from 20,000 taxis in Shenzhen. In the following section, the methodological framework for estimating aggregated dynamic route choice patterns is elaborated, and the results are presented. Finally, last section concludes the chapter with future work directions.

## 4.2 Data analysis

The data set consists of GPS tracks of around 20,000 taxis in a fast growing Chinese mega-city; Shenzhen. The rapid investment created one of the fastest-growing cities in the world with a population close to 11 million and, as expected, large congestion problems both in the urban and freeway system of the city. The network structure includes 28647 nodes and 35099 links, out of which 3354 are freeway links. The data set consists of trips (on the same day) from taxis equipped with a GPS sensor that stores its location every 10-40 seconds. For every GPS point, it is also known whether the taxi carries a passenger or not, which allows us to

distinguish between trips with and without passengers. Assuming that taxi passengers follow routes similar to regular cars in the network, we only focus on taxi trips with passengers. Even if taxi drivers might seek non-standard paths, we expect that speed estimates based on taxis with passengers are a good representation of all vehicles and that aggregated patterns are not influenced much by local low-level route choices.

In order to identify traffic conditions in the network, we first map-match GPS observation with the closest link in the network, and estimate the link speeds. To make a similar analogy with (Yildirimoglu and Geroliminis, 2014), we also aggregate the GPS observations that are matched with urban links inside the 1x1 km regions and compute the average speed for the urban components inside them. Following equations present the estimated space-mean link and region speed, respectively.

$$v_l(t) = \sum_{j \in \mathcal{J}} s_j(l, t) / \sum_{j \in \mathcal{J}} h_j(l, t) \tag{4.1}$$

$$v_r(t) = \sum_{j \in \mathcal{J}} s_j(r, t) / \sum_{j \in \mathcal{J}} h_j(r, t) \tag{4.2}$$

where $\mathcal{J}$ is the set of all journeys, $s_j(l, t)$ - $s_j(r, t)$ and $h_j(l, t)$ - $h_j(r, t)$ are respectively the distance traveled and the time spent by journey $j$ in link $l$ - in urban region $r$ at time period $t$. Note that in case there is no GPS observation on a link, the corresponding value is replaced with the average speed of neighboring links.

Figure 4.1(a) depicts the average speed in all the links throughout a 24-hr period (i.e. link-level matching, Eq. 4.1 applied for all links), while Figure 4.1(c) presents the average speed in the freeway links and 1x1 km cells that represent urban components of the network (i.e. regional matching, Eq. 4.1 and 4.2 applied for freeway links and urban cells, respectively). Note that, in the regional matching case, all urban links inside the same cell are assigned the same average speed. Clearly, the speed on freeway links is higher than the speed on surrounding urban links in both estimation methods, which points to the hierarchy of roads in mixed urban/freeway systems. Note that the average speed calculation could be done for smaller time periods, e.g. 1-hr. Figure 4.1(a) and 4.1(c) represents only the time-independent average traffic conditions in a day. In addition, Figure 4.1(b) and 4.1(d) show the percentage distribution of speeds among the urban and freeway links. Although the distribution of speed among the freeway links remains the same (freeway link speed estimation is done with link-level matching in both approaches), the distribution of urban link speeds shows certain differences. While link-level matching produces rather extreme speed estimations (close to 0km/h or more than 50 km/h), regional matching smoothes the estimations and return a higher number of links with moderate speed measurements. In fact, regional matching changes the way the perception of travelers is considered in the analysis; travelers are assumed to know only the average traffic conditions in the subnetworks, which may be more realistic from a traveler information perspective. We have noticed that a significant portion of the links in the urban network have very few observations, which makes the estimation of link speeds quite noisy. Although the available taxi data set is one of the most dense worldwide, penetration rate is rather small and

does not allow for accurate estimations at the link level. This observation as well motivates our approach for regional speed aggregation and investigation of aggregated patterns.
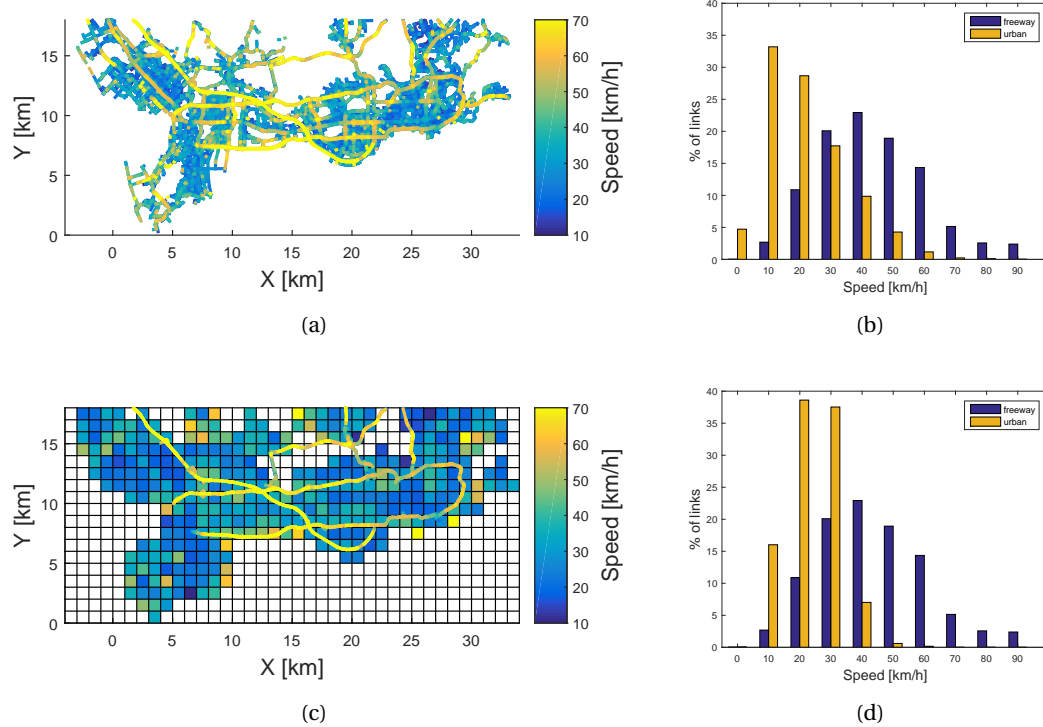


Figure 4.1: Estimation of speed through probe vehicles (a) speed in all links, (b) distribution of speed in link-level matching, (c) speed in urban regions and freeway links, (d) distribution of speed in regional matching.

Figure 4.2 presents the average and the variance of speeds estimated for every 1-hr period with two different matching approaches. As freeway speed is estimated in the same way in both approaches, there is only one curve for the mean and the variance of freeway speed throughout the day. Although the curve of the mean urban speed is quite similar in link-level and regional matching cases, variance values are significantly different. Regional matching approach smoothes the results in urban areas especially with low number of samples, and significantly decreases the variance level. Note that at the link level, in many cases, speed has to be estimated from measurements of neighboring links. Figure 4.2(a) also presents the ratio of average urban and freeway speeds, which is more or less constant throughout the day. This implies that relative speeds in two systems remain the same, and congestion builds up simultaneously.

Figure 4.3 introduces the distribution of average number of observations in 1-hr period throughout the day. Figure 4.3(a) presents the results from link-level matching; around 50% of urban links do not have any measurement the whole day. Although around 40% of freeway links lack measurement as well, interpolation of downstream/upstream link speeds is

(a)

(b)
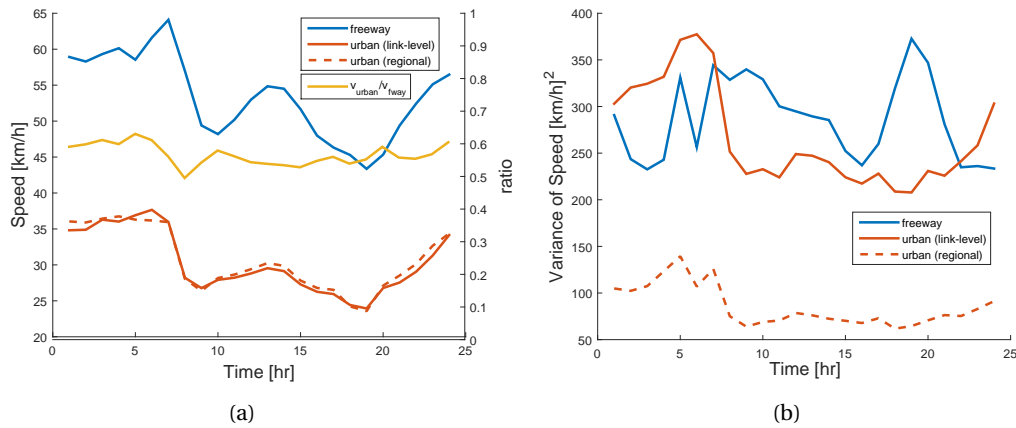
Figure 4.2:  Mean and variance of speed in network subsystems (a) Evolution of mean speeds and , (b) variance of speeds in freeway and urban systems.

expected to perform well in freeways as it is consistent with traffic flow dynamics. While the high percentage in the freeway is due to a very accurate mapping of short sections (e.g. road curvatures, arc-shaped ramps, etc.), arterials are problematic as there are sequences of roads with lack of data due to low demand. On the other hand, Figure 4.3(b) shows the distribution of observation on the links (freeway) or in the regions (urban) in regional matching approach. Apparently, the aggregation of observations within 1x1 km cells removes the low sampling problem; around 70% of urban cells have more than 20 observations/hr.



(a)

(b)
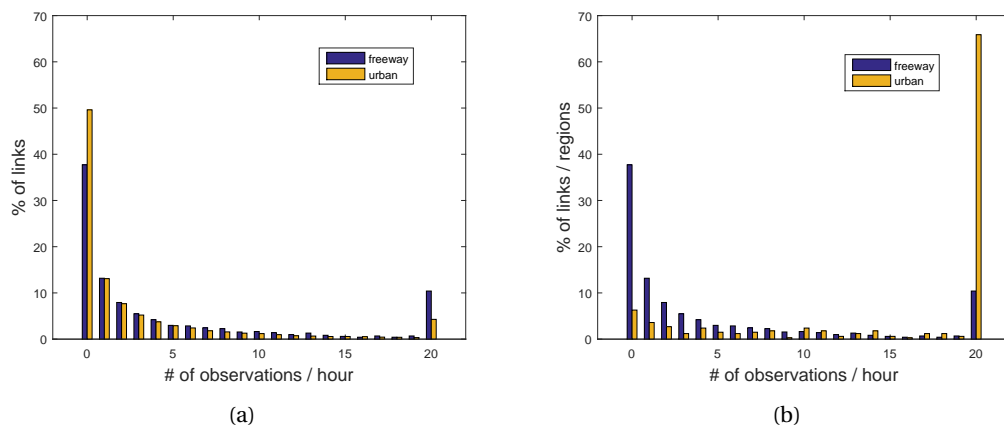
Figure 4.3:  Distribution of GPS observations in network subsystems (a) distribution of link counts in link-level matching , (b) distribution of freeway link and urban region counts in regional matching.

## 4.3  Methodological framework

In this section, we create aggregated route choice patterns through individual trip trajectories, and attempt to reconstruct them with shortest path results. In other words, we compare the patterns identified through the observed paths and the shortest paths that result from a few different travel cost considerations. The aggregated patterns we are interested in include average distance crossed in certain subnetworks, regional split ratios that define the proportion of accumulation aiming for a particular neighboring subnetwork and distance traveled in urban/freeway systems. The former two variables are strongly related with dynamic equations of MFD modeling, which will be further discussed in the next subsection. In addition, they are very important for developing mixed urban/freeway control strategies, such as integrated corridor management (Diakaki et al., 1997) or mixed traffic signal/ramp metering schemes (Haddad et al., 2013).

Travel cost functions that we consider in the shortest path calculation are;

- distance (regardless of hierarchy of roads)

- estimated travel time (a single constant speed for each speed limit zone, abbreviated as est. tt.)

- time-independent regional travel times (average values over 24-hr, abbreviated as reg. tt.)

- time-independent link travel times (average values over 24-hr, abbreviated as link tt.)

- time-dependent regional travel times (average values every 1-hr, abbreviated as dyn. reg. tt.)

- time-dependent link travel times (average values every 1-hr, abbreviated as dyn. link tt.)

The comparison of time-independent (static) and time-dependent (dynamic) cost scenarios may reveal the importance of dynamic traffic conditions in the network, and it may explain if and how people adapt to varying traffic conditions. In time-dependent cost scenarios, we employ the travel costs that correspond to departure time of trips in the shortest path calculation. On the other hand, the difference between regional and link-level matching may expose the significance of traveler perception regarding the travel costs along the network. Note that the number of observations per link especially in the urban system can be quite low, which may lead to under- or over-estimated link travel times. Regional matching puts all the measurements in the urban system together and assigns the smoothed average value to the links inside the same area.

71

### 4.3.1  MFD modeling

[1]Let us assume that the urban network is partitioned into several regions with well-defined MFDs. Let $Q_I(t)$ [veh/s] denotes the exogenous traffic flow demand generated in region $I$, $N_I(t)$ [veh] be the vehicle accumulation in region $I$. The traffic flow conservation equations are as follows:

$$\frac{\mathrm{d}N_I(t)}{\mathrm{d}t} = Q_I(t) - \sum_{H \in \mathcal{V}_I} M_{IH}(t) + \sum_{H \in \mathcal{V}_I} M_{HI}(t) \tag{4.3}$$

where $\mathcal{V}_I$ is the set of regions that are directly reachable from region $I$, i.e. adjacent regions to region $I$ including region $I$ itself. $M_{II}(t)$ [veh/s] would be the *internal trip completion rate* for accumulation in $I$ with destination $I$ (without going through another region), while the *transfer flow* for accumulation in $I$ to neighboring region $H$ is denoted by $M_{IH}(t)$ [veh/s].

Internal trip completion rates ($I = H$) and transfer flows ($I \neq H$) are estimated corresponding to the ratio between accumulations as:

$$M_{IH}(t) = \theta_{IH}(t) \cdot \frac{F_I\big(N_I(t), \sigma(N_I(t))\big)}{L_{IH}(t)} \tag{4.4}$$

where $F_I(\cdot)$ [veh.m/s] is the production MFD of region $I$ (i.e. the total distance traveled per unit time in the region) that is a function of the region accumulation, $N_I(t)$, and the link density *heterogeneity* across all region $I$ links, $\sigma(N_I(t))$. Moreover, $L_{IH}(t)$ [m] is the average trip length corresponding to transfer trips from region $I$ to its neighbor region $H \in \mathcal{V}_I$, and $\theta_{IH}(t)$ is the percentage of accumulation in region $I$ going through neighboring region $H$; hence $\sum_{H \in \mathcal{V}_I} \theta_{IH}(t) = 1$. Modeling the region link density heterogeneity, i.e. $\sigma(N_I(t))$, is investigated in Ramezani et al. (2015). Previous work assumes $L_{IH}(t)$ and $\theta_{IH}(t)$ are constant permanently or over the prediction horizon, and identify the control actions by respecting this assumption. In this chapter, we aim for providing the average trip length $L_{IH}(t)$ and split ratio $\theta_{IH}(t)$ using the available OD demand and shortest path assignment. This new piece of information could lead to a better design of traffic control/management strategies. Note that average trip length $L_{IH}(t)$ and split ratio $\theta_{IH}(t)$ represent the aggregated effect of route choice decisions in the region $I$, and they will be referred as the route choice variables of MFD modeling in the remainder of the chapter.

### 4.3.2  Aggregating route choice patterns

Let us assume that a rectangular urban region $I$ (presented in Figure 4.4(a)) is connected to other homogeneous regions through its edges. In addition, there are freeway connections at the boundary of the region, which might be considered as a separate subnetwork. Therefore,

---

[1]This section presents a simpler class of traffic model presented in Section 3.2.1. The main difference is that the new model does not account for explicit destination information in the corresponding accumulation or demand variables. While this difference allows for simpler expression, the accuracy of such a model is beyond the scope of this chapter.

$\mathcal{V}_I$ is a set of neighboring regions, freeway subnetwork and region $I$ itself. The following formulas estimate the average trip length $L_{IH}(t)$ and split ratio $\theta_{IH}(t)$, respectively.

$$L_{IH}(t) = \sum_{j \in \mathcal{J}} s_j(I, H, t) \, / \, \sum_{j \in \mathcal{J}} \mathbf{1}_{\mathcal{R}^+}(s_j(I, H, t)) \quad H \in \mathcal{V}_I \tag{4.5}$$

$$\theta_{IH}(t) = \sum_{j \in \mathcal{J}} \mathbf{1}_{\mathcal{R}^+}(s_j(I, H, t)) \, / \, \sum_{j \in \mathcal{J}} \mathbf{1}_{\mathcal{R}^+}(s_j(I, t)) \quad H \in \mathcal{V}_I \tag{4.6}$$

where $s_j(I, H, t)$ is the distance crossed by journey $j$ in region $I$ till the boundary of region $H$ (see Figure 4.4(a)), and $\mathbf{1}_{\mathcal{R}^+}(.)$ is an indicator function with value 1 in case $(.) \in \mathcal{R}^+$, 0 otherwise. In other words, Eq. 4.5 represents the average distance for all vehicles that cross a non-negative distance in region $I$ to go to neighboring region $H$, while Eq. 4.6 indicates the portion of vehicles going to region $H$ among all vehicles that cross a non-negative distance in region $I$. Note that $s_j(.)$ can be coming from the observed GPS tracks or shortest path trajectories depending on the scenario analyzed.
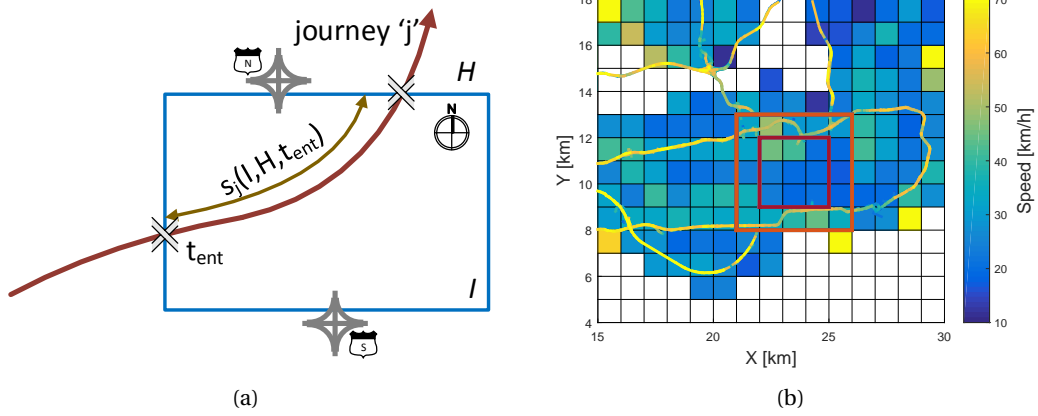


Figure 4.4: Schematic of the study area (a) a trajectory through the area of interest, (b) study area.

In addition to route choice variables needed in MFD modeling, one can estimate vehicle distance traveled (VDT) in urban and freeway systems. This information allows us to define the traffic load in different components of the mixed urban/freeway system. Let us denote $\mathcal{L}_F$ and $\mathcal{L}_U$ the set of freeway and urban links in the network, respectively. The following formula compute VDT in freeway and urban system.

$$dt_f(t) = \sum_{j \in \mathcal{J}} \sum_{l \in \mathcal{L}_f} s_j(l, t) \tag{4.7}$$

$$dt_u(t) = \sum_{j \in \mathcal{J}} \sum_{l \in \mathcal{L}_u} s_j(l, t) \tag{4.8}$$

Traffic equilibrium can be formulated as a fixed-point problem, and in mathematics, a fixed point of a function is the function's domain that is mapped to itself. That is, $x^*$ is a fixed point

of function $f$, if and only if $x^* = f(x^*)$. For instance, in traffic equilibrium, $x^*$ can represent the route choice decisions, and $f$ may represent the mapping of route choice decisions into themselves, see Bottom (2000) for a detailed discussion. The mapping starts with a set of route choice decisions (e.g. set of $L_{IH}(t)$ and $\theta_{IH}(t)$), computes resulting traffic conditions (e.g. set of $v_l(t)$ or $v_r(t)$), and returns a new set of route choice decisions as a response to new conditions. Note that mapping function can be defined with respect to traffic conditions as well, where $x^*$ would be traffic conditions and $f$ the mapping function of them. In addition, these two mappings would be equivalent with respect to fixed-point solution. If one has a fixed-point, the other one has too, and if one does not then none does. Within this framework, assuming DUE exists, we observe fixed-point aggregated route choices (i.e. set of fixed $L_{IH}^*(t)$ and $\theta_{IH}^*(t)$) and traffic conditions (i.e. set of fixed $v_l^*(t)$ or $v_r^*(t)$). Using the supposed fixed-point traffic conditions, we assign the traffic demand to the shortest paths and reconstruct the aggregated route patterns. The existence of fixed-point solution or equilibrium conditions would require estimated values calculated this way to be very close to observed ones. Therefore, this analysis allows us to test the validity of DUE assumptions at the aggregated network level.

## 4.4   Results

In order to produce the route choice variables required in MFD modeling, we choose an area $I$ of 3x3 km (presented with red lines in Figure 4.4(b)). The area in the west of the square is the major city center with strong trip attractions and high level of congestion. Note that there are two freeway connections from the chosen urban area; one in the north, one in the south. For the sake of simplicity, we consider the vehicles that take the freeway via these interchanges as one single traveler group. This assumption can easily be relaxed later in order to calculate inflows to specific freeway junctions. Although a detailed partitioning algorithm is not yet conducted, we assume the chosen area is homogenous enough to assign MFD properties (variance of speed in this region is significantly smaller than the overall variance shown in Figure 4.2(b)). We also assume there are 4 other homogeneous urban regions connected to the chosen area through its edges in 4 directions (i.e.north, south, west, east). Therefore, $\mathcal{V}_I$ is a set of six components; 4 neighboring regions, freeway subnetwork and region $I$ itself. In other words, we define 6 traveler groups with distinct choices; heading for one of 4 neighboring regions, taking the freeway in the north or south, and finishing the trip inside the region. We apply Eq. 4.5-4.6 for all GPS trajectories and shortest paths that result from a set of travel cost definitions described in previous section. Since all OD pairs in the whole data set have been considered in the analysis, the number of trips that actually cross region $I$ can be quite different for the observed trajectories and shortest path results. This might significantly affect the accuracy of estimated variables.

Figure 4.5 introduces observed and estimated (or reconstructed) split ratios and average trip lengths. Note that solid curves in Figure 4.5 represent observed trajectories, while dashed curves are estimated with shortest path trajectories. Travel costs that have been tested here are time-dependent regional and link travel times. Not all but certain components of split

ratios and average trip lengths reveal a strong dynamic behavior, see for example $\theta_{I,inside}$ (the fraction of trips that finish within the region) and $L_{I,west}$ (the average trip length for trips exiting through the west boundary), and these patterns have been roughly followed by estimated variables. Note that, between 11h and 13h, in dynamic link travel time case, $L_{I,north}$ estimations are not very consistent with neighboring time periods (see Figure 4.5(d)). This is again probably due to low sampling and inaccurate travel time estimation for certain links. The network is much less dense in the northern part of the square and few destinations hit towards this direction (less than 3% as shown in Figure 4.5(a) and 4.5(c)).



(a)

(b)

(c)

(d)
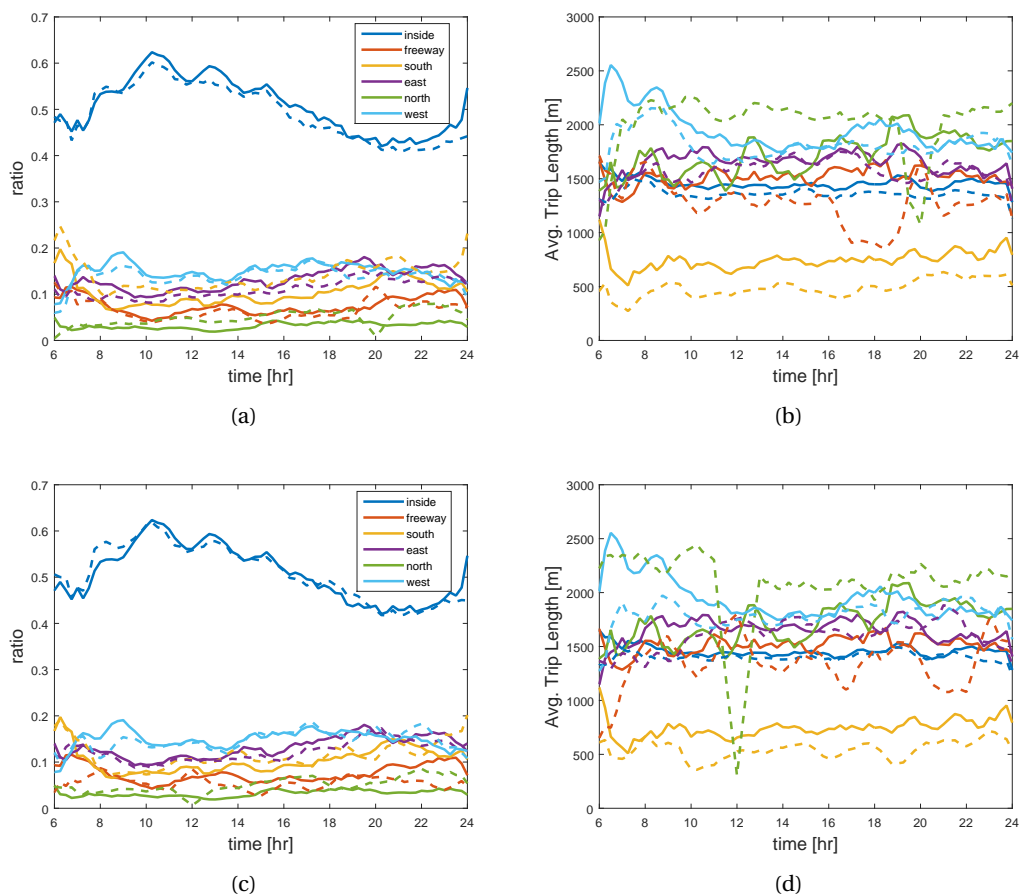
Figure 4.5: (a) Split ratios with dynamic regional travel time, (b) average trip lengths with dynamic regional travel time, (c) split ratios with dynamic link travel time, (d) average trip lengths with dynamic link travel time.

In order to calculate $dt_f$ and $dt_u$ values, we choose an area of 5x5 km with both freeway and urban links (see the orange lines in Figure 4.4(b)). We apply Eq. 4.7 and 4.8, and calculate the observed and estimated VDT values using actual and shortest path trajectories, respectively. Figure 4.6 depicts the resulting curves for static/dynamic regional/link travel times. Note that, in all scenarios, freeway load has been overestimated (for about 10%), and consequently

urban system use has been underestimated. In fact, one of the major freeways present in this analysis is a toll freeway. Therefore, travel cost on this specific freeway section should be a certain combination of toll fee and travel time. However, this analysis will be reported in an extended version of this article; this issue requires further investigation. Another interesting observation is the similarity between time-dependent and -independent scenarios. Except the afternoon peak period, both scenarios produce quasi-identical results. This implies that relative speeds in urban and freeway systems remain approximately constant throughout the day except the afternoon peak.
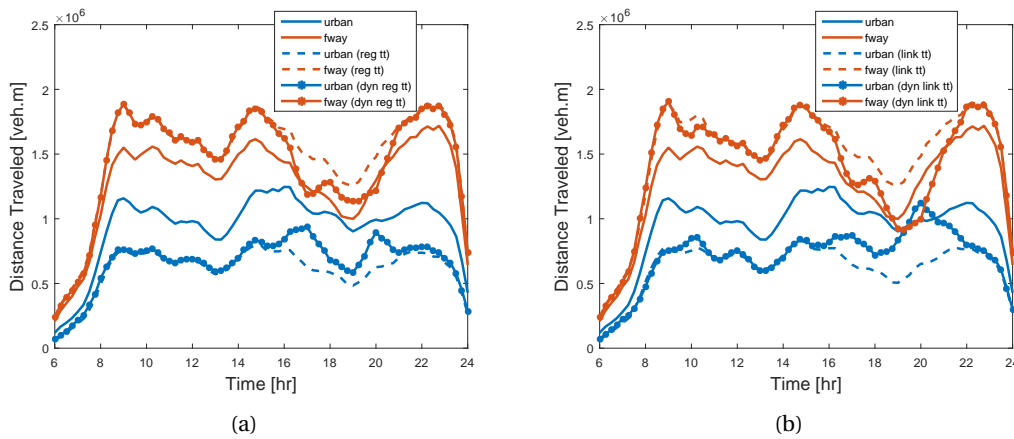


Figure 4.6: Total distance traveled (a) with regional travel times, (b) with link travel times

Intuitively, it is known that travelers tend to use freeways (if available) because of its uninterrupted nature in order to cross a long distance in a mixed urban/freeway network. For each trip in the data set, we define an urban share parameter that provides the percentage of distance crossed in urban links. In other words, an urban share parameter close to 0 represents a trip mostly done on the freeway, while a parameter close to 1 indicates a trip that solely crosses urban links. Figure 4.7(a) depicts the number of trips and average urban share parameter that correspond to total trip distance values. Note that Figure 4.7(a) represents the aggregated values for trips that are within certain distance range. For example, the left end of the curves represent the average urban share and total number for all trips that have a trip distance between ]0, 0.5]km range, so on so forth. As expected, the average urban share is quite close to 1 for short distance trips; travelers do not tend to use the freeway to cross a short distance. Average urban share is almost monotonically decreasing with increasing distance, and converges to a value around 0.4. The access/egress distance from/to freeway system from the urban links, where origin and destination points are located, does not allow the urban share value to be less than 0.4 in this specific network example. It might be interesting to investigate this phenomenon in other mixed urban/freeway environments and compare network properties. On the other hand, Figure 4.7(b) introduces the patterns that result from shortest path algorithm with different travel cost considerations. Except the travel cost based on distance, all other cost definitions are able to follow the trend. Note that the

results from regional and link-level travel times (for time-dependent and time-independent cases) are fairly close to each other in Figure 4.5-4.7, which points to the fact that regional traffic information knowledge, as well, can reproduce aggregated route choice patterns. Lower number of observations per link may turn the problem to the advantage of regional travel time consideration. Less observations may lead to inaccurate link travel time estimations, while regional travel time is not expected to significantly deteriorate due to smoothing effect. This analysis should be a future research priority.
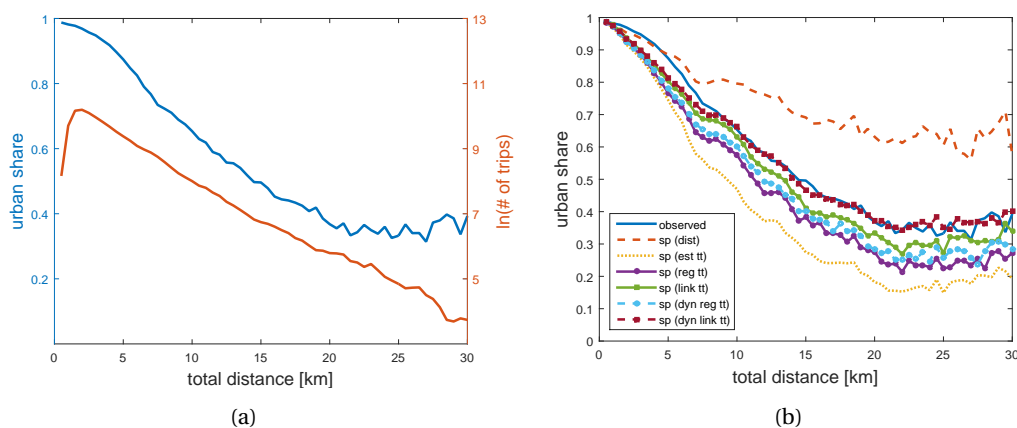


Figure 4.7:  (a) Urban share and number of trips vs. total trip distance, (b) urban share with shortest path

Table 4.1 presents the mean absolute errors (MAE) for the estimation of variables presented in Figure 4.5-4.7.  For each subcomponent (e.g.  traveler groups for $\theta_{IH}$ and $L_{IH}$, urban and freeway subsystems for VDT, and no subcomponent for urban share estimation) and each observation (e.g. time periods for $\theta_{IH}$, $L_{IH}$, VDT, and total distance values for urban share estimation), we simply calculate the absolute difference between the observed and estimated (or reconstructed) variables, and compute the average of all to report MAE. In overall, estimations that are based on distance or estimated travel time produce quite inaccurate results compared with other travel cost definitions. Although time-dependent regional travel time produces slightly better results than its time-independent version in case of $\theta_{IH}$ and $L_{IH}$ estimation, dynamic link travel time fails to improve the results of its static counterpart. The probable reason behind it might be the lack of sufficient number of observations for certain links. Note that the best estimation has been achieved with time-independent link travel time computation for both $\theta_{IH}$ and $L_{IH}$. In case of $dt_f$, $dt_u$ and urban share estimation, dynamic cost definitions produce better results than their static counterparts, and the best results have been obtained with time-dependent link-level travel times.

Table 4.1: Mean Absolute Error (MAE) of estimated variables

| | Travel Cost | | | | | |
|---|---|---|---|---|---|---|
| | distance | est. tt | reg. tt | link tt | dyn. reg. tt | dyn. link tt |
| $\theta_{IH}$ | 0.038 | 0.042 | 0.023 | 0.015 | 0.019 | 0.015 |
| $L_{IH}$ [m] | 395 | 372 | 199 | 191 | 195 | 201 |
| VDT [veh.m.1e5] | 4.35 | 2.96 | 2.30 | 2.20 | 1.86 | 1.77 |
| urban share | 0.18 | 0.16 | 0.10 | 0.05 | 0.08 | 0.02 |

## 4.5 Summary

This chapter is a first attempt, to the authors' knowledge, to identify and to reproduce aggregated route choice patterns in a mixed urban/freeway environment. An extensive GPS data set of 20,000 taxis from Shenzhen, China make it possible to monitor, understand and analyze aggregated route choice patterns. The variables that we focus on in this chapter include regional split ratios, average trip lengths, VDT in different systems and urban/freeway shares. Note that the former two variables can be embedded in MFD dynamics, and may lead to better traffic predictions. However, testing of such hypothesis requires data from both loop detectors and probe vehicles, which is not available in this case. An important objective of this work is to reproduce the aggregated patterns through shortest path assumption, and to test the validity of equilibrium assumptions at an aggregated level. We replace each observed trajectory with a shortest path that connects the same origin and destination points, and reproduce aggregated variables that result from different travel cost considerations. Although the extent of GPS observations allows only limited traffic state observability, shortest path implementations with link-level and regional speed calculations seem to produce decent replications of aggregated patterns. Nevertheless, further research is needed to finalize this ongoing study and draw rigorous conclusions.

# Part II

# 5 Experienced travel time prediction for congested freeways

T RAVEL time is an important performance measure for transportation systems, and dissemination of travel time information can help travelers make reliable travel decisions such as route choice or departure time. Since the traffic data collected in real time reflects the past or current conditions on the roadway, a predictive travel time methodology is needed to derive the information to be disseminated. However, an important part of the literature either uses instantaneous travel time assumption, and sums the travel time of roadway segments at the starting time of the trip, or uses statistical forecasting algorithms to predict the future travel time. This study benefits from both the available traffic flow fundamentals (e.g. shockwave analysis, bottleneck identification) and advanced data mining techniques to provide travel time prediction. The methodological framework of this approach sequentially includes a bottleneck identification algorithm, clustering of traffic data in traffic regimes with similar characteristics, development of stochastic congestion maps for clustered data and an online congestion search algorithm, which combines historical data analysis and real-time data to predict experienced travel times at the starting time of the trip. The experimental results based on the loop detector data on Californian freeways indicate that the proposed method provides promising travel time predictions under varying traffic conditions.

## 5.1   Introduction

Predictive travel time is valuable information required by drivers and transportation managers to improve the quality of travel and to make control decisions. The provision of travel time information through Advanced Traveler Information Systems (ATIS) enables drivers to make decisions, such as route choice and departure time. In addition, besides the fundamental relation with traffic flow modeling, travel time can be used by transportation agencies to deploy efficient control measures and to prevent potential traffic congestion. Besides its direct implementation for users and practitioners, travel time experiences strong fluctuations and stochastic traffic phenomena that make its reliable estimation and prediction a challenging physical and mathematical task. Thus, its modeling and estimation requires a combination of

81

correct physics and strong statistical tools.

There are two general methods for obtaining travel time; direct measurement and estimation (Yeon et al., 2008). Direct measurement of travel time can be obtained through test vehicles, license plate matching techniques (automatic vehicle identification, AVI) and ITS probe vehicle techniques. Direct measurement techniques may be misleading in the case of low sampling rates and existence of outlier travel time observations. In order to suppress noise signals, Dion and Rakha (2006) developed an adaptive filtering algorithm, which adjusts its validity window by tracking average travel times. On the other hand, travel time estimation is conducted using the data taken from loop detectors, smart phones or global positioning systems (GPS) devices. As numerous freeways around the world are equipped with loop detectors that collect flow, speed and occupancy information, a vast literature of travel time estimation in freeways relies on them. Travel time estimation can be either based on local velocity measurements, or more sophisticated models that attempt to correlate vehicle observations at multiple locations (Coifman, 2002; Coifman and Krishnamurthy, 2007). In addition, estimation models tend to underestimate travel times under congested conditions because of the queue dynamics which cannot be adequately represented in the model. To address this problem, Yeon et al. (2008) made use of discrete time Markov Chains, where the states correspond to whether or not a link is congested, and computed the expected route travel time for several adjacent short links. Furthermore, GPS data provide new opportunities for traffic state estimation and they can be incorporated in estimation algorithms for travel time (Herrera and Bayen, 2010). Mazaré et al. (2012), using the experimental probe data from a field experiment and loop detector data from California Performance Measurement System (PeMS), evaluates the trade-offs between the two types of data. To produce an improved estimate of velocity field, speed measurements from GPS or loop detectors are combined using a mathematical traffic model equivalent to Cell Transmission Model and a traffic state estimation algorithm, the ensemble Kalman filtering. Resulting velocity fields are used to compute travel time, assuming that a vehicle travels at the mean speed reported in each cell. However, the essential problem with travel time information is that it always has to refer to future conditions in the roadway. On the contrary, traffic data collected in real time reflect past or current conditions in the roadway.

Using traffic speed information, there are two ways to compute travel time; instantaneous and experienced. Instantaneous travel time is calculated combining the speed measurements in different locations at the departure time of a trip. On the other hand, experienced travel time is calculated by traveling a trajectory through the velocity field. The time it takes to traverse each segment is calculated, and the speed measurement at the time when the trajectory reaches the next segment is used to compute its travel time. Mathematically speaking, if a freeway is divided into $i=1,\ldots,I$ sections ($I$ is the most downstream section), and $\tau_i(t_d)$ is the travel time of section $i$ for starting time $t_d$, then the instantaneous $T^{in}_{S,I}(t_d)$ and experienced, $T^{ex}_{S,I}(t_d)$ travel times to traverse all sections between $S$ and $I$ for departure time $t_d$ are estimated as

follows ($T_{S,I}^{ex} = 0, for S \geq I$):

$$T_{S,I}^{in}(t_d) = \sum_{i=S}^{I} \tau_i(t_d) \tag{5.1}$$

$$T_{S,I}^{ex}(t_d) = \sum_{i=S}^{I-1} \tau_i(t_d + T_{S,i}^{ex}(t_d)) \tag{5.2}$$

To further motivate this research direction, a speed contour plot is presented for a section in freeway I-5S in California in Figure 5.1. This plot is constructed with loop detector data for a congested *Friday* of 2011. A few active bottlenecks can be seen in the site that start at different times and propagate upstream. Travel trajectories for instantaneous and experienced travel time approaches are constructed using the speed measurements at the fixed detectors. Space-time *(x,t)* points on the trajectories are calculated using Eq. 5.1-5.2 and replacing *I* and *S* with the corresponding section numbers. Figure 5.1 clearly shows the difference between instantaneous and experienced travel time by plotting a few vehicle trajectories for the two estimators. Note that these differences can be quite significant especially during the congestion onset and dissipation. This indicates that estimation of travel time should not be solely based on the traffic data collected in real time, but also the future recurrent traffic conditions should be integrated from historical data.
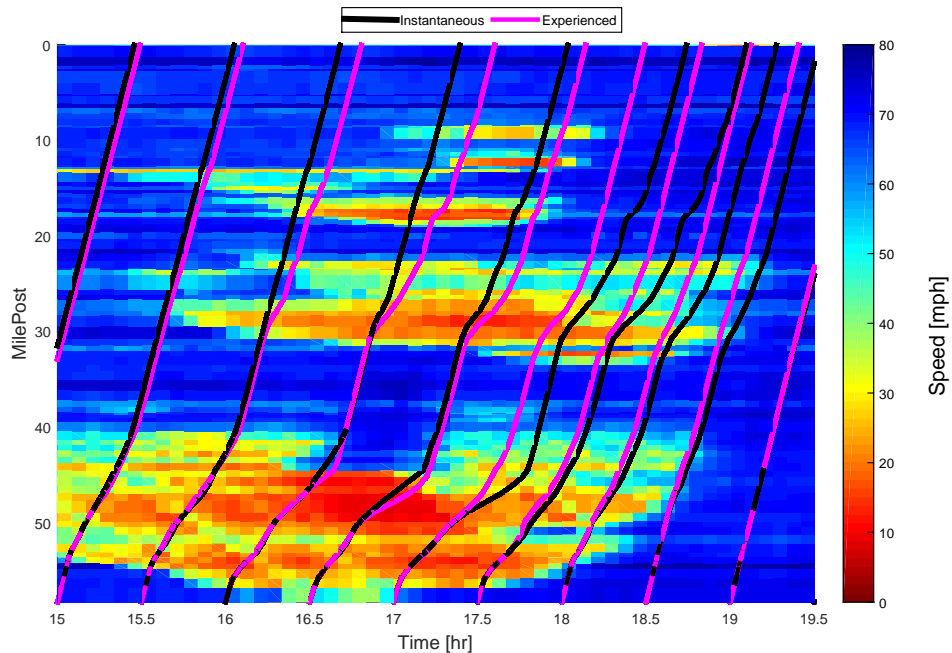


Figure 5.1: Speed contour plot and trajectories for a congested day (15:00-19:30) on I5-S freeway.

83

Nevertheless, simple historical average provides large estimation errors, which are due to the stochastic characteristics of traffic especially under congested conditions (formation of queues, demand uncertainty, etc.). To emphasize more the high variation of travel times from day to day and the high error of historical average as an estimator, Figure 5.2 plots for different departure times, the median, 10th and 90th percentile of experienced travel time for all *'Tuesdays'*, *'Wednesdays'* and *'Thursdays'* of one year in the same study site. From the time series graph, it is clear that even uncongested off-peak periods according to the median travel time value in the early afternoon have some significant probability to experience strong delays. Distributions for four different departure times are illustrated as well. Thus, there is a need for development of an accurate short-term traffic state prediction to be integrated in the estimation of experienced travel times.
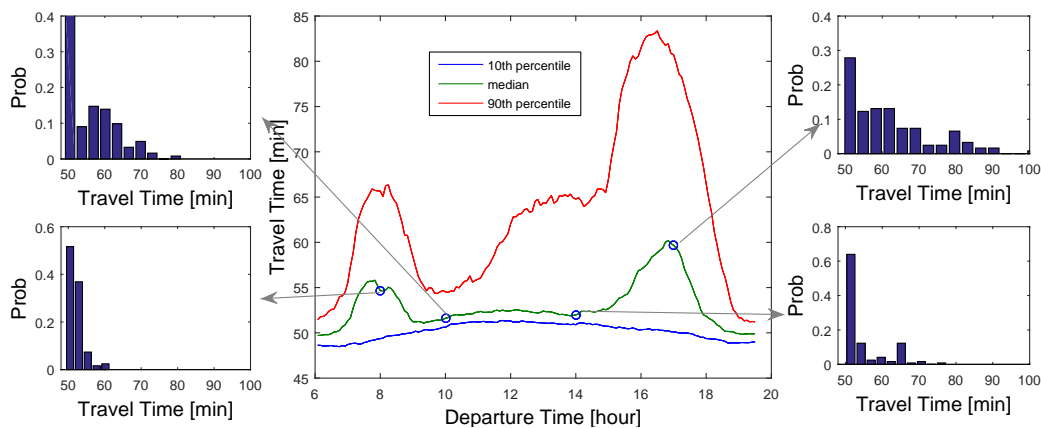


Figure 5.2: Median, 10th and 90th percentiles of experienced travel times for *'Tuesday-Thursday'* set in 2011 for different departure times in I5-S

The need for short-term traffic prediction led to the development of various forecasting algorithms. These methods can be broadly classified in two major categories; parametric methods (e.g. linear regression (Zhang and Rice, 2003), time series models (Yang, 2005; Min and Wynter, 2011), Kalman filtering (Okutani and Stephanedes, 1984; Van Lint, 2008)) and non- parametric methods (neural network models (Ledoux, 1997; Vlahogianni et al., 2005; Van Lint, 2006), support vector regression (Vanajakshi and Rilett, 2007), simulation models (Liu et al., 2006)). In the past years, neural network models have gained attention in transportation field and are frequently applied in traffic state prediction. The majority of transportation applications of neural networks is based on simple back-propagation algorithm (e.g. Adeli (2001)). However, other computing algorithms such as counter-propagation neural networks have the potential to improve prediction results (Dharia and Adeli, 2003). Additionally, many previous studies utilized macroscopic traffic flow models along with Kalman filtering technique to predict traffic states (Nanthawichit et al., 2003; Wang et al., 2008). Moreover, Fei et al. (2011), with a Bayesian framework, and Du et al. (2012), with an information fusion model, attempt to predict short-term travel time distribution considering the fact that the mean of a short-term

travel time distribution may not be an accurate tracking indicator.

Models where travel time is directly used as the state variable, suffer from the fact that travel time in the previous time interval is needed to predict future travel time. However, in practice, trip travel time is usually greater than the prediction interval. Therefore, experienced travel time in the previous interval cannot be used to calculate future travel times. On the other hand, instantaneous travel time, which does not consider congestion or speed evolution, is available at the departure time to predict instantaneous travel time in the following time interval. Data-driven approaches, which make use of instantaneous travel time, are consistent under some cases with the transitional physics of traffic flow and they are capable of constructing the underlying behavior of traffic without strong assumptions on its temporal evolution, see for example Jiang and Adeli (2004); Vlahogianni et al. (2005). However, data-driven approaches cannot explicitly infer knowledge from point measurements for estimating link performance measures (Vlahogianni et al., 2008). This is because spatiotemporal traffic flow dynamics are mainly governed by the queue formation and dissipation at point bottlenecks. Abrupt changes of traffic phenomena (e.g. lane changes, capacity drop, merge behavior, oscillations) can affect congestion development and propagation in various ways that require physical than statistical models to be explained, see for example Leclercq et al. (2011); Li et al. (2010a); Treiber et al. (2010). These characteristics of traffic's transitional behavior and the existence of variant traffic regimes may not be identified by statistically-oriented or data-driven approaches and increase their estimation and prediction errors.

In contrast to the aforementioned existing methodologies, the approach presented in this chapter benefits from the available traffic flow essentials (e.g. shockwaves, bottlenecks). The proposed method makes use of both historical and real time traffic information to provide travel time prediction. Instead of identifying traffic flow patterns using statistical methods (that sometimes might not succeed to capture complex phenomena of traffic flow), we propose to integrate in the methodology, identification of traffic patterns with traffic flow theory fundamentals, for example with shockwave analysis and bottleneck identification. The aim of the proposed model is to respond to changes in traffic pattern in real-time and to provide the expected 'experienced' travel time reflecting both a priori knowledge (i.e. historical dataset) and real-time traffic data through an incremental learning approach.

Considering the lag associated with experienced travel time in real-time information, this chapter focuses on congestion evolution to develop spatiotemporal traffic state maps and to construct predicted travel trajectories on them, leading eventually to predicted experienced travel time and the travel trajectory. The current and historical speed data are utilized in the prediction framework. However, due to the strong variability and difficulty in predicting them, speed data is first processed through an algorithm that would identify congested space-time domains. This approach, therefore, prefers to consider congestion evolution rather than speed evolution, because of the ease in detecting congestion evolution pattern in real time. An existing bottleneck identification algorithm is utilized to determine the location and spatial extent of the bottlenecks (Chen et al., 2004). The algorithm is used in this study

to store the major traffic events likely to be observed on the roadway (in historical data) and to track real-time conditions (in current data). Using the shockwave phenomena and identified bottleneck locations in real-time, the impact of a bottleneck can be predicted before it completely develops. Historical information can be useful to determine the characteristics of the bottlenecks (i.e. spatial extent and duration) and so, predict their impacts. Nevertheless, as we will show later, traffic conditions significantly vary from day to day (even for similar demand conditions) and as a result the size of a bottleneck in the time-space domain and travel speed of vehicles in this domain experience strong fluctuations. Hence, a simple prediction based on historical average or a partitioning of traffic conditions based on days (weekdays-weekends) or times of day (AM or PM peak) might introduce significant estimation errors.

This study partitions the historical dataset in clusters with similar characteristics based on the traffic patterns observed in the roadway. The building block of the methodology is the development of stochastic congestion maps, which identify the probability that a time-space domain is congested. This probability might have strong fluctuations for days with significantly different level of congestion and as a result it can decrease the performance of the prediction. Hence, this study develops a congestion search algorithm to revise the state of a priori knowledge according to the newly available information in real-time. Finally, the proposed approach develops a speed profile to construct travel trajectories using the predicted congestion evolution pattern and to calculate expected experienced travel time downstream for a given starting trip time.

The remainder of the chapter is organized as follows: Section 5.2 develops the methodological framework, Section 5.3 and 5.4 provide an implementation of the methodology in a real case study by utilizing one year of data for a congested Californian freeway (I5-S), while Section 5.5 includes discussion and future directions.

## 5.2  Methodology

The methodological framework of this approach includes (i) an existing bottleneck identification algorithm, (ii) clustering of data in traffic regimes with similar characteristics, (iii) development of stochastic congestion maps, (iv) an online congestion search algorithm, which combines historical and real-time data, and (v) a simple speed profile. While the proposed methodology uses loop detector data, it is not constrained to other sources of data (e.g. GPS data), given that this data can be utilized for bottleneck identification.

### 5.2.1  Bottleneck identification algorithm: a review

Chen et al. (2004) developed an algorithm to automatically identify bottleneck locations, their activation and deactivation times, and their spatial extents using loop detector data and focusing on speed measurements. Our methodology, which is described in the following sections, is not constrained by the specific algorithm. This method compares each pair of

adjacent detectors and determines the existence of bottleneck between them when

- Speed difference between upstream and downstream detectors is above the minimum speed differential, $\Delta v_{\min}$ threshold.

- Speed at upstream detector is below the maximum speed threshold, $v_{\max}$.

Chen et al. (2004) choose values of $v_{\max}$=40mph and $\Delta v_{\min}$=20mph with data aggregated at 5min intervals taken from California freeways. These parameters may need to be adjusted depending on the application. Wieczorek et al. (2010) discusses the effect of parameters on the model results, develops assessment criteria to select the optimal configuration of parameters, and by using the results of the bottleneck identification algorithm, maps recurrent congestion in time and space. Note that the algorithm cannot identify exact location of active bottlenecks, as it is limited by the nature of available data (e.g. loop detectors). The algorithm declares that there is an active bottleneck between two loop detector locations, and we constantly associate it with downstream detector location in order to be consistent. As long as the spacing between detectors is not very long, this assumption is not expected to cause significant errors.

The algorithm has also an offline part to identify the sustained bottleneck locations. This part smoothens the results of the online part, and fills in the small gaps in bottleneck detection at a particular detector. Basically, if several consecutive time periods are identified as bottleneck points, but one point in the middle failed to be identified so, offline part fills in this gap. Considering the fact that deactivation and reactivation of a bottleneck is not possible in such a short period of time, offline part accounts for the problems that could arise from the selection of parameters or missing data. The congested region affected by an active bottleneck can be defined using the speed measurements at upstream locations. This description is slightly modified in our estimation from the original algorithm. A congested region associated with a bottleneck ends at the detector location where two consecutive upstream detectors have more than $v_{\max}$, while a single detector with more than $v_{\max}$ is sufficient to enclose the congested region in the original algorithm. We note that with this alternation, the methodology provides better identification especially in the offset of congestion.

Identification of congested sections in an automated way allows to restore the major traffic events that occur on the roadway (in historical data) and to keep track of traffic conditions in real time (in current data). However, since the algorithm has an offline part, it is not possible to smooth the results in real-time. The following methodological parts are not constrained to the specific bottleneck identification algorithm or data. The choice was based on the small computational effort, combined with its proved accuracy to estimate congested conditions. Any type of data and algorithm that can provide accurate bottleneck locations and formation of queues in a time-space domain can be directly utilized in the remaining of the chapter.

## 5.2.2   Clustering of days with similar traffic patterns

Historical traffic patterns are crucial to the development of a travel time prediction framework due to the recurrence of traffic events. To use the historical dataset in a useful and efficient manner, days with similar traffic patterns (i.e. speed profiles) should be identified to decrease the randomness of traffic conditions. Otherwise large variations and temporal bias might be experienced by utilizing very heterogeneous data. Clustering techniques have been already used in transportation field to analyze traffic flow patterns, see for example Weijermars and Van Berkum (2005) and Ji and Geroliminis (2012). Since travel times are computed using local velocity measurements in this study, time-dependent speed measurements along the roadway can be used in the clustering step. Note that traffic speed data is processed to identify congested space-time domains, and the current and historical congestion information is the input to travel time prediction framework. Considering the high variability in traffic speed data and the difficulty in predicting them, our methodology mainly focuses on congestion evolution rather than speed evolution. In other words, this study uses a binary approach (i.e. congested or free-flow) to predict traffic states. However, as identification of traffic patterns involves only the use of historical data, this binary approach is not needed in clustering; traffic speed data can be directly used in the clustering step. Without clustering the variance of travel time for a given departure time is significantly larger and this has a direct erroneous effect in the prediction.

Time-dependent local velocity measurements at different locations and multiple days are the input to this step. Each variable in the input dataset represents a velocity measurement for a particular time period and a particular roadway section. Since a high number of sections on the roadway and time periods in a day lead to a large number of variables (e.g.  "180 time periods per day" $\times$ "89 roadway sections" = 16020 variables for the study site of the chapter), it is not straightforward to define a metric to compare and cluster days with strong similarity for the freeway route under consideration. Reduction in the variable size is required to proceed with clustering. Additionally, results of clustering can be corrupted by the noises in the original data (Milligan, 1980); speed data itself may contain noises which are not useful to explain general trends. Thus, clustering may benefit from a preprocessing step of variable selection which results in de-noising of data. Few leading principal components identified by Principal Component Analysis (PCA) can help us remove the noise in the dataset, reduce the dimensions and so apply more rigorous clustering techniques. First, original dataset is processed through PCA to reduce its dimensions and to remove the noise. Second, resulting principal components are used to cluster the days with similar traffic patterns.

**Principal component analysis (PCA)**

PCA is a well-established technique to reduce the dimensions of the dataset and to compress the data, see for example Nagendra and Khare (2003). PCA, using the orthogonal transformation, converts a set of observations with correlated variables into a set of observations with linearly uncorrelated variables, which are called principal components (PC). In other words, it

transforms the data into a new space which has most of the information (or energy) of the original data, but with a lower dimension. It lists PC's in the descending order of the variance associated with them. With respect to our problem, a freeway route might have detectors installed every a few hundred meters that provide speed and flow measurements every a few minutes. This creates an immense data set that cannot be directly utilized to cluster different days.

Suppose matrix $\mathbf{X}(m \times n)$ is the original data set with rows corresponding to observations (e.g. different days) and columns corresponding to variables (e.g. time- and space-dependent local velocity measurements). The variables are correlated, and there exists another set of uncorrelated variables $S_{m \times n}$, which is a linear combination of $X_{m \times n}$;

$$\mathbf{S} = \mathbf{P} * \mathbf{X} \tag{5.3}$$

where $P$ is the $(m \times m)$ projection matrix. The aim of PCA is to find a projected space $S$ whose covariance matrix is diagonal, or in other words variables are uncorrelated. The first row in the projection matrix represents a new axis in the uncorrelated set, and the resulting values from the first row create a new variable whose variance is the maximum among all possible choices (i.e. first principal component). The second row has the same properties for the set without the first principal component; so on and so forth.

PCA algorithm can be summarized as follows;

1. Subtract from each element the mean value of the corresponding column:
   $X[j, i] \leftarrow X[j, i] - E_i[\mathbf{X}]$

2. Compute correlation matrix $\mathbf{C} = \mathbf{X}^T \mathbf{X}$

3. Compute eigenvalues $|\mathbf{C} - \lambda_i \mathbf{I}_n| = 0, i = 1, 2.., n$

4. Compute eigenvectors $\mathbf{C} e^i = \lambda_i e^i$

5. Choose $p < n$ eigenvectors: $e^1, \dots, e^p$ with $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$

6. Project data into new space $\mathbf{S} = \mathbf{P} * \mathbf{X}, \mathbf{P} = \begin{bmatrix} e_1^1 & \cdots & e_1^p \\ \vdots & \ddots & \vdots \\ e_n^1 & \cdots & e_n^p \end{bmatrix}$

The eigenvalues represent the energy or the variance of the dataset along the eigenvector directions. The cumulative energy value $e$ for the $n$-th eigenvector is the sum of the energy across the eigenvectors from 1 to $n$, which can be formulated as $e[n] = \sum_{i=1}^{n} \lambda_i$. While selecting a subset of eigenvectors as the basis vector, the goal is to choose the minimum number of components that achieves reasonably high value of $e$ on a percentage basis. The number of components $p$ can be determined using the inequality $e[p]/e[n] > \kappa$ and setting an accuracy level $\kappa$ (e.g. 0.95).

**Gaussian mixture modeling (GMM)**

After reducing the dimensions of the dataset, Gaussian Mixture Model (GMM) is applied to create clusters in the historical dataset. GMM is the combination of multivariate normal density components, and it estimates normal distribution parameters using expectation maximization (EM) algorithm. GMM is often used for clustering purposes, and unlike other clustering methods, it is not solely based on the distance between the observations, but it is based on the distribution of data points. GMM is a more appropriate method than $k$-means clustering, when clusters have different sizes and correlation within them, which is the case for traffic data.

Consider now $(m \times p)$ matrix $S = \{s_j^i\}_{j=1..p}^{i=1..m}$ with rows corresponding to observations (e.g. days) and columns corresponding to the principal components. The probability density function (PDF) of $S$ will be modeled as a mixture of $K$ Gaussian distributions.

$$p(S) = \sum_{k=1}^{K} \alpha_k \, p(S|\mu_k, \Sigma_k) \tag{5.4}$$

where PDF of each Gaussian: $p(S|\mu_k, \Sigma_k) = N(\mu_k, \Sigma_k)$,
mixing coefficients: $\sum_{k=1}^{K} \alpha_k = 1$,
mean and covariance matrices of Gaussian $k$=1,..., $K$: $\mu_k, \Sigma_k$,
probability that the data is explained by Gaussian $k$: $\alpha_k = p(k) = \sum_{i=1}^{m} p(k|s^i)$.

The parameters of GMM are the means, covariance matrices and mixing coefficients;

$$\Theta = \mu_1, \ldots, \mu_k, \Sigma_1, \ldots, \Sigma_k, \alpha_1, \ldots, \alpha_K. \tag{5.5}$$

However, the parameters of GMM cannot be directly estimated because of unobserved latent variables. EM is an iterative method to find maximum-likelihood estimates of the parameters, where the model depends on unobserved latent variables (Dempster et al., 1977). EM attempts to find the optimum of the likelihood of the model given the data;

$$\max_{\Theta} L(\Theta|S) = \max_{\Theta} p(S|\Theta) \tag{5.6a}$$

$$\max_{\Theta} p(S\Theta) = \max_{\Theta} \prod_{i=1}^{m} \sum_{k=1}^{K} \alpha_k \cdot p(s^i|\mu_k, \Sigma_k) \tag{5.6b}$$

$$\max_{\Theta} \log p(S\Theta) = \max_{\Theta} \sum_{i=1}^{m} \log \left( \sum_{k=1}^{K} \alpha_k \cdot p(s^i|\mu_k, \Sigma_k) \right) \tag{5.6c}$$

At each estimation step $l$, parameters are updated as;

$$\alpha_k^{(l+1)} = \frac{1}{m} \sum_i p(k|s^i, \Theta^{(l)}) \tag{5.7a}$$

$$\mu_k^{(l+1)} = \frac{\sum_i p(k|s^i, \Theta^{(l)}) s^i}{\sum_i p(k|s^i, \Theta^{(l)})} \tag{5.7b}$$

$$\Sigma_k^{(l+1)} = \frac{\sum_i p(k|s^i, \Theta^{(l)})(s^i - \mu_k^{(l+1)})(s^i - \mu_k^{(l+1)})^T}{\sum_i p(k|s^i, \Theta^{(l)})} \tag{5.7c}$$

The parameter set $\Theta$ is updated iteratively until the log likelihood is increased by less than a certain threshold value.

Optimal number of clusters (K) can be determined by the use of average silhouette width (Rousseeuw, 1987) or information measures such as Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Silhouette is a common technique to validate clusters of data (Kaufman and Rousseeuw, 2009); it provides a distinct silhouette width value representing how well each observation belongs to its cluster and how dissimilar it is from the other clusters. In this work, optimal number of clusters will be chosen according to the average silhouette width values that they produce. Although an explicit formulation is not provided here, interested reader can refer to (Rousseeuw, 1987). In addition, the stability of clustering results is crucial to the selection of optimal number of clusters. GMM, whose initialization is random or based on $k$-means results, should return the same results every time it is repeated to ensure the accuracy and the robustness of the algorithm. Therefore, by varying the number of components in clustering, we can identify the optimal value that creates both high partitioning performance and robust clusters.

### 5.2.3 Stochastic congestion maps

To add a probabilistic flavor in the formation of bottlenecks from day to day that will contribute in the accurate prediction of experienced travel times, we introduce a new physical concept of describing spatiotemporal traffic patterns, the *stochastic congestion map*. It represents the likelihood of congestion occurrence at a given space-time point based on the observations for many days of the same cluster in the historical dataset. The probability of observing congestion at roadway segment $i$ at time interval $t$ is calculated separately for each cluster $k$.

$$p^k(i,t) = \frac{1}{D_k} \sum_{m=1}^{D_k} f(i,t,m) \tag{5.8a}$$

$$f(i,t,m) = \begin{cases} 1 & \text{if segment } i \text{ is congested at time } t \text{ on day } m \\ 0 & \text{otherwise} \end{cases} \tag{5.8b}$$

where $D_k$ is the number of days in cluster $k$. Function $f$ is estimated with the bottleneck

algorithm of Section 5.2.1. We consider the tool of stochastic congestion maps appropriate for various research methodologies, such as travel time reliability and predictive control.

Once cluster analysis and bottleneck identification algorithm are applied, a stochastic congestion map showing the probability of congestion occurrence at a particular space-time domain is estimated for each cluster. Two arbitrary clusters are shown in Figure 5.3a and c. Each cluster is divided into subsets using certain threshold probability values (e.g. from 0.05 to 1), and blocks corresponding to a certain likelihood of occurrence is created for each probability value. Blocks associated with the lowest threshold are constructed by the congested points with probability greater than 0.05 (i.e. occurring more than 5% of the analyzed days), blocks associated with second lowest threshold is constructed by the points with probability greater than 0.10, and so on so forth (like a cumulative 2D distribution). Thus, heavy congestion observations are associated with low threshold values, while light congestion situations are associated with high threshold values. Figure 5.3b and d represent the blocks along with threshold values associated with two clusters having different levels of congestion. Note that the block associated with the lowest threshold is a superset of all the other blocks, the one associated with second lowest is a superset of the blocks associated with third and higher lowest thresholds, so on so forth. In other words, light colored dots (e.g. green dots; block associated with the lowest threshold) in Figure 5.3b and d do exist beneath dark colored dots (e.g. any dot darker than green; blocks associated with second and higher lowest thresholds), although they are not visible. In addition, the difference between blocks is roughly in the shape of asymmetric rings around a given core, which can be considered as separate bottlenecks. These results are of great importance to our analysis because they show that even the location and duration of bottlenecks are roughly known a priori, a more careful look identifies strong stochastic phenomena that can vary travel times from one day to the other, which is also observed in Wieczorek et al. (2010). For example, bottleneck #5 in Figure 5.3b starts at location with milepost 40, but its extension in time and space varies from day to day. For the remainder of the chapter, bottleneck will refer to a varying spatiotemporal congested area caused by a particular active bottleneck location, and block will refer to a bottleneck subset that corresponds to certain probability of occurrence. In Section 5.3.3, we discuss an alternative method to construct stochastic congestion maps by introducing physical bounds on shockwave speeds.

The proposed approach does not distinguish between recurrent and non-recurrent phenomena because of two reasons. First, the available incident dataset does not allow us to quantify their effects on traffic conditions. Second, traffic prediction in non-recurrent conditions requires a different approach considering incident characteristics, such as the clearance time of the incident, the loss of capacity and others. Newell's 3-detector (Newell, 1993) model can be one of the ways to incorporate non-recurrent conditions in this framework. Based on the vehicle counts at two detector stations, vehicle counts at some intermediate location are computed. However, Newell's model cannot tolerate ramp flows between the two detectors. With respect to our problem, in case of unexpected congestion detection, the 3-detector model can be activated to determine the bottleneck extent. However, such a model would require

(a)

(b)

(c)

(d)
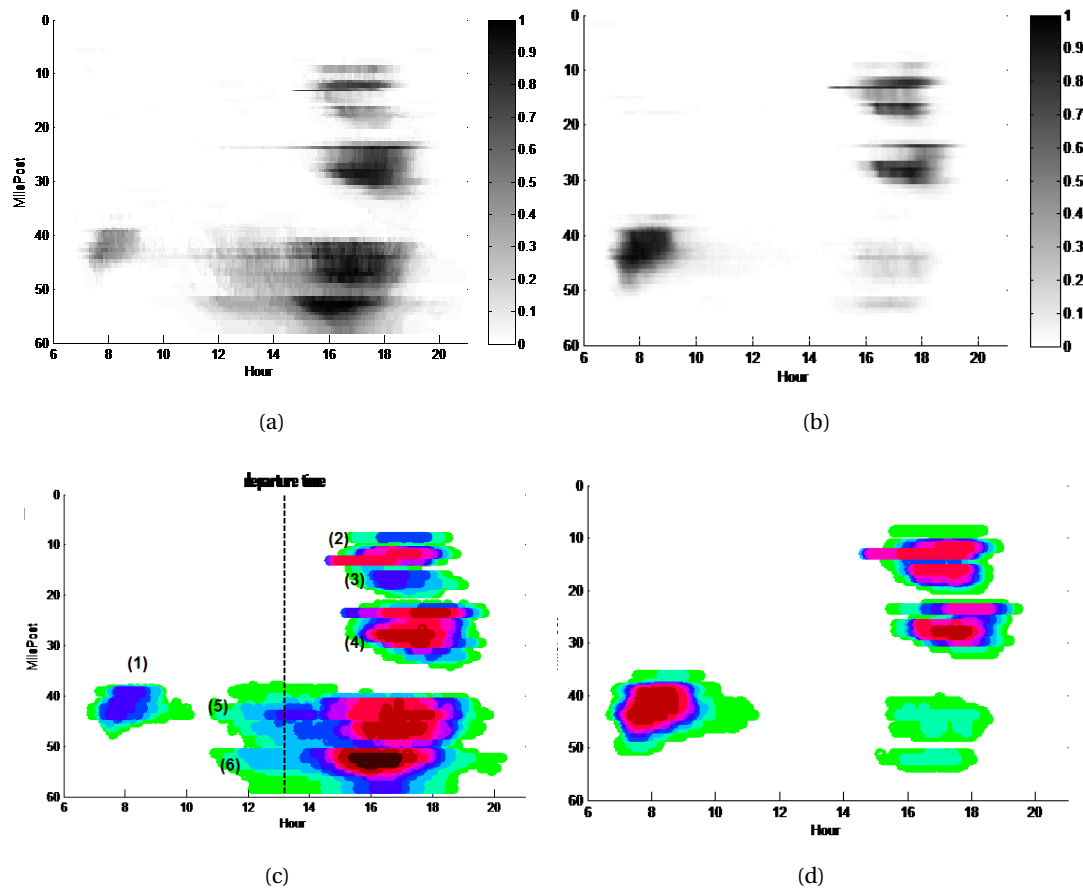
Figure 5.3: (a)-(b) Stochastic congestion maps, (c)-(d) blocks in stochastic congestion map,

prediction of incident duration (i.e. duration in which bottleneck is active), prediction of departure curve at the bottleneck location and prediction of arrival curve at downstream of the freeway (right before the closest ramp to the bottleneck location). In our case, non-recurring events are indirectly addressed through stochastic congestion maps with probability values as low as 0.05, which is not very recurrent. For example, if an accident creates much higher delays than recurrent bottlenecks (as this is expressed by low values of thresholds in the stochastic congestion maps), the travel time prediction can integrate some component of this oversized delay. However, note that travel time for a specific departure time can significantly vary even under recurrent conditions. To address the travel time variability and to provide more accurate travel time information, an online congestion search algorithm is developed in this study that allows us to switch between blocks and congestion evolution patterns represented by them.

### 5.2.4   Online congestion search algorithm

The aim of the online congestion search algorithm is to provide the connection between real-time and historical information. Binary information (congested or not congested) col-

lected in real-time is compared with the possible bottleneck shapes (i.e. blocks in stochastic congestion map) in the historical database, and the best representing block is selected. Note that stochastic congestion map, which incorporates continuous probability values, is already discretized to construct blocks, which consist of binary congestion data and are comparable with real-time information. Before delving into the details of the algorithm, note that there are multiple cores (bottlenecks) in the congestion map. In addition, although there may be a correlation between the size of the bottlenecks for a given day, it is intuitive that they are not highly dependent, as they occur at different locations and the propagation of the one does not significantly affect the propagation of the others. Therefore, the congestion map is divided into six separate bottlenecks as shown in Figure 5.3b, and the congestion search algorithm is applied separately for each of them to select the threshold value that best represents the real-time traffic conditions on the spatiotemporal area covered by the corresponding bottleneck.

The idea behind the online congestion search algorithm is to provide travel times based on the expected traffic conditions at the very beginning and integrate real-time congestion information to specify the shape of the bottleneck in future time periods. If there is no real-time information about the given bottleneck at the departure time, the expected threshold value (i.e. probability of 0.5) is used to compute the predicted travel time. However, if the departure time is later than the starting time of a downstream bottleneck, then we are able to compare real-time traffic information with congestion maps for particular bottlenecks (e.g. for a departure time at 13:00 in Figure 5.3b bottlenecks #5 and #6 are already active and thresholds can be updated). The algorithm determines the threshold value or block that would best represent the real-time congestion information obtained till the departure time. The threshold value or block estimated by the congestion search algorithm for each bottleneck is applied to construct the predicted congestion map, and travel time for the given departure time is computed on this predicted congestion map. A graphical representation of the algorithm is provided later.

For a given bottleneck, selection of threshold value is done by the following similarity metric;

$$\max_p \big(g(p,j,k) + h(p,j,k)\big) \tag{5.9a}$$

$$g(p,j,k) = TP(c_{p,j}^k, RTI) \tag{5.9b}$$

$$h(p,j,k) = TN(c_{p,j}^k, RTI, c_{1,j}^k) \tag{5.9c}$$

$p$ is the threshold (block) index,
$j$ is the bottleneck index,
$k$ is the cluster index,
$RTI$ real-time information (the set of congested points observed till the departure time),
$c_{p,j}^k$ is the set of points defined by bottleneck $j$, threshold $p$ and cluster $k$,
$TP$ (true positive) is the number of correctly classified congested points of $RTI$ by $c_{p,j}^k$,

$TN$ (true negative) is the number of correctly classified non-congested points of $RTI$ by $c_{p,j}^k$ in the set defined by $c_{1,j}^k$.

Note that the set of points used to determine $TN$, is defined by the maximum size that the given bottleneck can get (i.e. block associated with lowest threshold; $c_{1,j}^k$). Maximum size of the bottleneck is needed to identify the basis where number of rightly classified non-congested points can be calculated. Since real-time information is available only till departure time, the points up to the departure time are used in the search mechanism. In addition, a moving time window of two hours is used to keep track of varying conditions. If the same performance value is computed for multiple threshold values, then the one that is closest to the expected threshold value is chosen. This indicates that the algorithm always selects the conditions that are most likely to be observed.

### 5.2.5 Online cluster switch

Partitioning of historical dataset may give some hints about how to assign the days to the clusters in a predetermined way. If a cluster mainly consists of particular days of the week (e.g. Mondays only), the corresponding days can be pre-assigned to that cluster. However, it is very likely that the cluster might contain other days of the week. Hence, traffic prediction should be done in a flexible way that allows the algorithm to switch between clusters. On the contrary, the choice of cluster should not change at each time step, because it may bring unrealistic fluctuations in travel time. It should be considered as a tactical level decision, while the choice of threshold (block) is an operational level decision.

Therefore, online cluster switch is implemented if the following condition applies;

$$TP(c_{1,j}^{alt}, RTI[t-2, t]) > \alpha \cdot TP(c_{1,j}^{pre}, RTI[t-2, t]) \tag{5.10}$$

where $c_{1,j}^{alt}$ and $c_{1,j}^{pre}$ are the biggest blocks of an alternative cluster and the predetermined cluster, respectively. $RTI[t-2, t]$ is the real-time information with a time window of two hours, $\alpha > 1$ is an empirical coefficient that restricts strong fluctuations in the choice of cluster.

In other words, the algorithm switches to the alternative cluster when the number of congested points identified by the alternative cluster is $\alpha$ times greater than the ones identified by the predetermined cluster. Note that cluster switch is implemented separately for each defined bottleneck, and the sensitivity analysis for the parameter $\alpha$ is presented in Section 5.4.2.

Note that in Figure 5.3b and d, which present the blocks for different clusters, there are many congested points in one cluster that do not exist in the other. For instance, if the predetermined cluster (Figure 5.3d) is not able to address for the congested points observed in real-time (e.g. left part of the bottlenecks 5 and 6 in Figure 5.3b) and Eq. 5.10 is satisfied, then the algorithm switches to the alternative cluster. The cluster switch can also address a component of non-recurrent events by switching to a more congested cluster.

### 5.2.6 Speed profile

The developed methodology to predict the experienced travel times requires two pieces of future information (i) prediction of congestion development and propagation on the roadway (e.g. bottlenecks) and (ii) prediction of speed profiles. By the methodology described so far, the future speed profile inside and outside the bottleneck time-space domain is still unknown in the beginning of a trip. To estimate the trajectory of a vehicle which runs in the predicted time-space domain of a congestion map, different speeds are considered for congested, $V_c$, and uncongested conditions, $V_f$.

Speed information at the departure time can bring valuable insight into this problem and estimate $V_c$ and $V_f$.

$$V_c = \frac{1}{N_c} \sum_{i=1}^{N_c} v_i^{t_d} \quad \text{and} \quad V_f = \frac{1}{N_f} \sum_{i=1}^{N_f} v_i^{t_d} \tag{5.11}$$

where $t_d$ is the departure time, $v_i^{t_d}$ is the speed measurement on roadway segment $i$ at $t_d$, $N_c$ is the number of sections registered as congested at $t_d$, $N_f$ is the number of sections registered as uncongested at $t_d$.

However, the speed profile defined by Eq. 5.11 does not account for the speed variability along the roadway and it can produce erroneous results, especially when congested speeds vary with different roadway sections. Therefore, the congested speed, $V_c$, is modified for each roadway segment $r$ as follows. If a particular roadway section $r$ is registered as congested at the departure time, $t_d$, and if it remains congested in the predicted congestion map, its speed at the time of arrival $t_a$, estimated by Eq. 5.2 for time $[t_d + T_{S,r}^{ex}(t_d)]$. Speed $v_r^{t_a}$ is the average of the congested speed and the speed of the section at the time of departure, i.e. $v_r^{t_a} = \frac{1}{2}(V_c + v_r^{t_d})$. In this way, local congestion phenomena are integrated in the prediction and results are further improved. A more detailed analysis could estimate $V_c$ as a function of the threshold index $p$ for each bottleneck.

## 5.3 Case study

For the application of the methodological framework, data from PeMS is used. PeMS collects 30-sec loop detector flow and occupancy data throughout the Californian state. Then, it processes them and fills in the missing detector data to compute 5-minute flow, occupancy and speed averages Chen et al. (2001). For this study, a 60 mile section of I-5S in the district of San Diego/Imperial is selected, between mileage 0 and 60. A 20 miles portion of the freeway, along with the detector locations is presented in Figure 5.4. Considering the detector quality and the effect of strong recurrent congestion in multiple locations, the selected roadway section is a challenging study to evaluate the methodological framework. 5-minute loop detector data is collected through PeMS for the whole year of 2011. The dataset is divided randomly into two parts; Training Set (~80%-289 days) and Testing Set (~20%-76 days).
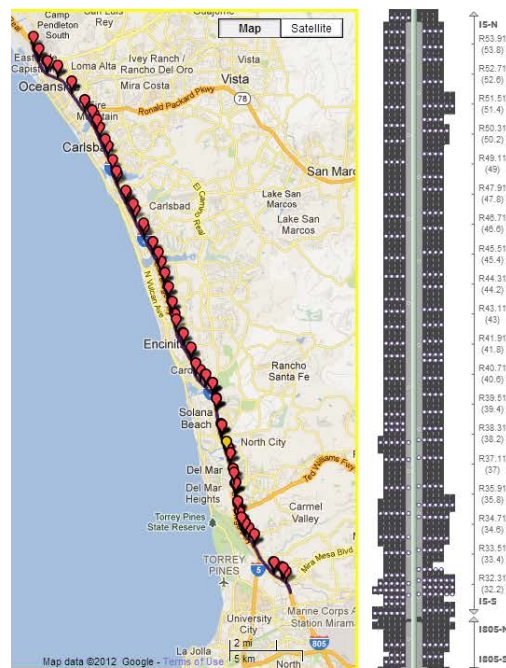
Figure 5.4: I-5 corridor in San Diego (source: pems.dot.ca.gov)

### 5.3.1 Bottleneck identification algorithm

Bottleneck identification algorithm is applied for the training dataset and stochastic congestion maps are estimated in the next steps of the methodology. Figure 5.5 presents the speed contour plot for a single day and the congested regions identified by the algorithm. It is clear that the identification is reliable both in the onset and offset of congestion.

### 5.3.2 Clustering of days with similar traffic conditions

This approach provides a way to cluster data based on their congestion profile. A qualitative approach (e.g. all Mondays are the same) is not appropriate given the stochastic characteristics of traffic especially under congested conditions (formation of queues, demand and capacity uncertainty, etc.). Clustering is crucial to the prediction performance, because it defines the library on which the learning scheme is implemented.

Before clustering, PCA, as described in Section 5.2.2, is applied to reduce the dimensions of the dataset and to remove the noise. PCA analysis shows that 100 principal components carry 95% of the variance in the original data of 16020 variables ("89 detectors"×" 180 5-min time periods" between 6AM and 9PM). Therefore, the rest of the clustering operation is carried out with the reduced dataset of 100 variables (components).

Average of silhouette width values in the data set can be used to determine optimal number of clusters. Figure 5.6 presents 100 realizations of clustering for several numbers of clusters, and

Figure 5.5: Identification of congested sections (20-Jul-2011)

the average silhouette width values that result from these realizations. The (vertical) range of average silhouette width values computed for the same cluster numbers clearly shows that results are stable only for three clusters. As the initialization of 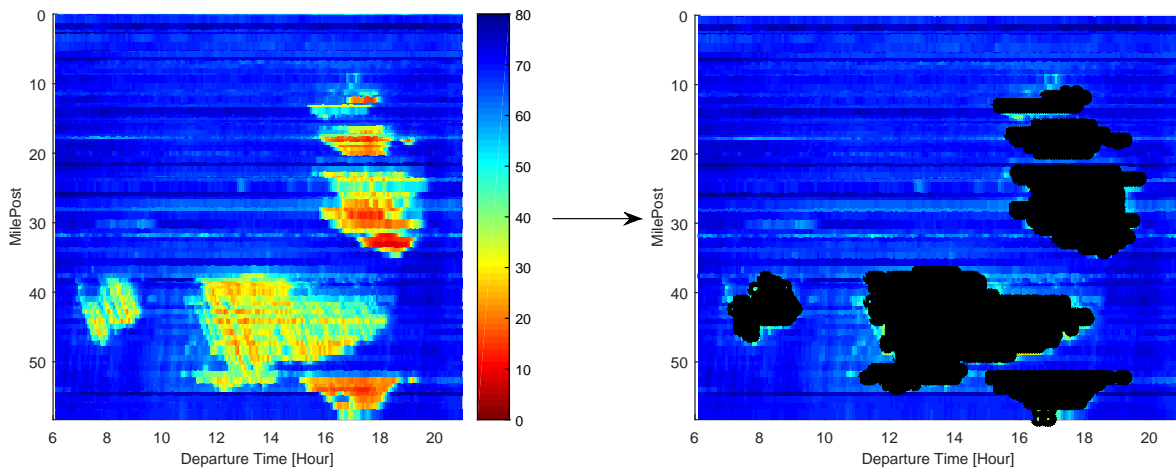GMM is random, each run of the algorithm can result in different clusters. Thus, the wide range observed in Figure 5.6 implies significant difference between the computed clusters, i.e. instability. In addition, mean average silhouette width (average of 100 realizations) reaches its maximum value at three clusters. Therefore, considering both the stability of the results and performance of partitions (i.e. average silhouette width), optimal number of clusters is selected to be three. We have noticed that by utilizing a higher number of clusters in the methodological framework, the experienced travel time prediction does not improve. Note that a similar analysis can be conducted with AIC and BIC values to determine optimal number of clusters.

Determining the number of clusters is often ambiguous, and it is a separate problem from actually solving the clustering problem. The optimal choice of clusters seeks a balance between the maximum compression of data using a single cluster, and the maximum accuracy by assigning each observation to a separate cluster. Suppose for example that the dataset consists of some dense and distant clusters (e.g. colors), whose number is not known a priori. If we set the number of clusters too low, the algorithm will combine some natural clusters (i.e. different colors) to reduce the total number of groups to the user-specified number of clusters. Additionally, since the initialization of the algorithm is random, each run of the algorithm can result in combination of different natural clusters, which causes instability. On the other hand, if we set the number of clusters too high, then some natural clusters have to be divided in an artificial way, in order to obey the specified number of groups, which also causes instability due to the randomness of artificial cuts. Therefore, correct choice of cluster number must be associated with best performance and stability. With respect to our problem, natural clusters may represent uncongested, moderately congested and highly congested traffic conditions. In that case, the clustering algorithm would perform at its optimum with the correct choice of

98
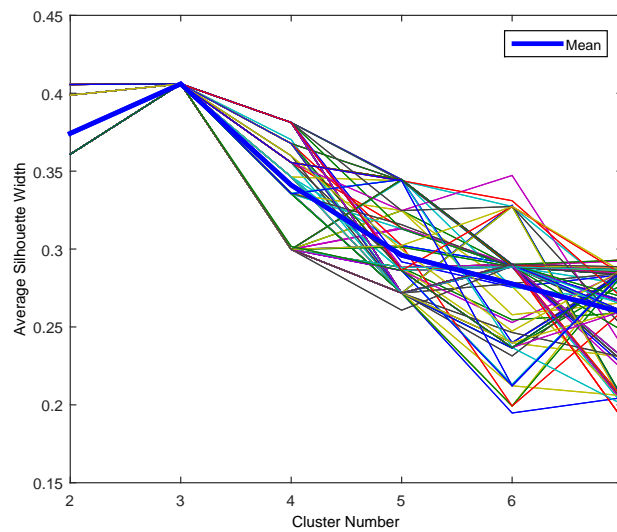
Figure 5.6: Average silhouette width vs. cluster number

three groups. However, another freeway where natural clusters can be described differently may require a different choice of the optimal number of clusters.

Figure 5.7 presents the distribution of the days along the first two principal components and the GMM results. Ellipses in Figure 5.7 represent 50% and 90% of the variance of the clusters along the dimensions of the two PC's. Note that the PC values do not have a physical meaning; they represent the values of the new uncorrelated variables. The results clearly show that the clusters are mainly dominated by certain features of days of the week, with some exceptions (about 15% of the days). The first cluster shown in Figure 5.7b is dominated by weekend days, the second by week days other than '*Fridays*', and the third by ' *Fridays*'.



Figure 5.7:  GMM results (a) based on days of the week, (b) based on clusters.

Table 5.1 presents the distribution of days along the clusters. Most of the week days classified in the first cluster are holidays that are not subject to a significant level of congestion. The second cluster is mainly composed of week days other than 'Fridays'. These days have significant level of congestion. However, the level of congestion is not as high as it is observed on the most congested cluster. Therefore, the clustering algorithm creates a separate cluster mainly for 'Fridays' (25% of this cluster includes days other than Friday). Although the clusters do not totally belong to a certain day of the week, this information is very useful to identify expected traffic conditions on the roadway for a particular day. Hence, each cluster is assigned to dominating days of the week in a predetermined way, and travel time prediction for a given day is executed within the corresponding cluster and its associated congestion map. The switch algorithm will provide the flexibility to switch to a more or less congested cluster if traffic conditions look very different than what was initially assumed.

Table 5.1: Distribution of days among the clusters
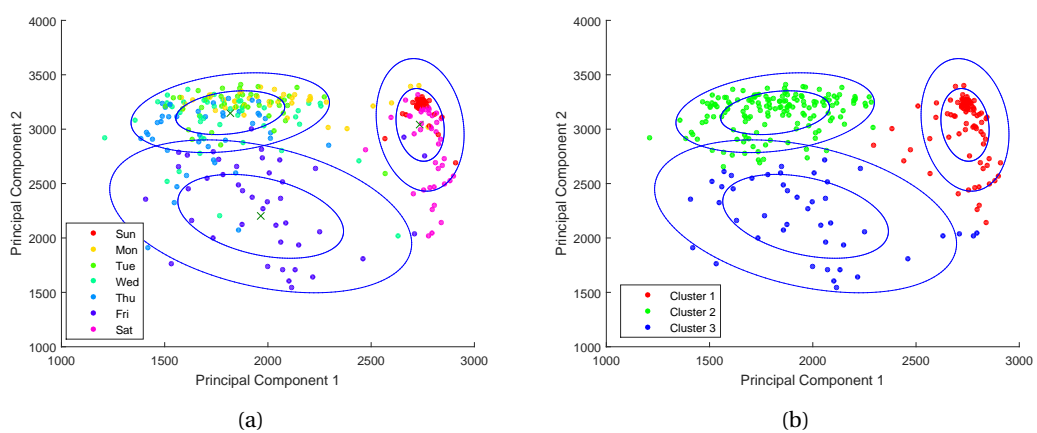
| Cluster | Days | | | | | | |
|---|---|---|---|---|---|---|---|
|  | Mon | Tue | Wed | Thu | Fri | Sat | Sun |
| 1 | 6 | 1 | 1 | 1 | 2 | 40 | 42 |
| 2 | 36 | 40 | 36 | 34 | 8 | 0 | 0 |
| 3 | 0 | 0 | 3 | 6 | 31 | 2 | 0 |

### 5.3.3 Stochastic congestion map

By combining the results obtained from bottleneck identification and clustering steps, stochastic congestion maps can now be created. They can be constructed by simply estimating the average number of congestion observation for a space-time point in each cluster. Figure 5.3(a) and 5.3(b) present the stochastic congestion maps for the third and the first cluster, respectively. Then, they are divided into blocks for different threshold probability values (see Figure 5.3(c) and 5.3(d).

Note that the shape of blocks in the congestion map may not be totally proper regarding the traffic flow essentials (e.g. shockwave speeds). This may be because of the flaws of the bottleneck identification algorithm or averaging different bottleneck spatio-temporal shapes. However, the purpose of this study is to predict experienced travel times or to predict the time spent in the queue, not to predict the shape of the bottlenecks. The important feature of an accurate methodology is to estimate with some confidence the number of congested points (in the time-space domain) that the predictive trajectory will intersect while travelling.

To test the influence of bottleneck shapes, stochastic congestion map is revised to incorporate physical bounds on shockwave speeds, and its effects on prediction results are investigated. For each bottleneck and for each subset represented by threshold probability values given in Figure 5.8(a), bottleneck shapes or congestion patterns have been modified to obey the shockwave speed bounds. In this study, maximum shockwave speed is calculated for a

triangular fundamental diagram and it is taken as 10 mph. Subset points are fitted, in the best possible way, into an area defined by polylines whose derivative is between 0 and 10 mph. Note that this modification leads to a certain change in the number of congested points defined by the subset, which is small indeed and does not change the accuracy of the algorithm. Revised congestion map for the $3^{rd}$ cluster, given in Figure 5.8(b), provides realistic and accurate bottleneck shapes compared to the original congestion map presented in Figure 5.8(a). Revised congestion map is evaluated in Section 5.4.1 as an alternative to the original approach.
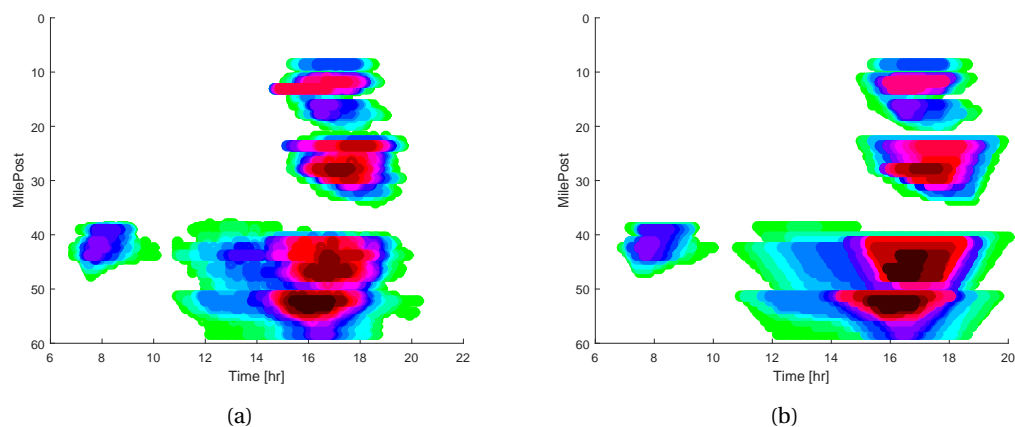


Figure 5.8: (a) Original stochastic congestion map, (b) revised stochastic congestion map.

### 5.3.4 Graphical representation of the algorithm

To further elaborate the implementation of the methodology, Figure 5.9 presents the mechanism of the algorithm on a particular day (*Wednesday*, 20-Jul-2011). Estimation of experienced travel times on this specific day is presented in the next section. Although the pre-assigned cluster for ' *Wednesday*'s indicates a moderate level of congestion, this particular day exhibits a higher level of congestion, which is a challenging case for our approach. Non-recurrent congestion on this particular day requires both cluster switch and adjustment of threshold values.

Figure 5.9 shows the results of the algorithm for bottlenecks #5 and #6 (refer to Figure 5.3(c) for the index of the bottlenecks). These two bottlenecks have very small size and very low probability in the pre-assigned cluster (see Figure 5.3(c) and 5.3(d)). Therefore, the algorithm switches to the highly congested cluster at 11:15am for the bottleneck #5, and at 11:30am for the bottleneck #6. Figure 5.9a, c and e show the speed contour plots and the congested points identified by the algorithm until the departure time. Three different departure times are shown, in the onset ($t_1$=12:00), during ($t_2$=14:30) and the offset ($t_3$=18:00) of congestion. Note that although the speed contour plot is given until 19:00, this information is not available at the departure time and the missing portion is shaded with a transparent gray rectangle.

(a)

(b)

(c)

(d)

(e)

(f)

(g)

Figure 5.9: Graphical representation of the algorithm (a)-(f)Current traffic information and prediction over the next 1-hr for different departure time periods, (g) chosen thereshold values for two bottleneck blocks

Figure 5.9b, d and f introduce the same information along with predicted congestion sections within the next one hour, and experienced and predicted trajectories for the trips that start at the specific departure time. Note that experienced and predicted trajectories are so similar that it is very difficult to distinguish between them especially in Figure 5.9d and f.

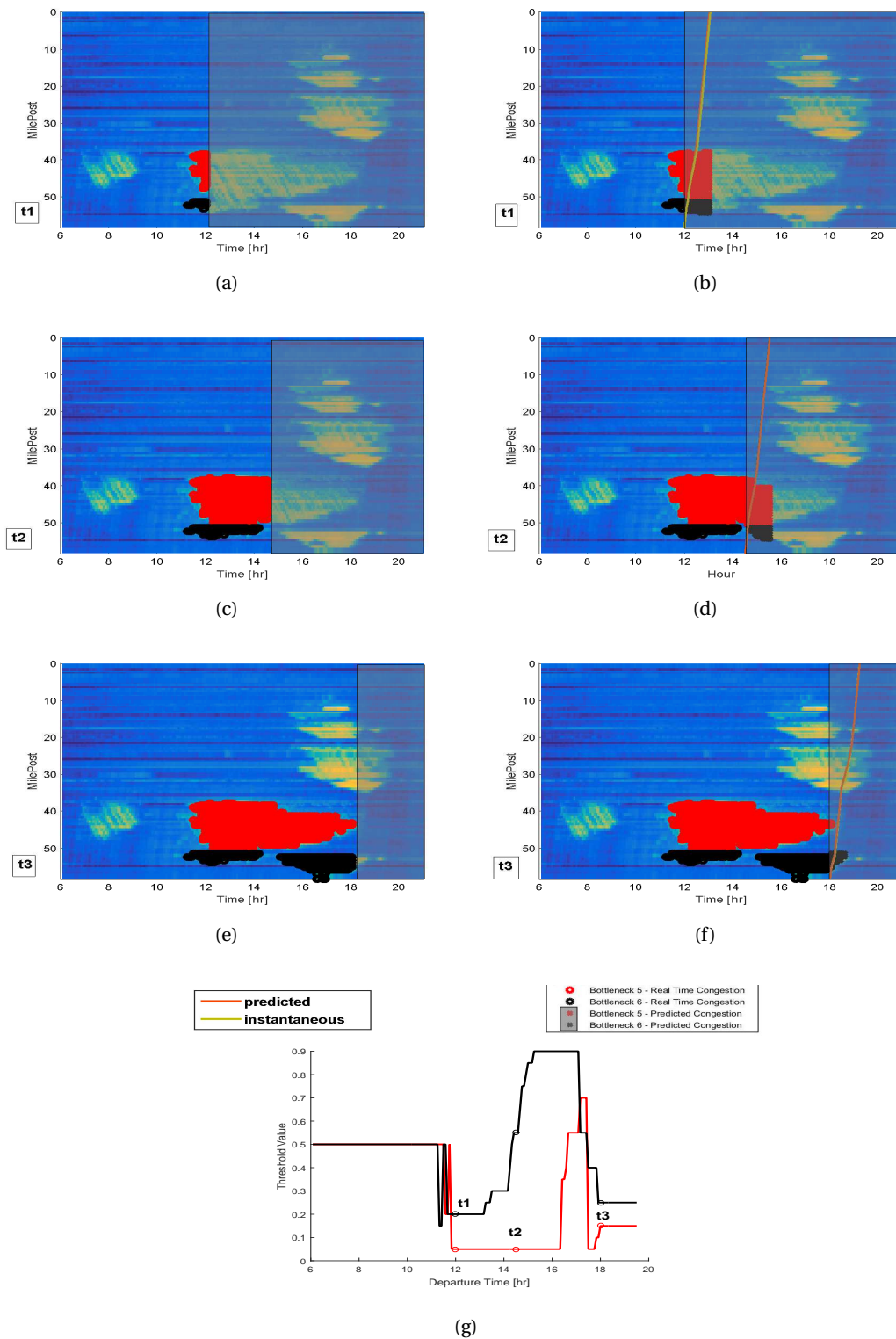Figure 5.9g introduces the results of the congestion search algorithm; threshold values for bottleneck #5 and bottleneck #6. Note that high values of thresholds represent lower levels of congestion, according to the definition in Section 5.2.3. Note that the output of the online congestion search algorithm is consistent with neighboring time periods. Threshold values are not subject to large variations (up-and-downs), which would represent sudden prediction changes. Figure 5.9g also gives the threshold values for the departure time periods $t_1$, $t_2$ and $t_3$. At 12:00 ($t_1$), bottlenecks are just starting to grow, and the congestion search algorithm chooses very low threshold values for both, considering the very early starting time of the congestion. Note that probability of having congestion at this departure time is very low even in the high congestion cluster (see Figure 5.3(a)). At 14:30 ($t_2$), the algorithm detects that bottleneck #6 has started to disappear and updates its decision by increasing the corresponding threshold value (from 0.3 to 0.5), while it insists on its decision for bottleneck #5. At 18:00 ($t_3$), the congestion search algorithm re-identifies congested sections around the bottleneck #6, it decreases its threshold value gradually to adjust the changes in the size and shape of the bottleneck. Note that while identified congested sections in real-time are shown for the whole bottleneck in the figures, a moving time window of the last two hours prior to the departure time is used in the search mechanism to adjust to rapidly changing traffic conditions.

## 5.4 Results

### 5.4.1 Evaluation of the proposed approach

The approach presented in this chapter (along with original stochastic congestion maps) is evaluated on the testing dataset (76 days). Since the weekend days are not subject to significant level of congestion, they are not considered in the evaluation step. Travel time, in this study, is computed using speed data from loop detectors and constant speed interpolation technique. For a link between two successive detectors, the speed measurement at downstream or upstream detector, or the average of two measurements can be used to represent the velocity. All constant speed interpolation methods imply instantaneous speed changes, which do not occur in real-time. However, considering the distance between the detectors (about 500m) in our study site, this phenomenon is not expected to largely affect the results. Kothuri et al. (2008) analyze travel time estimation errors that result from midpoint algorithm, and conclude that detector failure is the major cause of high estimation errors. Their analysis revealed that travel time estimates produced by the midpoint algorithm have a good accuracy compared with ground truth probe vehicle runs. Chen et al. (2003), where travel time is calculated from the speed measurements at single detectors, also indicate fair estimation accuracy. Considering the distance between the detectors (about 500m) and high detector quality in the study site,

midpoint algorithm is expected to produce accurate travel time estimates. Travel time can also be computed by using linear and quadratic speed interpolation methods, which do not require instantaneous speed changes. Alternatively, one could apply a more detailed traffic flow model (of first or higher order) to estimate speed between the detectors. Nevertheless, we do not expect the accuracy of the results to improve. Predicted travel time is calculated in the same manner as experienced travel time. However, instead of velocity field, which is unknown at the departure time, predictive trajectory travels through the predicted congestion map and uses an estimated speed profile as described by Eq. 5.11 to compute the time needed to traverse each segment. Historical travel times are also computed for each day of the week. The median value of the experienced travel times at a given departure time on a given day of the week is taken as historical average value.

To measure the effectiveness of the methods, two statistics, namely mean absolute error (MAE) and mean absolute percentage error (MAPE) are utilized;

$$MAE = \frac{1}{n} \sum_{t=1}^{n} |T(t) - \hat{T}(t)| \tag{5.12a}$$

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{T(t) - \hat{T}(t)}{T(t) - T_{free}} \right| \cdot 100 \tag{5.12b}$$

where $n$ is the number of observations (i.e. 5 min long departure time intervals) , $T_{free}$ is free-flow travel time, $T(t)$ is the experienced travel time computed with Eq. 5.2, and $\hat{T}(t)$ is the travel time provided by the methodology. Note that MAPE provides the percentage error in terms of delay.

Figure 5.10(b) presents MAE values of instantaneous and predictive travel time for 103 congested periods at least 30 min long. Since historical average performs clearly worse than the other methods, it is not shown in Figure 5.10(b). In 85 out 103 cases, predictive travel time methodology produces better results than instantaneous travel time assumption. Note that the performance of the proposed methodology is significantly better when conditions are more congested. Figure 5.10(b) presents the 103 congested periods irrespective of their lengths. To further elaborate this, congested periods are divided into two groups; periods where MAE is less or more than 2 min (see the rectangles in Figure 5.10(b)). Figure 5.10(a) provide the histogram of absolute errors for the two groups (considered as a whole) in a disaggregate way, which implicitly accounts now for the length of congested periods. Departure time intervals (of 5min) from each group are gathered together, and histogram plots are created using the error values associated with them. Note that although total number of congested periods in two groups is similar, number of departure time intervals (5 min long) within the periods is quite different (as indicated by the difference in total bin counts in Figure 5.10(a)-top and -down), which implies the difference in the length of congested periods in two groups. Average length of congested period is 180 min for MAE<2min and 275 min for MAE>=2min. Results indicate that short congested periods tend to be less problematic even under instantaneous travel time assumption. However, long congested periods are associated with high error for

the instantaneous estimation. Although error distribution in the first group is comparable for instantaneous and predictive travel time (Figure 5.10(a)-top), our approach outperforms instantaneous one in the congested group (Figure 5.10(a)-down).



(a)                                                              (b)
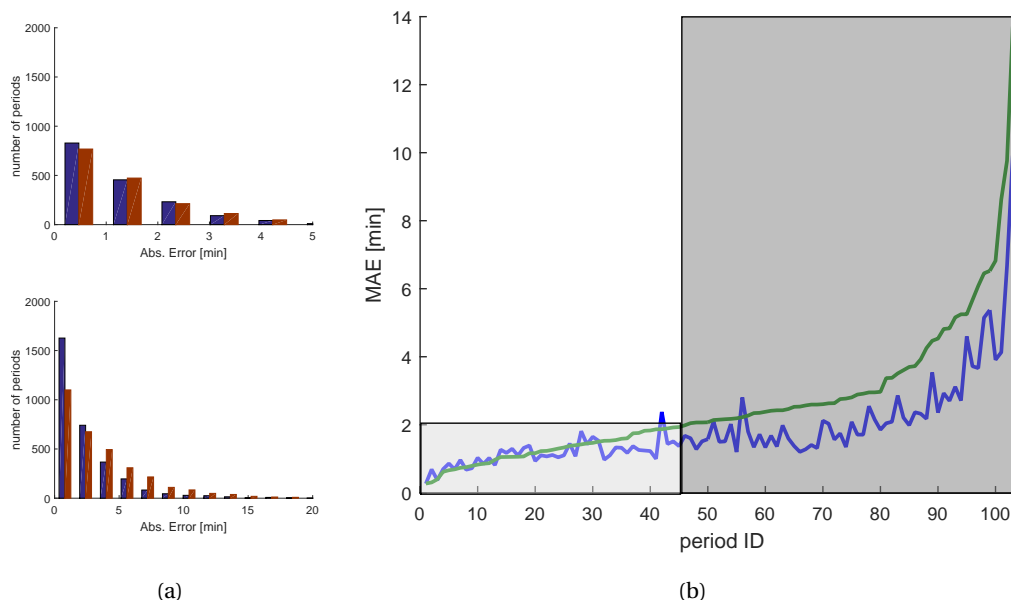
Figure 5.10: Performance of algorithms (a) Histogram of absolute errors, (b) model performance for congested periods

Figure 5.11 presents the travel times provided by the proposed methodology, the instantaneous approach (estimated by Eq. 5.1), the experienced travel time (which is considered the ground truth, estimated by Eq. 5.2) and the historical average method for 6 representative days. It clearly shows that historical average is not capable of producing accurate results under congested conditions. In overall, predictive travel time produces better results during both the onset and offset of the congestion. However, the morning peak in Figure 5.11(a) and the afternoon peaks in Figure 5.11(e) and 5.11(f) show that prediction model has a slightly better performance compared with the instantaneous approach during the congestion onset, while it has a significantly better performance during the offset. As newly available real-time information about the bottleneck becomes available, the predicted value is extremely close to the experienced during the congestion offset. Note that the day presented in Figure 5.11(d) is discussed in details in the graphical representation of the algorithm (Figure 5.9).

Table 5.4.1 provides MAE and MAPE values and the congestion duration for the days presented in Figure 5.11 along with the period ID that can be matched with Figure 5.10(b). MAE values indicate that predictive travel time outperforms the other two approaches except for two periods (30-Mar, morning peak; 31-Mar, morning peak) where all estimators (even the historical average) have small errors (around 1min). In addition, Table 5.4.1 provides the weighted average of MAE and MAPE values over all the congested periods with respect to their lengths (or durations). Results represent a clear improvement over the instantaneous travel time

Figure 5.11: Travel time predictions (a) 10-Feb-2011, (b) 30-Mar-2011, (c) 31-Mar-2011, (d) 20-Jul-2011, (e) 25-Aug-2011, (f) 22-Nov-2011

approach; delay percentage error (MAPE) decreases from 35% to 22% ($\bar{4}$0% improvement), and MAE reduces from 3.05 min to 2.10 min ($\bar{3}$0% improvement).
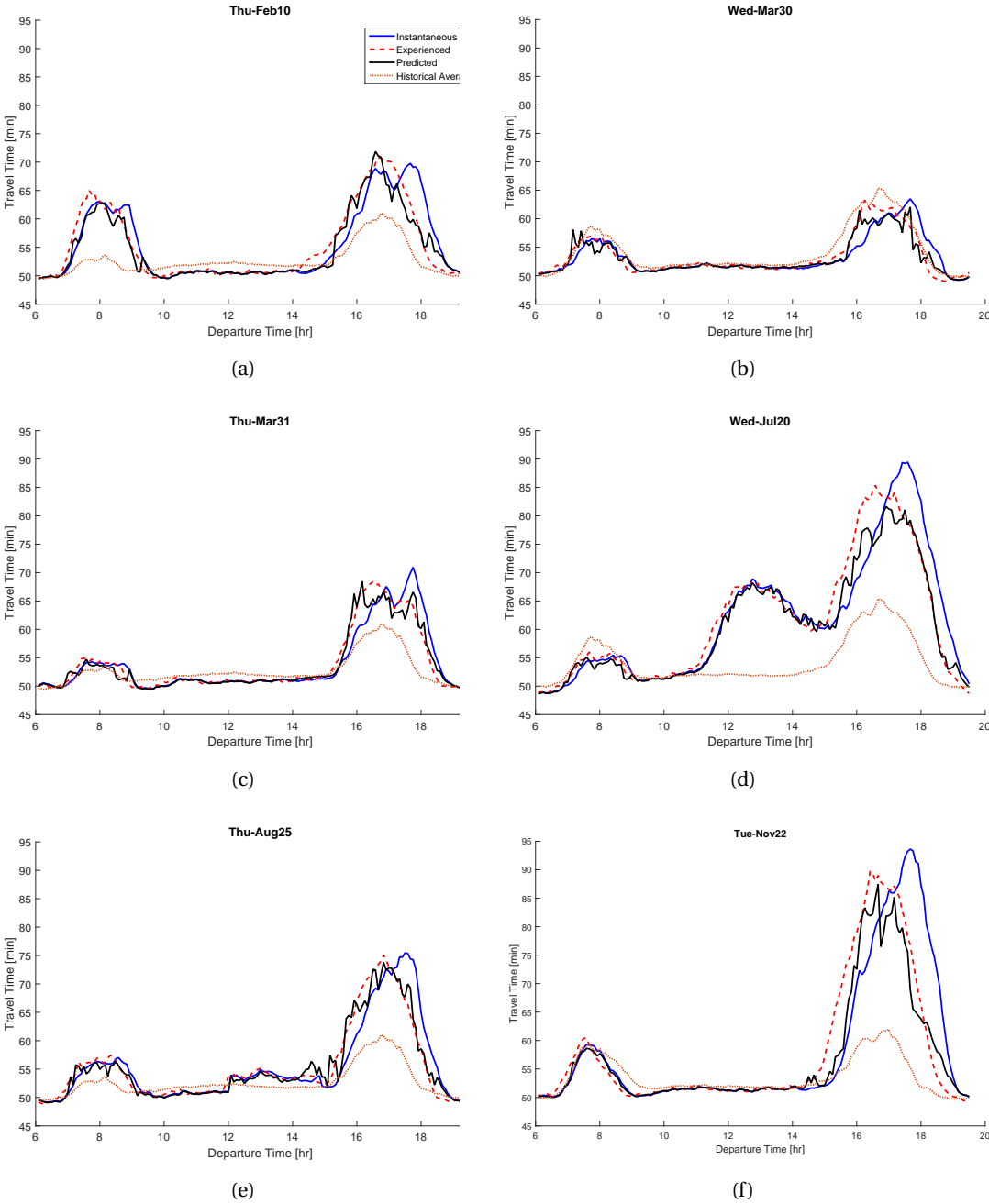
Table 5.2: MAE (min) and MAPE (%) of certain days in the testing set

| | | | *Prediction* | | *Instantaneous* | | *Hist. Avg.* | |
|---|---|---|---|---|---|---|---|---|
| *Date* | *Per. ID* | *Cong. Dur.* | *MAE* | *MAPE* | *MAE* | *MAPE* | *MAE* | *MAPE* |
| 10-Feb | 63 | 06:50-09:30 (2:40) | 1.98 | 23 | 2.43 | 31 | 6.85 | 60 |
| | 85 | 14:10-18:50 (4:40) | 1.99 | 27 | 3.70 | 48 | 5.16 | 39 |
| 30-Mar | 15 | 06:50-08:55 (1:55) | 1.17 | 18 | 1.05 | 18 | 1.30 | 22 |
| | 58 | 14:10-18:55 (4:45) | 1.34 | 29 | 2.34 | 57 | 1.68 | 35 |
| 31-Mar | 10 | 06:55-09:05 (2:10) | 1.03 | 25 | 0.83 | 25 | 1.31 | 28 |
| | 65 | 14:10-19:00 (4:50) | 1.38 | 18 | 2.53 | 30 | 4.04 | 32 |
| 20-Jul | 13 | 06:55-09:05 (2:10) | 0.81 | 15 | 0.97 | 20 | 1.60 | 29 |
| | 88 | 11:05-19:30 (8:25) | 2.18 | 17 | 4.25 | 36 | 12.19 | 65 |
| 25-Aug | 20 | 06:45-09:20 (2:35) | 0.93 | 17 | 1.16 | 24 | 2.60 | 38 |
| | 74 | 11:55-19:05 (7:10) | 1.35 | 23 | 2.77 | 37 | 4.74 | 40 |
| 22-Nov | 35 | 06:45-09:00 (2:15) | 1.32 | 22 | 1.59 | 27 | 2.28 | 53 |
| | 101 | 13:50-19:15 (5:25) | 4.13 | 34 | 8.62 | 76 | 12.34 | 51 |
| *Average of the whole testing set* | | | **2.10** | **22** | **3.05** | **35** | **5.83** | **45** |

The use of revised congestion map, presented in Figure 5.8 instead of its original has allowed us to investigate the influence of bottleneck shapes. Figure 5.12 provides evaluation results for 19 congested periods identified in 10 days (mostly 'Fridays') that exist in testing set and that belong to $3^{rd}$ cluster. Comparison of results reveals no significant improvement. Nevertheless, the motivation of this approach is to provide a methodology which is consistent with the physics of traffic and which can provide a more consistent alternative in case of missing or erroneous data. Although the revised congestion map has accurate bottleneck shapes, time spent in the queue is not significantly affected by this change. This might be the reason of the lack of improvement. Weighted average of MAE values for the periods presented in Figure 5.12 is 2.64, 2.70 and 3.90 min, respectively for predictive travel time with shockwave speed constraints, predictive travel time and instantaneous travel time.

### 5.4.2 Sensitivity analysis

Parameters used in the prediction framework are (i) the number of clusters (Figure 5.6), (ii) the number of components taken from PCA (Figure 5.13(a)) and (iii) the cluster switch parameter, $\alpha$ (Figure 5.13(b)). A sensitivity analysis is conducted to investigate the effect of these parameters on the performance of the proposed prediction model. First, to determine the optimal number of clusters in the dataset, average silhouette width value is computed for a range of possible cluster numbers (Figure 5.6). Note that, as the initialization of the algorithm is random, clustering has been implemented several times (100 in this case). As also explained in Section 5.3.2, two criteria namely stability and performance (i.e. average silhouette width)
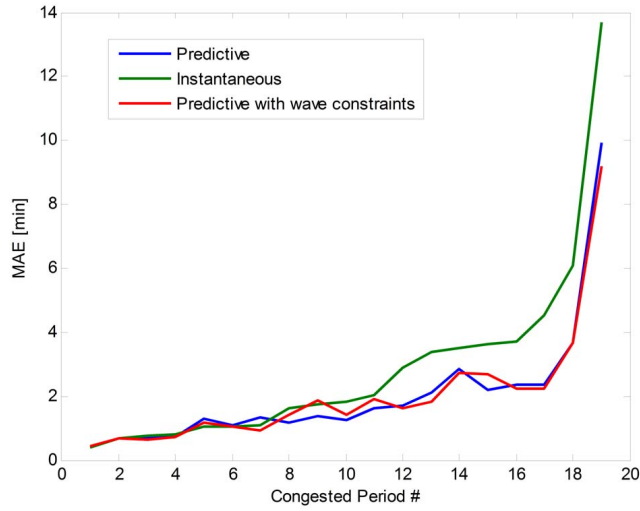
Figure 5.12: Comparison of the original and revised stochastic congestion maps

of results are considered together to determine the optimal number of clusters.



| (a) | (b) |
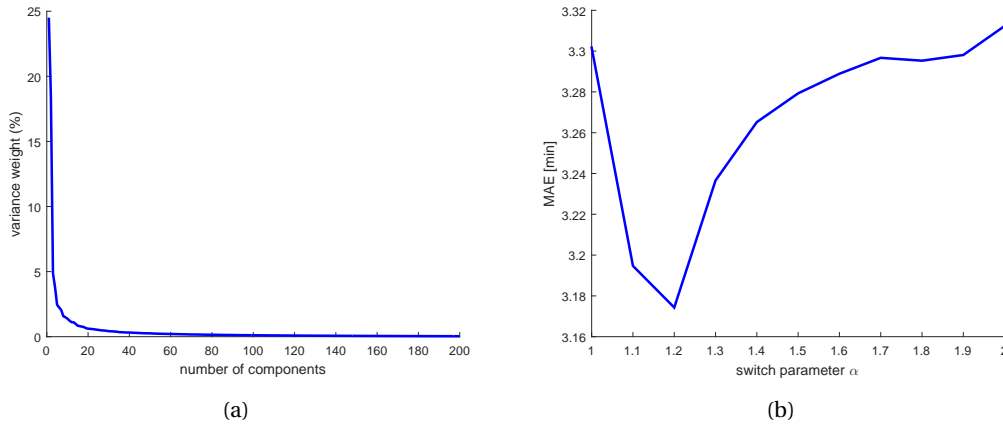
Figure 5.13: Sensitivity analysis (a) variance vs. number of components in PCA, (b) MAE vs. switch parameter $\alpha$

Second, to determine the number of principal components, variance weight along the eigenvector directions (which can be formulated as $\lambda_i / e[n]$ following the notation in Section 5.2.2) is considered. As it is seen in Figure 5.13(a), components beyond $25^{th}$ have insignificant contributions to the variance of the data set. However, cumulative variance reaches 95% at 100th component. Hence, the first 100 PCs are chosen to create matrix $S$ (see Eq. 5.3). Note that matrix $S$ is utilized only in the clustering step; the rest of the methodology is applied in the original traffic speed data (matrix $X$ in Eq. 5.3). A further analysis has shown that any number of components more than 25 results in identical clusters, which implies stability.

Third, to determine the switch parameter $\alpha$, days which belong to an alternative cluster instead of their original cluster (e.g. a 'Wednesday' which belongs to 3rd cluster) are identified, and MAE is estimated for different values of $\alpha$. Figure 5.13(a) presents the weighted average of MAE values for 11 congested periods that comply with the above description. It shows that MAE reaches its minimum at 1.2, which is the optimal value for this study. One can conduct a similar analysis to determine site-specific model parameters.

## 5.5 Summary

Dissemination of travel time information through ATIS or its use as in ATMS to deploy efficient control measures always requires the prediction of traffic conditions on the freeway. The aim of this chapter is to develop a methodology that exploits both traffic flow fundamentals and strong statistical tools.

First, an automated bottleneck identification algorithm is applied to detect the major traffic events that occur on the freeway. Then, the historical (or training) dataset is partitioned based on the clusters obtained through GMM. The results obtained from the first two parts are combined to create stochastic congestion maps for each cluster. Next, using the estimated speed profile, the congestion maps associated with threshold values and the congestion search algorithm that connects real-time and historical traffic data, this study predicts the experienced travel times. The experiment results based on the loop detector data of I-5S segment in California/San Diego indicate that the proposed method provides promising travel time predictions under varying traffic conditions.

In this study, there is no GPS data available. Experienced travel time, which is based on speed measurements at loop detectors and which is presented by Eq. 5.2, is used as ground truth travel time. Instantaneous, experienced and predicted travel times are all computed based on piecewise constant speed method. They may exhibit underestimation of travel times. However, the correction would apply to all methods compared. Therefore, it can be considered a systematic bias for all methods used and does not compromise the mutual comparison. This approach could also be compared with data driven approaches. However, this study attempts to produce experienced travel time in the following time interval, while data driven approaches aim to provide instantaneous travel time in future time periods. Hence, it would not be relevant to compare two distinct approaches.

# 6 Investigating empirical implications of hysteresis in day-to-day travel time variability

NOWADAYS, travelers not only seek to minimize their travel time on average, but also value its variation. The variation in the mean and the variance of travel time (across days, for the same departure time) has not been thoroughly investigated. A temporary decrease in capacity (e.g. congestion caused by an active bottleneck) leads to a quite significant difference in the variance of travel time for congestion onset and offset periods. This phenomenon results in hysteresis loops where the departure time periods in congestion offset exhibit a higher travel time variance than the ones in congestion onset with the same mean travel time. The aim of this chapter is to identify empirical implications that yield to the hysteresis phenomenon in day-to-day travel times. First, empirical hysteresis loop observations are provided from two different freeway sites. Second, we investigate the potential link with the hysteresis observed in traffic networks on macroscopic fundamental diagram (MFD). Third, we build a piecewise linear function that models the evolution of travel time within the day. This allows us to decompose the problem into its components, e.g. start time of congestion, peak travel time, etc. These components, along with their probability distribution functions, are employed in a Monte Carlo simulation model to investigate their partial effects on the existence of hysteresis. Correlation among critical variables is the most influential factor in this phenomenon, which should be further investigated regarding traffic flow and traffic equilibrium principles.

## 6.1 Introduction

Nowadays, there has been a growing interest in understanding travel time variability to make decisions such as route choice and departure time. Recent studies (Hollander, 2006; Asensio and Matas, 2008; Li et al., 2010b) show that people value both travel time and its predictability to choose their paths. In addition, researchers investigate the effect of travel time reliability in the scheduling of activities with respect to earliness or lateness cost (see for example Fosgerau and Karlstrom (2010), Brownstone and Small (2005)). Regarding this aspect, travel time reliability, a performance indicator of roadways, has an important effect on route and departure time choice, especially for time constrained trips (e.g. commute to work, trip to

airport).

Travel time variability, which develops the formal basis for reliability measures, can be investigated from several angles: vehicle-to-vehicle variability which corresponds to different vehicles traveling the same route at the same time, period-to-period variability corresponding to vehicles traveling the same route at different periods within a day, and day-to-day variability addressing the travel time variations of vehicles crossing the same route at the same period of time on different days (Noland and Polak, 2002). An accurate travel time reliability model might need to recognize these different types of variability at various levels. For instance, Kim and Mahmassani (2014) propose a compound distribution approach that distinguishes between day-to-day and vehicle-to-vehicle variability in modeling travel time reliability. However, they approximate the relation between mean and variance in day-to-day travel time distribution by a linear curve. In this chapter, we further investigate this relation that cannot be easily captured by a linear curve (see Figure 6.1) and identify empirical implications that lead to counter-clockwise 'hysteresis' loops. Note that details related to Figure 6.1 are later given in the text. Bates et al. (2004) is the first study to the authors' knowledge that investigates the relation between mean and variance in day-to-day travel times. They observe hysteresis loops in a single link and single bottleneck study area. This phenomenon results in counter-clockwise loops where the departure time periods in congestion offset exhibit a higher travel time variance than the ones in congestion onset with the same mean travel time.

Hysteresis in traffic domain has been first defined by Edie (1963) and Treiterer and Myers (1974) as the separation between acceleration and deceleration curves in speed-density diagrams. Asymmetric theories that provide explanations for the mechanisms of hysteresis are described in Newell (1965), Zhang (1999), Yeo and Skabardonis (2009) and others. Nevertheless, traffic hysteresis loop may disappear when lane-observed data are aggregated in 1-5 min intervals (Daganzo, 2002). The same reference attributed the hysteresis phenomenon to lane changing and the non-conservative nature of flow in a single lane. Hysteresis is also observed at the network level; Geroliminis and Sun (2011a) investigates the causes of the hysteresis in macroscopic fundamental diagram (MFD), where higher network flows are observed for the same network density in the onset and lower in the offset of the congestion. According to Geroliminis and Sun (2011a), the hysteresis phenomenon in the network level happens because of two reasons. The first reason is that distributions of individual occupancy measurements (congestion distribution) are different for the same level of network density, see also Geroliminis and Sun (2011b). The second reason is the synchronized occurrence of transient states and capacity drop at the individual detectors. Similar hysteresis loops are identified in Buisson and Ladier (2009) and Saberi and Mahmassani (2012) with empirical observations in different freeway systems. Saberi and Mahmassani (2012) explore also the relationship between size of the hysteresis loop and inhomogeneity of the congestion distribution. Gayah and Daganzo (2011) test the effects of driver adaptivity on MFD hysteresis loops in a simplified two-bin network. Daganzo (2011) provides theoretical explanation for traffic instabilities in freeway systems and proposes driver adaptation as a solution to this type of capacity loss during recovery of congestion.

In this chapter, we investigate the properties of hysteresis, and identify empirical implications that produce this phenomenon in day-to-day travel time distribution. These reasons are twofold in nature; due to network properties (i.e. traffic dynamics) and due to demand patterns. Fosgerau (2010) provides a theoretical proof for counter-clockwise loop using Vickrey bottleneck model with random service rate and assuming Nash equilibrium in arrival times. The equilibrium state in this model dictates a concave cumulative arrival function or an arrival rate which is decreasing. This implies a particular demand pattern in traffic networks and enables the theoretical proof of hysteresis loops in day-to-day travel times. However, empirical implications of hysteresis and which observed traffic variables influence the size of hysteresis loops are still not clear. On the other hand, Gayah et al. (2013b) analyzes micro-simulation data and suggests that hysteresis behavior in day-to-day travel times is linked to the hysteresis loops observed in macroscopic fundamental diagram (MFD). It is, therefore, a network property. Note that hysteresis loops in travel time differs from the one in MFD in the sense that variability is computed across time, not across space. However, analytical travel time model developed in Gayah et al. (2013b) does not explicitly account for the hysteresis in MFD, and therefore it is not possible to remove the MFD hysteresis from the model and to test its actual effect on the travel time hysteresis. Analysis based on real data will definitely shed more light towards this direction.

The chapter is structured as follows. In the next section, we analyze empirical data from loop detectors in two freeways in California and Greece. Consistent hysteresis loops are observed in both sites. In Section 6.3, we analyze the effect of network traffic properties on the hysteresis shape. We consider the hysteresis phenomenon which is very significant in freeway MFD, and investigate the effects of the existence or non-existence of MFD hysteresis on the relation between mean and variance in day-to-day travel times. In Section 6.4, we build a piecewise linear model that represents the evolution of travel time within the day. This model simply allows us to express the problem with new observable random variables, e.g. start time of congestion, peak travel time, etc. We can then employ these random variables and the distributions associated with them in a Monte Carlo simulation model in Section 6.5, and test the empirical properties which actually create the hysteresis in day-to-day travel time variability.

## 6.2 Hysteresis loop examples with real data

In this study, we have employed loop detector data from freeway performance measurement system (PeMS). PeMS collects 30-sec loop detector flow and occupancy data throughout the state of California. Then, it processes them and fills in the missing detector data to compute 5-minute flow, occupancy and speed averages (Chen et al., 2001). The first study site is a 30 mile section of I-5S in the district of San Diego/Imperial. There are 37 detectors located on the roadway, the distance between the detectors ranges from 0.05 mile to 1.47 mile. 148 days of 5-minute loop detector data is obtained through the PeMS database. Second study site is a 20-km section of Attiki Odos freeway in Athens, Greece. There are 61 loop detectors

installed on the roadway, and the distance between the detectors ranges from 0.05 to 0.7 km. In this study site, there are only 24 congested days that can be investigated in this chapter. For more detailed description of the site, the reader can refer to Antoniou et al. (2010). Note that both study sites suffer from one single active bottleneck, future research should focus on multi-bottleneck areas.

This study makes use of loop detector data to calculate instantaneous travel times for each day and each 5-minute departure time period (experienced travel time estimation as in Yildirimoglu and Geroliminis (2013) provides similar results). We employ speed measurements at the detectors in the departure time period, and apply the constant speed interpolation method (i.e. mean of the speed readings at two consecutive detectors) to calculate section travel times. The following equations indicate the mean and variance of travel time across days.

$$\mu(t) = \frac{1}{N} \sum_{i=1}^{N} \tau_i(t) \tag{6.1}$$

$$\sigma^2(t) = \frac{1}{N-1} \sum_{i=1}^{N} (\tau_i(t) - \mu(t))^2 \tag{6.2}$$

where $t$ is departure time period, $N$ is the number of days and $\tau_i(t)$ is the travel time along the freeway stretch on day $i$ and departure time $t$.

Figure 6.1 presents resulting mean-variance curve for the morning peak period in I-5S and Attiki Odos along with the departure time periods given next to the corresponding points. Note that the loop is counter-clockwise; the periods in congestion offset exhibit a higher travel time variance than the ones in congestion onset with the same mean travel time. In the remainder of the chapter, we employ Californian freeway site that has larger samples of data and that allows us to estimate distributions with statistical significance.

## 6.3  Investigating the link between hysteresis in MFD and day-to-day travel time distribution

This section investigates the potential link between the hysteresis observed on MFD (across space) and on the mean-variance curve of travel times (across time). We first build the aggregate (average) flow-density relationship using the loop detector data. Let us define the following weighted averages: $q_i(t) = \sum_j (q_{ij}(t) \cdot l_j) / \sum_j l_j$ for flow and $k_i(t) = \sum_j (k_{ij}(t) \cdot l_j) / \sum_j l_j$ for density, where $q_{ij}(t)$ and $k_{ij}(t)$ denote the flow and density measurements respectively for day $i$, detector $j$ and departure time $t$. In addition, $l_j$ is the distance associated with detector $j$ (i.e. spacing between successive detectors). Figure 6.2(a) presents the empirical relationship between space-mean flow and space-mean density from 37 detectors in I-5S for the period between 6am and 11am throughout 148 days (each point represents a 5 min period).

Figure 6.1: Mean-Variance Curve in the Morning Peak Period (a) in I-5S, (b) in Attiki Odos.

Figure 6.2(b) highlights hysteresis loops in the network level for 4 representative days from the same sample. As MFDs for individual days follow a clockwise hysteresis loop beyond the critical average density value (see Figure 6.2(b)), not all the points fall on the same curve; MFD exhibits a scattered behavior (see Figure 6.2(a)). In addition, we fit a 3rd degree polynomial function on the empirical data (see Figure 6.2(a)). Two different types of MFD defined here allow us to calculate travel times with and without the effect of hysteresis in the spatial aspect.
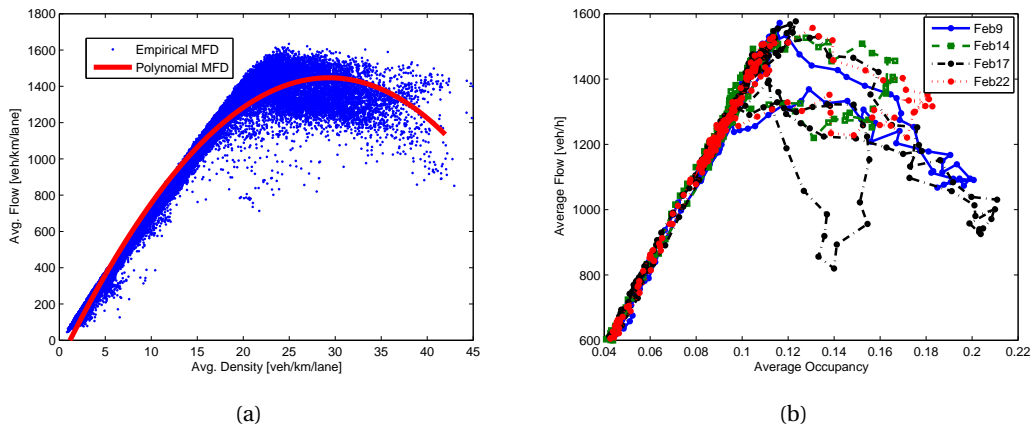


Figure 6.2: (a) Empirical and polynomial MFD, (b) empirical MFD for several days.

Average speed in the network on day $i$ and departure time $t$ can be calculated using the relation between average flow and average density observed through MFD; $v_i(t) = q_i(t)/k_i(t)$. Of more interest here, average unit travel time in the network can be calculated by simply taking the reverse of the average speed ($\varrho_i(t) = 1/v_i(t)$), and the unit travel time can be converted to the section travel time by multiplying it with the section length $l$ ($\bar{\tau}_i(t) = \varrho_i(t) \cdot l$). MFD relationship between average flow and average density is defined in two different ways; with

the empirical data collected from the detectors and with the polynomial function fitted on the empirical data. The former approach does not assign a functional relation between flow and density, and it includes the hysteresis effect described above, see Figure 6.2(b). The latter fits a polynomial function on the empirical data, and avoids the clockwise loop observed in MFD, see polynomial MFD in Figure 6.2(a). In the first approach, travel time is calculated using the actual average flow and density values in the freeway, while the second approach employs the actual density measurements and corresponding average flow values on the polynomial function. Of course, the second approach is not exactly consistent with traffic flow dynamics, but the purpose of this analysis is only to investigate the hysteresis phenomena observed in MFD and travel time data.  Figure 6.3(a) presents the curves that result from the actual travel time calculation and the two MFD based travel time definitions.  Although they are not identical, all three travel time considerations exhibit a counter-clockwise hysteresis loop, which implies that hysteresis on MFD is not the only reason of hysteresis in travel times.
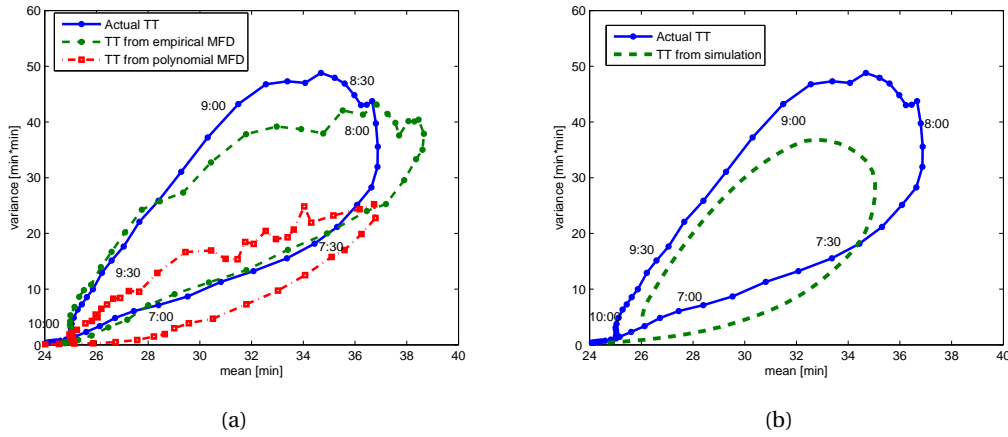


Figure 6.3: Mean - variance curves (a) from different MFD considerations, (b) from single-reservoir simulation.

As $q_i(t)$ and $k_i(t)$ values employed in polynomial MFD approach are not consistent regarding traffic flow dynamics (average flow on the polynomial curve would create different average density values in the next time periods), we further investigate the relationship between the hysteresis in MFD and in day-to-day travel times by considering a dynamic queueing system. Dynamic equation of the state variable or the accumulation, $n$, for a single reservoir system (Daganzo, 2007):

$$\frac{\mathrm{d}n}{\mathrm{d}t} = I'(t) - O(n(t)) \tag{6.3}$$

where $I'(t)$ is the exogenous demand or arrival flow, and $O(n(t))$ is the outflow that corresponds to the accumulation $n(t)$ on the MFD function. We incorporate arrival flow values, $I'(t)$, taken from the loop detectors located on the on-ramps and process aggregated traffic flow dynamics through Eq. 6.3. Note that the arrival flow, in this case, is the sum of entering flow from all

the on-ramps and the upstream point of the freeway section in hand. Similarly, the outflow is the sum of exiting flow from all the off-ramps and the downstream point of the freeway section. This analysis allows us to produce average flow and density values consistent with the traffic dynamics defined by MFD (without hysteresis), while the results from the polynomial MFD presented in Figure 6.3(a) are based on the actual density measurements and fictitious flows that correspond to them on the MFD function. This analysis is repeated for different days with arrival flow values that correspond to them. Resulting day-to-day time dependent accumulation values are converted to the average unit travel time, and mean-variance values of the section travel times are calculated. Figure 6.3(b) presents the mean-variance curve and the counter-clockwise loop that result from this single-reservoir simulation model. The existence of hysteresis simply proves that clockwise loop on MFD is not the only reason of the counter-clockwise loop in mean-variance relation on travel times. However, its existence might still affect the size and the shape of the hysteresis on travel times.

## 6.4 A piecewise linear model

This section builds a piecewise linear model that represents the evolution of travel time in a peak period. The parameters of the model constitute random variables (e.g. start time of congestion, peak travel time, etc.) that can be later investigated through a Monte Carlo simulation method if they have consistent statistical properties.

Figure 6.4(a) depicts the evolution of travel times across 148 days on I-5S for the morning peak. While the onset and offset of congestion follow a unimodal curve with a well defined maximum, there is strong stochasticity on the physical characteristics of this curve (start time of congestion, slope of onset/offset, etc.). To study the effect of uncertainty, we approximate the evolution of travel times on a given day by a 4-segment piecewise linear function (see Figure 6.4(b)). For each day $i$ in the data set, we apply Eq. 6.4, a curve fitting nonlinear optimization problem, to minimize the residual sum of squares (Eq. 6.4a) and to identify the attributes of the piecewise linear model (Eq. 6.4b). Note that day $i$ is omitted in Eq. 6.4 for notational simplicity. Eq. 6.4b indicates the mathematical formulation of travel time defined by the piecewise linear model, which is also the constraint of the optimization problem. The piecewise linear model consists of four sections; two edges outside the peak period ($p_0$), congestion onset ($p_{on}$) and congestion offset ($p_{off}$). Note the piecewise linear function presented in Eq. 6.4b has constant values at both edges ($p_0$), which represent steady state traffic conditions outside the peak period. Nonlinear programming produces six parameters (or six random variables for each day); start time of congestion $t_s$, starting travel time $\tau_s$, peak time $t_p$, peak travel time $\tau_p$, end time of congestion $t_e$ and ending travel time $\tau_e$. The initial conjecture is that stochastic nature of these variables influences the hysteresis of travel time curves. Note that travel time data set of each day consists of the same time periods between 6:00 and 11:00.

(a)                                                        (b)
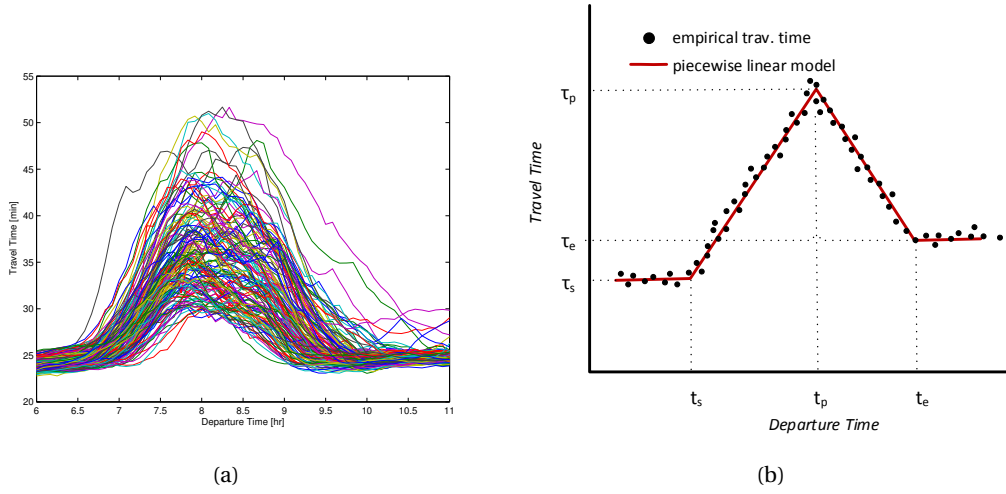
Figure 6.4: (a) travel time series of all days, (b) Piecewise linear model.

$$\underset{t_s,t_p,t_e,\tau_s,\tau_p,\tau_e}{\text{minimize}} \qquad \sum_t \big(f(t) - \tau(t)\big)^2 \tag{6.4a}$$

$$\text{subject to} \qquad f(t) = \begin{cases} \tau_s & \text{if } t \le t_s \ (p_0), \\[2mm] \tau_s + \dfrac{(t - t_s)}{(t_p - t_s)} \cdot (\tau_p - \tau_s) & \text{if } t_s < t \le t_p \ (p_{\text{on}}), \\[2mm] \tau_p - \dfrac{(t - t_p)}{(t_e - t_p)} \cdot (\tau_p - \tau_e) & \text{if } t_p < t \le t_e \ (p_{\text{off}}), \\[2mm] \tau_e & \text{if } t > t_e \ (p_0). \end{cases} \tag{6.4b}$$

Figure 6.5 presents the histograms of the six variables and the parametric marginal distributions fitted with maximum likelihood estimation. Note that parametric distributions presented in Figure 6.5 are the ones that produce best fitting (in terms of maximum likelihood values) among an extensive set of alternatives. The variables that indicate a point in time or that separate steady state, congestion onset and offset periods (i.e. $t_s, t_p, t_e$) are normally distributed, while the variables that represent travel time values at these points (i.e. $\tau_s, \tau_p, \tau_e$) are log-normally distributed. We have also used Kolmogorov-Smirnov test to assess whether the chosen distributions are suited to the data set. The null hypothesis that variables come from the fitted distributions are accepted in each case with 5% significance level. $p$-values are 0.24, 0.21, 0.37, 0.10, 0.19, 0.05 for $t_s, t_p, t_e, \tau_s, \tau_p, \tau_e$, respectively. Note that the range of variables $\tau_s$ and $\tau_e$ is actually quite short.

The parametric distributions presented in Figure 6.5 give probabilities of various values of variables without conditioning on other variables. In other words, these are marginal distributions. In order to establish the conditional relations between the random variables, dependence or correlation between them must be exploited. In fact, there are two distinct
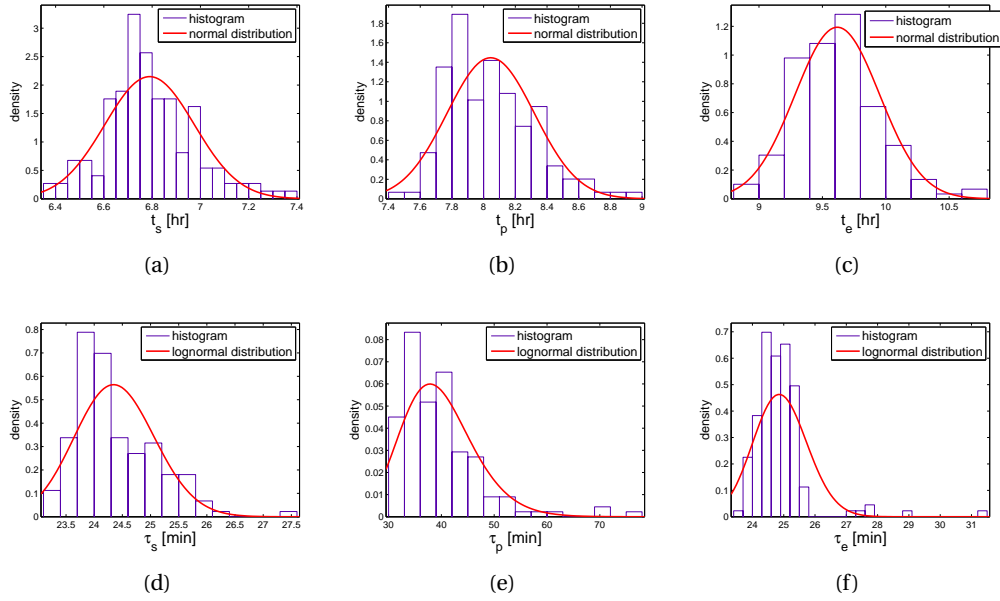
Figure 6.5: Histograms and distributions of random variables (a) starting time $t_s$, (b) peak time $t_p$, (c) ending time $t_e$, (d) starting travel time $\tau_s$, (e) peak travel time $\tau_p$, (f) ending travel time $\tau_e$.

approaches to model the dependence between the variables. The first is to determine the joint distribution of random variables, which is often limited due to the difficulty of computing the distributions. The second is to specify marginal distributions and correlations between the variables. Such a problem is actually not fully specified, because there may be many joint probability distributions satisfying the same marginal distribution and the same correlation values. However, often there is only enough data to estimate marginal distributions and correlations between variables. Table 6.1 presents Pearson correlation coefficient between any combination of six variables. Strong correlations exist for some pairs of variables that deserve further attention; the values in bold will be explained later in the chapter.

Table 6.1: Correlation coefficients

|            | $t_s$ | $t_p$ | $t_e$    | $\tau_s$ | $\tau_p$  | $\tau_e$ |
|------------|-------|-------|----------|----------|-----------|----------|
| $t_s$      | 1.00  | 0.08  | 0.05     | -0.03    | 0.20      | 0.21     |
| $t_p$      | -     | 1.00  | **0.48** | 0.03     | **0.33**  | 0.31     |
| $t_e$      | -     | -     | 1.00     | -0.07    | **0.45**  | 0.35     |
| $\tau_s$   | -     | -     | -        | 1.00     | 0.39      | 0.45     |
| $\tau_p$   | -     | -     | -        | -        | 1.00      | 0.48     |
| $\tau_e$   | -     | -     | -        | -        | -         | 1.00     |

Figure 6.6 presents, based on the analysis of the real data, the probability that a given time of departure $t$ belongs in any of the 3 time intervals of Eq. 6.4b, i.e. in the onset of congestion ($p_{\text{on}}$), the offset of congestion ($p_{\text{off}}$) or outside the congestion range ($p_0$). Note that there are

specific times of a day that belong to onset for some days and offset for others (e.g. 7:45-8:15) and other times that might experience no delay or a very high delay with significant probability. These observations highlight that existing theoretical analysis of morning commute fall short to describe stochastic behavior. This should be a future research priority.
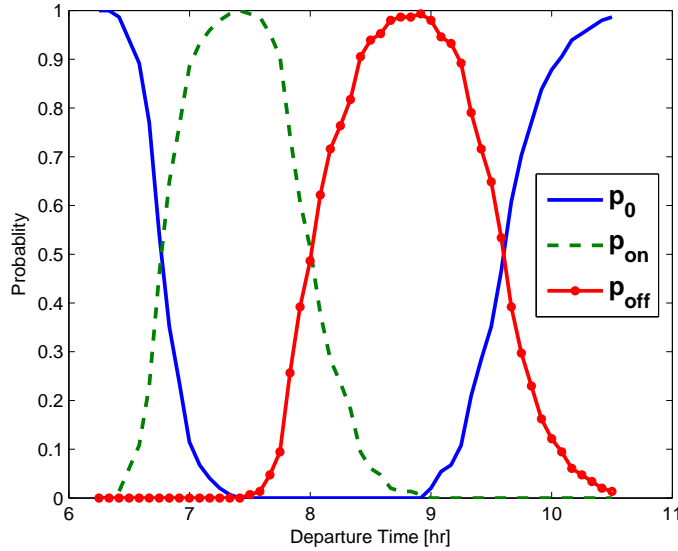


Figure 6.6: Probability of congestion intervals ($p_0$: outside congestion range, $p_{on}$: congestion onset, $p_{off}$:congestion offset)

## 6.5 Monte Carlo method

An analytical approximation of the mean and variance of travel time based on Eq. 6.4b would be very tedious due to the correlations among variables that does not allow for an elegant formulation. For example, Geroliminis and Sun (2011a) have estimated in a closed form solution the size of hysteresis loops for freeway MFDs, as the involved variables have small level of correlation. As we show later correlation is among the main reasons for the hysteresis loops and cannot be ignored. As an alternative way to investigate the empirical properties that create hysteresis in day-to-day travel time mean and variance, we employ the results presented in Figure 6.5 and Table 6.1, and build a Monte Carlo simulation method. This method runs simulation many times (each representing a different day in this case) by drawing a random variable from each one of the distributions defined above. In fact, the generation of independent random variables from a parametric distribution is quite straightforward. However, generation of correlated variables is not possible in all parametric distribution types. In case of correlated normal variables, Cholesky factorization of covariance matrix is quite useful. If $r_1, ..., r_n$ are independent draws from standard normal distribution $N(0,1)$, then $\mathbf{s} = \mathbf{a} + \mathbf{L} * \mathbf{r}$ is a vector of draws from $n$-variate normal $N(\mathbf{a}, \mathbf{LL}^T)$ where $\mathbf{L}$ is lower

triangular, and $\mathbf{LL}^T$ is the Cholesky factorization of the covariance matrix. Note that, in our case, correlation does not only exist between normal variables; there is significant correlation between lognormal-lognormal and normal-lognormal pairs as well (see Table 6.1). In fact, there is no special treatment to generate correlated lognormal variables. However, we can exploit the properties of a lognormal distribution.

If $\mathbf{x}$ be a $n$-variate lognormal random vector with mean $\boldsymbol{\mu_x}$ and covariance matrix $\Sigma_{\mathbf{x}}$, then we can write $\mathbf{y} = (ln(x_1), ..., ln(x_n)) = N(\boldsymbol{\mu_y}, \Sigma_{\mathbf{y}})$. Similarly, let $\mathbf{z} = N(\boldsymbol{\mu_z}, \Sigma_{\mathbf{z}})$ be a $m$-variate normal random vector with mean $\boldsymbol{\mu_z}$ and covariance matrix $\Sigma_{\mathbf{z}}$. To generate the random vectors $\mathbf{s_x}$ and $\mathbf{s_z}$ with the observed variables $\mathbf{x}$ and $\mathbf{z}$, we apply the following steps:

1. Create $\mathbf{V} = \begin{bmatrix} (\mathbf{y_1}...\mathbf{y_N}) \\ (\mathbf{z_1}...\mathbf{z_N}) \end{bmatrix} = \begin{bmatrix} \begin{pmatrix} y_{1,1}...y_{1,N} \\ ... \\ y_{n,1}...y_{n,N} \end{pmatrix} \\ \begin{pmatrix} z_{1,1}...z_{1,N} \\ ... \\ z_{m,1}...z_{m,N} \end{pmatrix} \end{bmatrix} = \begin{bmatrix} \begin{pmatrix} ln(x_{1,1})...ln(x_{1,N}) \\ ... \\ ln(x_{n,1})...ln(x_{n,N}) \end{pmatrix} \\ \begin{pmatrix} z_{1,1}...z_{1,N} \\ ... \\ z_{m,1}...z_{m,N} \end{pmatrix} \end{bmatrix}$

   $N$ is the number of days in the data set.

2. Calculate mean $\mathbf{m} = \begin{bmatrix} \boldsymbol{\mu_y} \\ \boldsymbol{\mu_z} \end{bmatrix}$ and covariance matrix $\Sigma = \begin{bmatrix} \Sigma_{\mathbf{y}} & \Sigma_{\mathbf{y,z}} \\ \Sigma_{\mathbf{y,z}} & \Sigma_{\mathbf{z}} \end{bmatrix}$. Note that $\Sigma_{\mathbf{y,z}}$ represents the covariance values between $\mathbf{y}$ and $\mathbf{z}$.

3. Apply Cholesky factorization $\Sigma = \mathbf{LL}^T$.

4. Generate $r_1, ..., r_{m+n}$ independent draws from $N(0,1)$, and calculate $\begin{bmatrix} \mathbf{s_Y} \\ \mathbf{s_Z} \end{bmatrix} = \mathbf{m} + \mathbf{L} \cdot \mathbf{r}$.

5. Calculate $\mathbf{s_x} = e^{\mathbf{s_y}}$.

In this study, $\mathbf{s_X}$ represents the lognormal variables (i.e. $\tau_s, \tau_p, \tau_e$), while $\mathbf{s_z}$ represents the normal variables (i.e. $t_s, t_p, t_e$). In the Monte Carlo method, we repeat the steps 4 and 5 defined above as many times as needed to reach stable results. In addition, we simply discard Monte Carlo realizations where physical requirements $t_s < t_p < t_e$ and $\tau_p > \tau_s, \tau_e$ are not satisfied. Note that once the random variables are computed, it is quite straightforward to calculate travel time values that correspond to different departure time periods. Figure 6.7 presents the variables to be generated in the Monte Carlo method and the linear lines connecting them.

We apply Monte Carlo simulation here with several correlation configurations to investigate the potential link between the hysteresis and correlation types between the random variables. First, we assume no dependence between variables, that is we skip the steps defined above, and draw independent random variables from the marginal distributions presented in Figure 6.5. This step can also be solved analytically with a closed form solution. Figure 6.8(a) presents the resulting day-to-day (or simulation-to-simulation) mean-variance curve from 5000 simulation
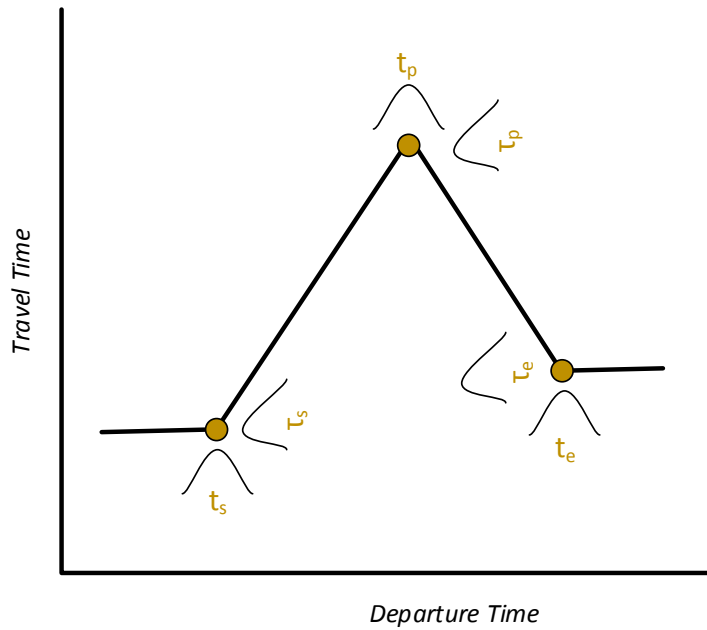
Figure 6.7: Random variables in the Monte Carlo method

runs with independent variables. Simulation results, in this case, do not exhibit an hysteresis behavior; mean-variance points fall almost exactly on the same curve in both congestion onset and offset periods. This is an important result, as it implies that ignoring the correlation yields to hysteresis-free curves. Note that mean and variance values presented in Figure 6.8 are not the sample mean and variance defined in Eq. 6.1-6.2. As Monte Carlo method is designed to produce the distribution of an unknown probabilistic entity, and not some sample properties of it, we fit a lognormal distribution on travel time values of each departure time across days (i.e. $\{\tau_i(t)|\forall i\}$), and present the mean-variance values of lognormal distributions in Figure 6.8. This simply allows us to distinguish between sample mean-variance and population mean-variance, which may be very different especially in limited data sets.

Second, we incorporate the correlation structure between the random variables. Figure 6.8(b) presents mean-variance curves that result from Monte Carlo simulation (each 5000 runs) with different correlation structures. Accounting for dependence between all six variables, we are able to approximate counter-clockwise pattern observed with actual travel time data. The difference between the observed and simulated data may be due to the limited size of the available data set. Simulation results indicate mean-variance values from 5000 runs, while the available data set has only 148 days. To understand the relative importance of the variables regarding the hysteresis existence, we have tried an extensive set of possible subset combinations for which the correlation structure is needed. Although we assume correlation only between three variables (i.e. $t_p, t_e, \tau_p$) and consider the other three (i.e. $t_s, \tau_s, \tau_e$) independent, we obtain a mean-variance curve that is very similar to the one that

results from six dependent variables (see Figure 6.8(b)). In fact, two of the three independent variables (i.e. $\tau_s, \tau_e$) are distributed in a very short range (see Figure 6.5(e)-(f)). In other words, travel time values outside the peak periods have approximately constant values over the days. Therefore, as a next step, we remove the randomness of the three independent variables, and replace them with their constant mean values. As represented by the black curve in Figure 6.8(b), this leads to the small underestimation of variance in the beginning of congestion onset due to the neglected distribution of $t_s, \tau_s$ and $\tau_e$. However, as the other variables come into play, the mean-variance curve becomes similar to the ones from previous scenarios. This analysis concludes that the variables $t_p, t_e, \tau_p$ and the correlation structure associated with them are the most influential factors of hysteresis existence in day-to-day travel time variability.
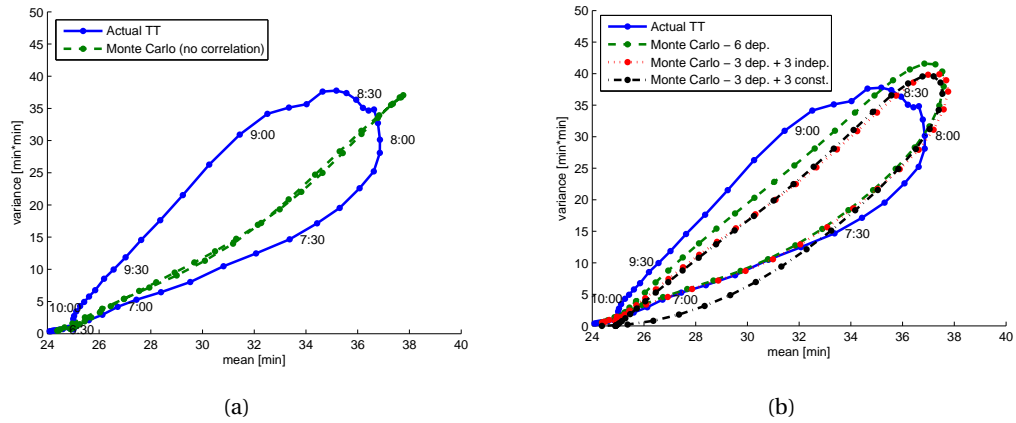


Figure 6.8: Monte Carlo results (a) with independent variables, (b) with different degrees of dependence.

Table 6.1 presents in bold the correlation coefficients between the three significant variables defined above (i.e. $t_p, t_e, \tau_p$). In fact, they are all positively correlated, which is quite reasonable from a traffic point of view. Based on the results presented in the chapter, we can claim that the following statements are the traffic flow properties that are behind hysteresis phenomenon.

- A late occurrence in peak time $t_p$ is accompanied by a late congestion ending time $t_e$.

- A late occurrence in peak time $t_p$ is accompanied by a high peak travel time $\tau_p$.

- A high peak travel time $\tau_p$ is accompanied by a late congestion ending time $t_e$.

In other words, the uncertainty in the length of congestion onset is strongly associated with the hysteresis loop. This can also be seen in Figure 6.6, where the beginning of the onset of congestion has small uncertainty across the days (see the sharp increase in $p_{on}$), while the uncertainty in the length of the onset (range where $p_{on} > 0$) and the peak time $t_p$ (where $p_{on}$ and $p_{off}$ have nonzero values) significantly influences the large range for the offset of

congestion (range where $p_{\text{off}} > 0$). Note that the probability of being in the offset of congestion for different times ($p_{\text{off}}$) has longer tails than being in the onset ($p_{\text{on}}$). A probabilistic model for the duration of congestion has also been studied by Stathopoulos and Karlaftis (2002), i.e. given the onset of congestion, what is the probability that it will end during the following time periods.

Results from the Monte Carlo method show that the dependence between variables plays a crucial role in the existence and size of hysteresis. However, the empirical correlations (or degree of dependence) between the random variables can vary for different roadway systems. Therefore, we run a sensitivity analysis to determine the changes in hysteresis shape in response to the changes in the variance and the correlation values. First of all, to identify the *magnitude of hysteresis*, we define a metric which represents the average variance difference between congestion onset and offset periods. Magnitude of hysteresis is simply the result of dividing the area of the polygon specified by the vertices of mean-variance values (or the area inside the mean-variance curve) and the difference between maximum and minimum mean travel time value. In other words, we approximate the area inside the curve with a rectangle of the same area, and we calculate the rectangle height that represents the average difference in variance between onset and offset periods. Second, we modify the variance and the correlation values for the generated random variables, and run Monte Carlo simulation. Note that in the sensitivity analysis we only use three variables (i.e. $t_p, t_e, \tau_p$) that are proven to be significant. Eq. 6.5 presents the modified covariance matrix.

$$
\Sigma = \begin{bmatrix} \sigma^2_{t_p}\alpha_\sigma & \alpha_\rho\,\alpha_\sigma\,\sigma_{t_p}\,\sigma_{t_e}\,\rho_{t_p,t_e} & \alpha_\rho\,\alpha_\sigma\,\sigma_{t_p}\,\sigma_{ln(\tau_p)}\,\rho_{t_p,ln(\tau_p)} \\ - & \sigma^2_{t_e}\alpha_\sigma & \alpha_\rho\,\alpha_\sigma\,\sigma_{t_e}\,\sigma_{ln(\tau_p)}\,\rho_{t_e,ln(\tau_p)} \\ - & - & \sigma^2_{ln(\tau_p)}\alpha_\sigma \end{bmatrix}
\tag{6.5}
$$

where $\sigma_i$ is the standard deviation of the variable $i$, $\rho_{i,j}$ is the correlation between variables $i$ and $j$, $\alpha_\sigma$ and $\alpha_\rho$ are the multiplication factors for the variance and correlation values, respectively. Eq. 6.5 simply creates a new covariance matrix based on the defined multiplication factors, which allows us to test the effect of higher and lower dependence between variables. Figure 6.9(a) depicts the magnitude of hysteresis with several combinations of variance and correlation factors. Note that if $\alpha_\sigma$ and $\alpha_\rho$ are equal to 1, we employ the original covariance matrix. The low values presented in lower left triangle in Figure 6.9(a) represent an insignificant hysteresis shape where the the difference in variance between onset and offset periods is very small. An increase in $\alpha_\sigma$, or $\alpha_\rho$ or both results in an increase in magnitude values. Figure 6.9(b) presents mean-variance curves that result from several combinations of variance and correlation factors. Note that magnitude of hysteresis calculated for these combinations are circled in red in Figure 6.9(a). Although the combinations $\alpha_\sigma = 1, \alpha_\rho = 1.6$ and $\alpha_\sigma = 1.6, \alpha_\rho = 1$ have very similar magnitude values due to the symmetry in Figure 6.9(a), resulting curves are very different from each other.
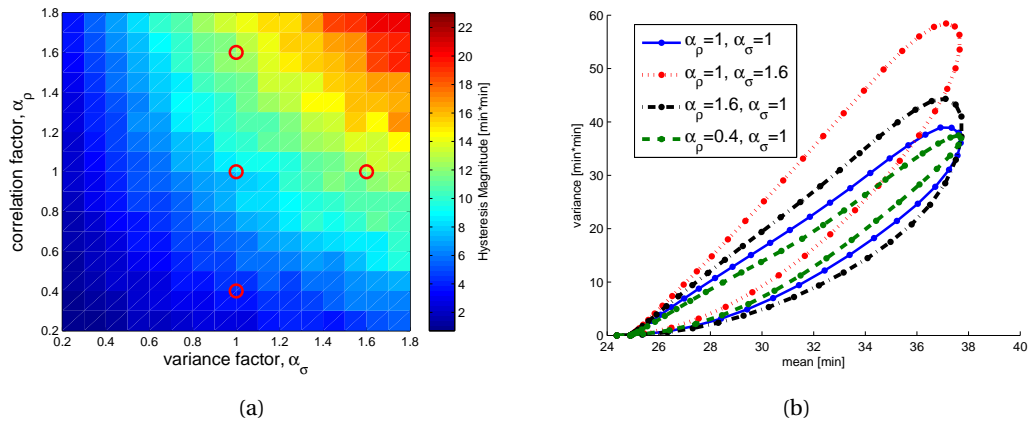
Figure 6.9: Sensitivity of results to (a) magnitude of hysteresis, (b) resulting hysteresis curves.

## 6.6 Summary

In this chapter, we investigate empirical implications of hysteresis loops in day-to-day travel times. An interesting observation is that there is no strong link between the hysteresis in the freeway MFD and the hysteresis in day-to-day travel time distribution. A simple dynamic queueing system based on polynomial MFD without hysteresis still produces significant hysteresis loops in the mean-variance relation. The two phenomena are not strongly related as the estimation of variables is based on averages across space for the MFD and across time for the travel time graph. Our analysis shows that the hysteresis in the day-to-day travel time distribution is mainly due to correlations and stochasticity among variables related to the duration and the occurrence of congestion across days (start and ending time, duration of offset and onset, maximum value of travel time). This explanation is based on the development of a piecewise linear model that represents the evolution of travel times and allows us to decompose the problem into its components, e.g. start time of congestion, peak travel time, etc. A Monte Carlo simulation with these variables shows that $t_p, t_e, \tau_p$ and the correlation structure associated with them are the critical traffic properties that lead to hysteresis in day-to-day travel time variability.

# 7 Conclusion and future research

THIS thesis thoroughly studies the incorporation of route choice effect into large-scale urban network modeling (Part I) and travel time prediction / variability models (Part II). The final chapter briefly summarizes the findings and the main contributions of the thesis and proposes new directions for future research. Detailed conclusions are also provided at the end of each chapter.

## 7.1 Approximating dynamic equilibrium conditions with macroscopic fundamental diagrams

The main contributions can be listed as follows:

- Incorporation of regional route choice into multi-region MFD modeling.

- Approximating dynamic stochastic user equilibrium conditions in large scale networks.

- Developing a loading procedure that can handle the stochastic components in MFD modeling.

- Integrating time-dependent average trip length with production-outflow derivations.

This study develops a DTA model that can be incorporated within MFD modeling in order to establish equilibrium conditions. As it considers the response of drivers to dynamic traffic conditions, the proposed model could lead to a more accurate network traffic modeling and better traffic predictions.

A future research direction is to design a large-scale traffic control strategy, e.g. perimeter control, that addresses the reaction of drivers to changing traffic conditions. The challenge is to incorporate the proposed assignment model into the design of control strategy, which might impose additional computational difficulties. Such a goal could also be achieved through a leader-follower approach, i.e. an iterative implementation of control design and traffic

assignment. A procedure that inherently addresses route choice behavior can maximize the performance of the urban network, which is the ultimate control goal. In addition, the model does not consider local congestion pockets within the region. Therefore, traffic management based on this model could negatively affect the homogeneity of traffic conditions inside the region. A hierarchical control approach could be a possible solution to this problem. Control or route guidance on the lower level could use the full instantaneous information from the links, while control on the upper level takes regional path decision deploying MFD dynamics and considering the evolution of traffic conditions in the network in the future time periods.

Another direction is to investigate how aggregated modeling describes network propagation in case of heavily directional flows and time- and space-dependent pockets of congestion. A combination of dynamic partitioning algorithms (extending the work of Ji and Geroliminis (2012) and Ji et al. (2014)) with real time perimeter control and route guidance strategies with multiple regions can provide efficient solutions.

## 7.2 Equilibrium analysis and route guidance in large-Scale networks with MFD dynamics

The main contributions are summarized as:

- Developing two MFD-based traffic models with different aggregation levels.

- Integrating user equilibrium and system optimum flows in the network MFD analysis.

- Developing a route guidance control system based on parsimonious MFD model.

This study extends the DTA model developed in Chapter 2 to a route guidance scheme, where travelers are advised to follow a certain series of subregions. The proposed model could be further exploited in a hierarchical traffic management scheme and alleviate the computational difficulties in detailed route guidance methods.

As a future research direction, the route guidance system can be integrated with the perimeter control strategy, which is expected to further improve homogeneity and network performance. This is challenging because the perimeter control decisions are optimized based on the region-based model, while the control decisions should be applied in the subregion-based model, and designing a realistic and accurate information feedback from the plant to the optimization or operation model needs further investigations. The interaction between the adaptive control and equilibrium state is an important issue to be addressed in future works. This problem could be overcome in a day-to-day assignment framework, where people adapt to unexpected traffic conditions (created by the new control strategy) by taking different route decisions over days in addition to en-route decision mechanisms.

The results of this study cannot be directly employed in a real-world route guidance scheme,

but it is possible with some extra effort to incorporate it into a hierarchical traffic guidance scheme. While the proposed model provides sequence of subregions, a link-level DTA model could provide the drivers with sequence of links inside the subregions. This hierarchical approach could remove the computational burden from the link-level DTA model, as the application area would be limited by the size of subregions. A field test would provide more insights about the applicability and implications of the proposed control strategy.

## 7.3 Aggregated dynamic route choice patterns for large-scale mixed urban/freeway networks

The list below summarizes the main contributions:

- Utilization of probe vehicles in order to extract human mobility patterns in a large metropolitan area.

- Observing and reconstructing aggregated route patterns.

- Testing equilibrium assumptions at the network scale.

This study focuses on observation and reconstruction of aggregated route choice patterns in large-scale mixed urban/freeway networks. The proposed method makes use of probe data to exploit the aggregated patterns in roadway networks. If integrated with MFD modeling, aggregated route patterns that result from this analysis may lead to better traffic predictions in large-scale networks.

Future work should focus on discrete choice models instead of shortest path algorithms. A better route choice estimation at individual trip level will definitely lead to better estimation of aggregated route choice patterns. Discrete choice models may also allow us to distinguish between toll freeways and others in an explicit way. This might overcome the issue of overestimation on freeway load. Another future research direction is to test the accuracy of aggregated route choice variables regarding MFD dynamics. Previous work in literature estimates average trip lengths through loop detector data. In other words, outflow value identified through loop detectors has been employed to estimate average trip lengths. However, in this chapter, we propose trajectory-based calculations. The integration between these two should be analyzed for a real network where both probe vehicle and loop detector data are available (this is not the case in Shenzhen).

## 7.4 Experienced travel time prediction for congested freeways

The main contributions can be listed as follows:

- Combining correct physics and advanced data mining tools to predict experienced

travel times.

- Reacting in real-time to abrupt changes in traffic pattern for freeway travel time prediction.

- Development of space–time stochastic congestion maps for freeway routes.

- Experimental validation based on loop detector data on Californian freeways.

This chapter develops a method to predict travel times in congested freeway routes. Predicted travel times can be integrated in monitoring tools to help practitioners evaluate the roadway network and to enable travelers to make informed decisions.

This study utilizes only loop detector data to estimate the speed and the travel time on the roadway. However, the proposed model could have a great potential with other sources of sensor technologies as well. Real-time traffic data can be collected from various sources including fixed loop detectors, license plate recognition systems, toll infrastructure, cell phones and probe vehicles. Especially the increased use of GPS-equipped cell phones or vehicles provide transportation researchers with a great potential of big data. However, the use of probe data for traffic estimation raises challenges because of errors in raw measurements, post-processing algorithms (e.g. map matching) and penetration rate. Further research is needed on fusion of probe data with existing sources of traffic data in order to complete the big picture of traffic state.

Estimation of travel time distributions for different departure times should also be a research priority, as reliability measures can improve the planning of travel trips for various users and provide tools to traffic management for more efficient control. Another research direction is to test the developed method on a signalized arterial where bottlenecks (or queue profiles) are affected by both demand - capacity interaction and traffic signals.

## 7.5 Investigating empirical implications of hysteresis in day-to-day travel time variability

The main contributions are as following:

- Observing hysteresis loops in mean–variance travel time plots for freeway routes.

- Investigating the link between MFD hysteresis and travel time variability hysteresis.

- Employing a Monte Carlo simulation to explain the magnitude the reasons for hysteresis.

- Developing a piecewise linear function to model the evolution of travel time within a day.

## 7.5. Investigating empirical implications of hysteresis in day-to-day travel time variability

This study investigates empirical causes of hysteresis loops in day-to-day travel time variability. The proposed method could lead to a better understanding of travel time reliability measures, where onset and offset of congestion are treated separately.

A future work should investigate the potential link between the empirical findings presented in this chapter and traffic equilibrium models for the morning and evening commute. The fact that onset and offset of congestion approximately follow a triangular form (as of Figure 6.4) possibly implies a strong connection between travel time reliability and scheduling of trip. While some recent work at the network level equilibrium (see for example Geroliminis and Levinson (2009), Arnott (2013), Gonzales and Daganzo (2012)) provides interesting findings towards this direction, a connection with reliability needs further investigation. Another interesting future direction is to investigate different traffic management schemes (e.g. perimeter control, ramp metering or congestion pricing) for arterial or freeway networks that treat travel time reliability as part of objective function to be improved.

# Bibliography

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

Adeli, H. 2001. Neural networks in civil engineering: 1989–2000. *Computer-Aided Civil and Infrastructure Engineering*, 16(2):126–142.

Antoniou, C., Kopelias, P., and Papadimitriou, F. 2010. Active traffic management applications in greece. the case of attica tollway. In *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*, pages 237–242. IEEE.

Arnott, R. 2013. A bathtub model of downtown traffic congestion. *Journal of Urban Economics*, 76(1):110–121.

Asensio, J. and Matas, A. 2008. Commuters' valuation of travel time variability. *Transportation Research Part E: Logistics and Transportation Review*, 44(6):1074–1085.

Barceló, J. and Casas, J. 2005. Dynamic network simulation with aimsun. In *Simulation Approaches in Transportation Analysis*, pages 57–98. Springer.

Bates, J., Fearon, J., and Black, I. 2004. Frameworks for modelling the variability of journey times on the highway network. Technical report, ARUP.

Bekhor, S., Ben-Akiva, M., and Scott Ramming, M. 2002. Adaptation of logit kernel to route choice situation. *Transportation Research Record: Journal of the Transportation Research Board*, (1805):78–85.

Ben-Akiva, M., Bierlaire, M., Bottom, J., Koutsopoulos, H., and Mishalani, R. 1997. Development of a route guidance generation system for real-time application. In *Proceedings of the IFAC Transportation Systems Conference, Chania*.

Ben-Akiva, M., De Palma, A., and Kanaroglou, P. 1986. Dynamic model of peak period traffic congestion with elastic arrival rates. *Transportation Science*, 20(3):164–181.

Ben-Akiva, M. E. and Bierlaire, M. 1999. Discrete choice methods and their applications to short-term travel decisions. In Hall, R., editor, *Handbook of Transportation Science*, chapter 2, pages 5–34. Kluwer, Dordrecht.

# Bibliography

Ben-Elia, E., Di Pace, R., Bifulco, G. N., and Shiftan, Y. 2013. The impact of travel information's accuracy on route-choice. *Transportation Research Part C: Emerging Technologies*, 26:146–159.

Bierlaire, M. and Crittin, F. 2006. Solving noisy, large-scale fixed-point problems and systems of nonlinear equations. *Transportation Science*, 40(1):44–63.

Bottom, J. A. 2000. *Consistent Anticipatory Route Guidance*. PhD thesis, Massachusetts Institute of Technology, Cambridge.

Brownstone, D. and Small, K. 2005. Valuing time and reliability: assessing the evidence from road pricing demonstrations. *Transportation Research Part A: Policy and Practice*, 39(4):279–293.

Buisson, C. and Ladier, C. 2009. Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams. *Transportation Research Record: Journal of the Transportation Research Board*, 2124(-1):127–136.

Cantarella, G. E., Improta, G., and Sforza, A. 1991. Iterative procedure for equilibrium network traffic signal setting. *Transportation Research Part A: General*, 25(5):241 – 249.

Cantillo, V., Heydecker, B., and de Dios Ortúzar, J. 2006. A discrete choice model incorporating thresholds for perception in attribute values. *Transportation Research Part B: Methodological*, 40(9):807–825.

Cascetta, E., Nuzzola, A., Russo, F., and Vitetta, A. 1996. A modified logit route choice model overcoming path overlapping problems: Specification and some calibration results for interurban networks. In Lesort, J. B., editor, *Proceedings of the 13th International Symposium on Transportation and Traffic Theory*, pages 697–711. Pergamon, Oxford.

Chabini, I. 1998. Discrete dynamic shortest path problems in transportation applications: Complexity and algorithms with optimal run time. *Transportation Research Record: Journal of the Transportation Research Board*, 1645(1):170–175.

Chen, C., Petty, K., Skabardonis, A., Varaiya, P., and Jia, Z. 2001. Freeway performance measurement system: mining loop detector data. *Transportation Research Record: Journal of the Transportation Research Board*, 1748(1):96–102.

Chen, C., Skabardonis, A., and Varaiya, P. 2003. Travel-time reliability as a measure of service. *Transportation Research Record: Journal of the Transportation Research Board*, (1855):74–79.

Chen, C., Skabardonis, A., and Varaiya, P. 2004. Systematic identification of freeway bottlenecks. *Transportation Research Record: Journal of the Transportation Research Board*, (1867):46–52.

Chiu, Y.-C., Bottom, J., Mahut, M., Paz, A., Balakrishna, R., Waller, T., and Hicks, J. 2011. Dynamic traffic assignment: A primer. *Transportation Research E-Circular*, (E-C153).

Coifman, B. 2002. Estimating travel times and vehicle trajectories on freeways using dual loop detectors. *Transportation Research Part A: Policy and Practice*, 36(4):351–364.

Coifman, B. and Krishnamurthy, S. 2007. Vehicle reidentification and travel time measurement across freeway junctions using the existing detector infrastructure. *Transportation Research Part C: Emerging Technologies*, 15(3):135–153.

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

Daganzo, C. F. 2002. A behavioral theory of multi-lane traffic flow. part i: Long homogeneous freeway sections. *Transportation Research Part B: Methodological*, 36(2):131–158.

Daganzo, C. F. 2007. Urban gridlock: macroscopic modeling and mitigation approaches. *Transportation Research Part B: Methodological*, 41(1):49–62.

Daganzo, C. F. 2011. On the macroscopic stability of freeway traffic. *Transportation Research Part B: Methodological*, 45(5):782–788.

Daganzo, C. F., Bouthelier, F., and Sheffi, Y. 1977. Multinomial probit and qualitative choice: A computationally efficient algorithm. *Transportation Science*, 11(4):338–358.

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

Daganzo, C. F. and Sheffi, Y. 1977. On stochastic models of traffic assignment. *Transportation Science*, 11(3):253–274.

De Palma, A., Ben-Akiva, M., Lefevre, C., and Litinas, N. 1983. Stochastic equilibrium model of peak period traffic congestion. *Transportation Science*, 17(4):430–453.

Dharia, A. and Adeli, H. 2003. Neural network model for rapid forecasting of freeway link travel time. *Engineering Applications of Artificial Intelligence*, 16(7):607–613.

Diakaki, C., Papageorgiou, M., and McLean, T. 1997. Simulation studies of integrated corridor control in glasgow. *Transportation Research Part C: Emerging Technologies*, 5(3):211–224.

Dial, R. B. 1971. A probabilistic multipath traffic assignment model which obviates path enumeration. *Transportation Research*, 5(2):83–111.

Dion, F. and Rakha, H. 2006. Estimating dynamic roadway travel times using automatic vehicle identification data for low sampling rates. *Transportation Research Part B: Methodological*, 40(9):745–766.

Doig, J. C., Gayah, V. V., and Cassidy, M. J. 2013. Inhomogeneous flow patterns in undersaturated road networks. *Transportation Research Record: Journal of the Transportation Research Board*, 2390(1):68–75.

## Bibliography

Du, J., Wong, S., Shu, C.-W., Xiong, T., Zhang, M., and Choi, K. 2013. Revisiting jiang's dynamic continuum model for urban cities. *Transportation Research Part B: Methodological*, 56:96–119.

Du, L., Peeta, S., and Kim, Y. H. 2012. An adaptive information fusion model to predict the short-term link travel time distribution in dynamic traffic networks. *Transportation Research Part B: Methodological*, 46(1):235–252.

Duranton, G. and Turner, M. A. 2011. The fundamental law of road congestion: Evidence from us cities. *The American Economic Review*, pages 2616–2652.

Edie, L. C. 1963. Discussion of traffic stream measurements and definitions. pages 139–154, OECD, Paris, France. Proceedings of the 2nd International Symposium on the Theory of Traffic Flow.

European Commission, E. 2011. Roadmap to a single european transport area towards a competitive and resource efficient transport system. *White Paper*.

Fei, X., Lu, C.-C., and Liu, K. 2011. A bayesian dynamic linear model approach for real-time short-term freeway travel time prediction. *Transportation Research Part C: Emerging Technologies*, 19(6):1306–1318.

Fosgerau, M. 2010. On the relation between the mean and variance of delay in dynamic queues with random capacity and demand. *Journal of Economic Dynamics and Control*, 34(4):598–603.

Fosgerau, M. and Karlstrom, A. 2010. The value of reliability. *Transportation Research Part B: Methodological*, 44(1):38–49.

Frejinger, E. and Bierlaire, M. 2007. Capturing correlation with subnetworks in route choice models. *Transportation Research Part B: Methodological*, 41(3):363–378.

Gayah, V., Dixit, V., and Guler, S. 2013a. Relationship between mean and day-to-day variation in travel time in urban networks. *EURO Journal on Transportation and Logistics*, pages 1–17.

Gayah, V., Dixit, V., and Guler, S. 2013b. Relationship between mean and day-to-day variation in travel time in urban networks. *EURO Journal on Transportation and Logistics*, pages 1–17.

Gayah, V. V. and Daganzo, C. F. 2011. Clockwise hysteresis loops in the macroscopic fundamental diagram: An effect of network instability. *Transportation Research Part B: Methodological*, 45(4):643–655.

Gayah, V. V. and Dixit, V. V. 2013a. Using mobile probe data and the macroscopic fundamental diagram to estimate network densities. *Transportation Research Record: Journal of the Transportation Research Board*, 2390(1):76–86.

136

Gayah, V. V. and Dixit, V. V. 2013b. Using mobile probe data and the macroscopic fundamental diagram to estimate network densities. *Transportation Research Record: Journal of the Transportation Research Board*, 2390(1):76–86.

Gayah, V. V., Gao, X. S., and Nagle, A. S. 2014. On the impacts of locally adaptive signal control on urban network stability and the macroscopic fundamental diagram. *Transportation Research Part B: Methodological*, 70(0):255 – 268.

Geroliminis, N. and Daganzo, C. F. 2008. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological*, 42(9):759–770.

Geroliminis, N., Haddad, J., and Ramezani, M. 2013. Optimal perimeter control for two urban regions with macroscopic fundamental diagrams: A model predictive approach. *IEEE Transactions on Intelligent Transportation Systems*, 14(1):348–359.

Geroliminis, N. and Levinson, D. 2009. Cordon pricing consistent with the physics of overcrowding. *Transportation and Traffic Theory*, 1(1):219–240.

Geroliminis, N. and Sun, J. 2011a. Hysteresis phenomena of a macroscopic fundamental diagram in freeway networks. *Transportation Research Part A: Policy and Practice*, 45(9):966–979.

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

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

Gonzales, E. and Daganzo, C. 2012. Morning commute with competing modes and distributed demand: user equilibrium, system optimum, and pricing. *Journal of Urban Economics*, 46(10):1519–1534.

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

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

Haddad, J. and Shraiber, A. 2014. Robust perimeter control design for an urban region. *Transportation Research Part B: Methodological*, 68:315–332.

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

Herrera, J. C. and Bayen, A. M. 2010. Incorporation of lagrangian measurements in freeway traffic state estimation. *Transportation Research Part B: Methodological*, 44(4):460–481.

## Bibliography

Hollander, Y. 2006. Direct versus indirect models for the effects of unreliability. *Transportation Research Part A: Policy and Practice*, 40(9):699–711.

Hoogendoorn, S. P. and Bovy, P. H. 2004. Dynamic user-optimal assignment in continuous time and space. *Transportation Research Part B: Methodological*, 38(7):571–592.

Jahn, O., Möhring, R. H., Schulz, A. S., and Stier-Moses, N. E. 2005. System-optimal routing of traffic flows with user constraints in networks with congestion. *Operations research*, 53(4):600–616.

Jayakrishnan, R., Tsai, W. K., and Chen, A. 1995. A dynamic traffic assignment model with traffic-flow relationships. *Transportation Research Part C: Emerging Technologies*, 3(1):51–72.

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

Ji, Y., Luo, J., and Geroliminis, N. 2014. Empirical observations of congestion propagation and dynamic partitioning with probe data for large-scale systems. *Transportation Research Record: Journal of the Transportation Research Board*, (2422):1–11.

Jiang, X. and Adeli, H. 2004. Wavelet packet-autocorrelation function method for traffic flow pattern analysis. *Computer-Aided Civil and Infrastructure Engineering*, 19(5):324–337.

Jiang, Y., Wong, S., Ho, H., Zhang, P., Liu, R., and Sumalee, A. 2011. A dynamic traffic assignment model for a continuum transportation system. *Transportation Research Part B: Methodological*, 45(2):343–363.

Kaufman, L. and Rousseeuw, P. J. 2009. *Finding groups in data: an introduction to cluster analysis*, volume 344. John Wiley & Sons.

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

Keyvan-Ekbatani, M., Yildirimoglu, M., Geroliminis, N., and Papageorgiou, M. 2013. Traffic signal perimeter control with multiple boundaries for large urban networks. In *16th International IEEE Conference on Intelligent Transportation Systems (ITSC)*.

Keyvan-Ekbatani, M., Yildirimoglu, M., Geroliminis, N., and Papageorgiou, M. 2015. Multiple concentric gating traffic control in large-scale urban networks. *Intelligent Transportation Systems, IEEE Transactions on*, 16(4):2141–2154.

Kim, J. and Mahmassani, H. S. 2014. A finite mixture model of vehicle-to-vehicle and day-to-day variability of traffic network travel times. *Transportation Research Part C: Emerging Technologies*, 46:83 – 97.

Knoop, V., Hoogendoorn, S., and Van Lint, J. W. 2012. Routing strategies based on macroscopic fundamental diagram. *Transportation Research Record: Journal of the Transportation Research Board*, 2315(1):1–10.

Knoop, V. L., Hoogendoorn, S., and Van Lint, J. 2013. Impact of traffic dynamics on macroscopic fundamental diagram. In *Transportation Research Board 92nd Annual Meeting*.

Kothuri, S., Tufte, K., Fayed, E., and Bertini, R. 2008. Toward understanding and reducing errors in real-time estimation of travel times. *Transportation Research Record: Journal of the Transportation Research Board*, (2049):21–28.

Kuwahara, M. and Akamatsu, T. 1997. Decomposition of the reactive dynamic assignments with queues for a many-to-many origin-destination pattern. *Transportation Research Part B: Methodological*, 31(1):1 – 10.

Leclercq, L., Chiabaut, N., and Trinquier, B. 2014. Macroscopic fundamental diagrams: A cross-comparison of estimation methods. *Transportation Research Part B: Methodological*, 62(0):1 – 12.

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

Leclercq, L., Laval, J. A., and Chiabaut, N. 2011. Capacity drops at merges: an endogenous model. *Transportation research part B: methodological*, 45(9):1302–1313.

Ledoux, C. 1997. An urban traffic flow model integrating neural networks. *Transportation Research Part C: Emerging Technologies*, 5(5):287–300.

Li, X., Peng, F., and Ouyang, Y. 2010a. Measurement and estimation of traffic oscillation properties. *Transportation Research Part B: Methodological*, 44(1):1–14.

Li, Z., Hensher, D. A., and Rose, J. M. 2010b. Willingness to pay for travel time reliability in passenger transport: a review and some new empirical evidence. *Transportation research part E: logistics and transportation review*, 46(3):384–403.

Lighthill, M. and Whitham, G. 1955a. On kinematic waves. i. flood movement in long rivers. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 229(1178):281–316.

Lighthill, M. and Whitham, G. 1955b. On kinematic waves. ii. a theory of traffic flow on long crowded roads. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 229(1178):317–345.

Little, J. D. 1961. A proof for the queuing formula: L= $\lambda$w. *Operations research*, 9(3):383–387.

Liu, H. X., He, X., and He, B. 2009. Method of successive weighted averages (mswa) and self-regulated averaging schemes for solving stochastic user equilibrium problem. *Networks and Spatial Economics*, 9(4):485–503.

## Bibliography

Liu, Y., Lin, P.-W., Lai, X., Chang, G.-L., and Marquess, A. 2006. Developments and applications of simulation-based online travel time prediction system: traveling to ocean city, maryland. *Transportation Research Record: Journal of the Transportation Research Board*, (1959):92–104.

Lo, H. K. and Szeto, W. 2002. A cell-based variational inequality formulation of the dynamic user optimal assignment problem. *Transportation Research Part B: Methodological*, 36(5):421–443.

Magnanti, T. L. and Perakis, G. 1997. Averaging schemes for variational inequalities and systems of equations. *Mathematics of Operations Research*, 22(3):568–587.

Mahmassani, H., Williams, J., and Herman, R. 1984. Investigation of network-level traffic flow relationships: some simulation results. *Transportation Research Record*, 971:121–130.

Mahmassani, H. S., Hou, T., and Saberi, M. 2013a. Connecting networkwide travel time reliability and the network fundamental diagram of traffic flow. *Transportation Research Record: Journal of the Transportation Research Board*, 2391(1):80–91.

Mahmassani, H. S. and Peeta, S. 1993. Network performance under system optimal and user equilibrium dynamic assignments: Implications for advanced traveler information systems. *Transportation Research Record: Journal of the Transportation Research Board*, 1408:83–93.

Mahmassani, H. S., Saberi, M., and Zockaie, A. 2013b. Urban network gridlock: Theory, characteristics, and dynamics. *Transportation Research Part C: Emerging Technologies*, page (In Press).

Mahmassani, H. S., Saberi, M., and Zockaie, A. 2013c. Urban network gridlock: Theory, characteristics, and dynamics. *Transportation Research Part C: Emerging Technologies*, 36(0):480 – 497.

Mazaré, P.-E., Tossavainen, O.-P, Bayen, A., and Work, D. 2012. Trade-offs between inductive loops and gps probe vehicles for travel time estimation: A mobile century case study. In *Transportation Research Board 91st Annual Meeting*, volume 349.

Mazloumian, A., Geroliminis, N., Helbing, D., Mazloumian, A., Geroliminis, N., and Helbing, D. 2010. The spatial variability of vehicle densities as determinant of urban network capacity. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1928):4627–4647.

Merchant, D. K. and Nemhauser, G. L. 1978. A model and an algorithm for the dynamic traffic assignment problems. *Transportation Science*, pages 183–199.

Min, W. and Wynter, L. 2011. Real-time road traffic prediction with spatio-temporal correlations. *Transportation Research Part C: Emerging Technologies*, 19(4):606–616.

Nagendra, S. S. and Khare, M. 2003. Principal component analysis of urban traffic characteristics and meteorological data. *Transportation Research Part D: Transport and Environment*, 8(4):285–297.

Nagle, A. S. and Gayah, V. V. 2013. The accuracy of network-wide traffic state estimations using mobile probe data. *Transportation Research Record, in press.*

Nanthawichit, C., Nakatsuji, T., and Suzuki, H. 2003. Application of probe-vehicle data for real-time traffic-state estimation and short-term travel-time prediction on a freeway. *Transportation Research Record: Journal of the Transportation Research Board*, (1855):49–59.

Newell, G. F. 1965. Instability in dense highway traffic: A review. pages 73–83, London, UK. Proceedings of the 2nd International Symposium on the Theory of Road Traffic Flow,.

Newell, G. F. 1993. A simplified theory of kinematic waves in highway traffic, part i: General theory. *Transportation Research Part B: Methodological*, 27(4):281–287.

Noland, R. B. and Polak, J. W. 2002. Travel time variability: a review of theoretical and empirical issues. *Transport Reviews*, 22(1):39–54.

Okutani, I. and Stephanedes, Y. J. 1984. Dynamic prediction of traffic volume through kalman filtering theory. *Transportation Research Part B: Methodological*, 18(1):1–11.

Ortigosa, J., Menendez, M., and Tapia, H. 2013a. 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*, pages 1–22.

Ortigosa, J., Menendez, M., and Tapia, H. 2013b. 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*, pages 1–22.

Papageorgiou, M. 1990. Dynamic modeling, assignment, and route guidance in traffic networks. *Transportation Research Part B: Methodological*, 24(6):471 – 495.

Papageorgiou, M. 1998. Some remarks on macroscopic traffic flow modelling. *Transportation Research Part A: Policy and Practice*, 32(5):323–329.

Papageorgiou, M., Diakaki, C., Dinopoulou, V., Kotsialos, A., and Wang, Y. 2003. Review of road traffic control strategies. *Proceedings of the IEEE*, 91(12):2043–2067.

Peeta, S. and Mahmassani, H. S. 1995. System optimal and user equilibrium time-dependent traffic assignment in congested networks. *Annals of Operations Research*, 60(1):81–113.

Peeta, S. and Ziliaskopoulos, A. K. 2001. Foundations of dynamic traffic assignment: The past, the present and the future. *Networks and Spatial Economics*, 1(3):233–265.

Ramezani, M., Haddad, J., and Geroliminis, N. 2015. Dynamics of heterogeneity in urban networks: aggregated traffic modeling and hierarchical control. *Transportation Research Part B: Methodological*, 74:1–19.

# Bibliography

Ramming, M. S. 2001. *Network knowledge and route choice*. PhD thesis, Massachusetts Institute of Technology.

Ran, B. and Boyce, D. 1996. *Modeling dynamic transportation networks: an intelligent transportation system oriented approach*. Springer.

Ran, B., Hall, R. W., and Boyce, D. E. 1996. A link-based variational inequality model for dynamic departure time/route choice. *Transportation Research Part B: Methodological*, 30(1):31 – 46.

Rousseeuw, P. J. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65.

Saberi, M. and Mahmassani, H. S. 2012. Exploring properties of networkwide flow-density relations in a freeway network. *Transportation Research Record: Journal of the Transportation Research Board*, 2315(1):153–163.

Saeedmanesh, M. and Geroliminis, N. 2015. Clustering of heterogeneous networks with directional flows based on "snake" similarities. In *Transportation Research Board 94th Annual Meeting*, number 15-1354.

Sheffi, Y. 1985. *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*. Prentice Hall, Englewood Cliffs.

Stathopoulos, A. and Karlaftis, M. 2002. Modeling duration of urban traffic congestion. *Journal of Transportation Engineering*, 128(6):587–590.

Treiber, M., Kesting, A., and Helbing, D. 2010. Three-phase traffic theory and two-phase models with a fundamental diagram in the light of empirical stylized facts. *Transportation Research Part B: Methodological*, 44(8):983–1000.

Treiterer, J. and Myers, J. 1974. The hysteresis phenomenon in traffic flow. *Proceedings of the 6th International Symposium on Transportation and Traffic Theory*, 6:13–38.

Van Lint, J. 2006. Reliable real-time framework for short-term freeway travel time prediction. *Journal of transportation engineering*, 132(12):921–932.

Van Lint, J. 2008. Online learning solutions for freeway travel time prediction. *Intelligent Transportation Systems, IEEE Transactions on*, 9(1):38–47.

Vanajakshi, L. and Rilett, L. 2007. Support vector machine technique for the short term prediction of travel time. In *Intelligent Vehicles Symposium, 2007 IEEE*, pages 600–605. IEEE.

Vlahogianni, E. I., Geroliminis, N., and Skabardonis, A. 2008. Empirical and analytical investigation of traffic flow regimes and transitions in signalized arterials. *Journal of Transportation Engineering*, 134(12):512–522.

Vlahogianni, E. I., Karlaftis, M. G., and Golias, J. C. 2005. Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach. *Transportation Research Part C: Emerging Technologies*, 13(3):211–234.

Vovsha, P. and Bekhor, S. 1998. Link-nested logit model of route choice: overcoming route overlapping problem. *Transportation Research Record: Journal of the Transportation Research Board*, (1645):133–142.

Wang, Y., Papageorgiou, M., and Messmer, A. 2008. Real-time freeway traffic state estimation based on extended kalman filter: Adaptive capabilities and real data testing. *Transportation Research Part A: Policy and Practice*, 42(10):1340–1358.

Weijermars, W. and Van Berkum, E. 2005. Analyzing highway flow patterns using cluster analysis. In *Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE*, pages 308–313. IEEE.

Wieczorek, J., Fernández-Moctezuma, R., and Bertini, R. 2010. Techniques for validating an automatic bottleneck detection tool using archived freeway sensor data. *Transportation Research Record: Journal of the Transportation Research Board*, (2160):87–95.

Yang, H. and Yagar, S. 1995. Traffic assignment and signal control in saturated road networks. *Transportation Research Part A: Policy and Practice*, 29(2):125 – 139.

Yang, J.-S. 2005. A study of travel time modeling via time series analysis. In *Control Applications, 2005. CCA 2005. Proceedings of 2005 IEEE Conference on*, pages 855–860. IEEE.

Yeo, H. and Skabardonis, A. 2009. Understanding stop-and-go traffic in view of asymmetric traffic theory. In *Transportation and Traffic Theory 2009: Golden Jubilee*, pages 99–115. Springer.

Yeon, J., Elefteriadou, L., and Lawphongpanich, S. 2008. Travel time estimation on a freeway using discrete time markov chains. *Transportation Research Part B: Methodological*, 42(4):325–338.

Yildirimoglu, M. and Geroliminis, N. 2013. Experienced travel time prediction for congested freeways. *Transportation Research Part B: Methodological*, 53:45–63.

Yildirimoglu, M. and Geroliminis, N. 2014. Approximating dynamic equilibrium conditions with macroscopic fundamental diagrams. *Transportation Research Part B: Methodological*, 70:186–200.

Yildirimoglu, M., Limniati, Y., and Geroliminis, N. 2015a. Investigating empirical implications of hysteresis in day-to-day travel time variability. *Transportation Research Part C: Emerging Technologies*, 55:340–350.

Yildirimoglu, M., Ramezani, M., and Geroliminis, N. 2015b. Equilibrium analysis and route guidance in large-scale networks with mfd dynamics. *Transportation Research Part C: Emerging Technologies (in press)*.

## Bibliography

Zhang, H. M. 1999. A mathematical theory of traffic hysteresis. *Transportation research part B: Methodological*, 33(1):1–23.

Zhang, X. and Rice, J. A. 2003. Short-term travel time prediction. *Transportation Research Part C: Emerging Technologies*, 11(3):187–210.

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

Zhu, F. and Ukkusuri, S. V. 2013. A cell based dynamic system optimum model with non-holding back flows. *Transportation Research Part C: Emerging Technologies*, 36:367–380.

Ziliaskopoulos, A. K. 2000. A linear programming model for the single destination system optimum dynamic traffic assignment problem. *Transportation science*, 34(1):37–49.

144

# Mehmet Yildirimoglu

## Personal Data

|  |  |
|---|---|
| Place and Date of Birth: | Turkey \| 6 October 1985 |
| Address: | Ave. de Montchoisi 25, 1006, Lausanne, Switzerland |
| Phone: | +41 78 734 3288 |
| email: | MEHMET.YILDIRIMOGLU@EPFL.CH |

## Research Interests

Intelligent transportation systems, Traffic operations and control, Traffic flow theory, Complex systems, Big data, Sustainable mobility, Transit and logistics

## Education

**2011-2015**     Ph.D. in Civil Engineering,
**Ecole Polytechnique Federale de Lausanne (EPFL)**, Switzerland
Thesis: "Traffic Modeling, Forecasting and Assignment in Large-Scale Networks"
Advisor: Prof. Nikolas Geroliminis

**2009-2011**     M.S. Degree in Civil Engineering
**Rutgers University**, New Jersey
Thesis: "Travel Time Estimation and Prediction in Closed Toll Highway"
Advisor: Prof. Kaan Ozbay

**2004-2009**     B.S. Degree in Civil Engineering
**Middle East Technical University (METU)**, Turkey

## Research Experience

**2011-*Current*** | **Research Assistant at EPFL, Switzerland**

**2012** | **Master Thesis Co-supervisor**
Anne Koymans, "Travel time distributions for congested transportation networks"

**2012-2014** | **Semester project Supervisor**
Paul Anderson, "Ramp metering and variable speed limit for coordinated control"
Benjamin Zimeray; Joao P. Medeiros, "Travel time reliability for congested freeways"
Cyprien Say; Mathieu Plourde, "Coordinated ramp metering for Swiss freeways"

**2009-2011** | **Research Assistant at Rutgers University, New Jersey**

## Teaching Experience

| | |
|---|---|
| 2011-2014 | **Teaching Assistant at** EPFL, **Switzerland**<br>*Transportation Systems Engineering*<br>Gave lecture sessions, Taught classes consisting of 110+ students with diverse backgrounds, helped the design of curriculum, led a team of 3 teaching assistants in managing the class materials, supervised exercise sessions. |
| 2014 | **Teaching Assistant at** EPFL, **Switzerland**<br>*Fundamentals of Traffic Operation and Control*<br>Supervised exercise sessions for classes of 70+ students, prepared exercise materials. |

## Honors and Ranks

| | |
|---|---|
| 2009 | Graduated 2nd in class (among 240 people – METU Civil Engineering) |
| 2005-2009 | Received 8 High Honor degrees during the undergraduate study in METU |

## Peer Reviewed Journal Papers

1. **Yildirimoglu, M.**, Ramezani, M., and Geroliminis, N. "Equilibrium Analysis and Route Guidance in Large-Scale Networks with MFD Dynamics", Transportation Research Part C: Emerging Technologies (available online).

2. **Yildirimoglu, M.**, Limniati, Y., and Geroliminis, N. "Investigating empirical implications of hysteresis in day-to-day travel time variability", Transportation Research Part C: Emerging Technologies, 55, 2015, pp. 340-350.

3. Keyvan-Ekbatani, M., **Yildirimoglu, M.***, Geroliminis, N., and Papageorgiou, M. "Multiple Concentric Gating Traffic Control in Large-Scale Urban Networks", IEEE Transactions on Intelligent Transportation Systems, PP (99), 2015, pp. 1-14. (*: joint first author).

4. **Yildirimoglu, M.**, and Geroliminis, N. "Approximating dynamic equilibrium conditions with macroscopic fundamental diagrams", Transportation Research Part B: Methodological, 70, 2014, pp. 186-200.

5. Yang, H., Ozbay, K., Ozturk, O., and **Yildirimoglu, M.** "Modeling work zone crash frequency by quantifying measurement errors in work zone length", Accident Analysis & Prevention, 55, 2013, pp. 192-201.

6. **Yildirimoglu, M.**, and Geroliminis, N. "Experienced travel time prediction for congested freeways", Transportation Research Part B: Methodological, 53, 2013, pp. 45-63.

**(Articles in progress)**

- Ramezani, M., **Yildirimoglu, M.**, and Geroliminis, N. "Within and Day to day dynamics for large scale networks with perimeter control and route choice", to be submitted to Transportation Research Part B.

- **Yildirimoglu, M.**, and Geroliminis, N. "Aggregated dynamic route choice patterns for large-scale mixed urban/freeway networks", to be submitted to Transportation Research Part B: Emerging Technologies.

# Peer Reviewed Conference Papers

1. **Yildirimoglu, M.**, Petit, A., Geroliminis, N., and Yanfeng, O. "Bus Service Design Under Demand Diversion and Dynamic Roadway Congestion Based on Aggregated Network Models", Accepted for Presentation at 95th Annual Meeting of the Transportation Research Board, Washington, D.C., 2016.

2. **Yildirimoglu, M.**, and Geroliminis, N. "Observing and Reconstructing Aggregated Dynamic Route Choice Patterns for Large-Scale Mixed Urban/Freeway Networks", Accepted for Presentation at 95th Annual Meeting of the Transportation Research Board, Washington, D.C., 2016.

3. **Yildirimoglu, M.**, and Geroliminis, N. "Regional Route Choice Behavior in Large-Scale Urban Networks", Presented at 4th Symposium of the European Association for Research in Transportation, Copenhagen, Denmark, 2015.

4. **Yildirimoglu, M.**, Ramezani, M., and Geroliminis, N. "Equilibrium Analysis and Route Guidance in Large-Scale Networks with MFD Dynamics", to be presented in 21st International Symposium on Transportation and Traffic Theory (ISTTT), Kobe, Japan, August 2015.

5. **Yildirimoglu, M.**, Ramezani, M., and Geroliminis, N. "A Rolling Horizon Approach for Route Guidance in Large-Scale Networks", Presented at 94th Annual Meeting of the Transportation Research Board, Washington, D.C., 2015.

6. **Yildirimoglu, M.**, Limniati, Y., and Geroliminis, N. "Linking the Properties of Mean-Variance Relations in Travel Times with the Hysteresis in Macroscopic Fundamental Diagram", Presented at 3rd Symposium of the European Association for Research in Transportation, Leeds, UK, 2014.

7. **Yildirimoglu, M.**, and Geroliminis, N. "Approximating Dynamic Equilibrium Conditions with Macroscopic Fundamental Diagrams", Presented at 93rd Annual Meeting of the Transportation Research Board, Washington, D.C., 2014.

8. Keyvan-Ekbatani, M., **Yildirimoglu, M.**, Geroliminis, N., and Papageorgiou, M. "Traffic Signal Perimeter Control with Multiple Boundaries for Large Urban Networks", Presented at Intelligent Transportation Systems (ITSC), 16th International IEEE Conference, The Hague, Netherlands, 2013.

9. **Yildirimoglu, M.**, and Geroliminis, N. "Approximating Dynamic Equilibrium Conditions with Macroscopic Fundamental Diagrams", Presented at 2nd Symposium of the European Association for Research in Transportation, Stockholm, Sweden, 2013.

10. **Yildirimoglu, M.**, and Geroliminis, N. "Experienced Travel Time Prediction in Congested Freeway Routes", Presented at 92nd Annual Meeting of the Transportation Research Board, Washington, D.C., 2013.

11. **Yildirimoglu, M.**, Koymans, A., and Geroliminis, N. "Exploring the Properties of Mean-Variance Relations in Freeway Travel Times", Presented at 92nd Annual Meeting of the Transportation Research Board, Washington, D.C., 2013.

12. **Yildirimoglu, M.**, and Geroliminis, N., "Experienced Travel Time Prediction in Freeway Systems", Presented at Intelligent Transportation Systems (ITSC), 15th International IEEE Conference, Anchorage, AK, 2012.

13. **Yildirimoglu, M.**, and Ozbay, K., " Comparison of Two Novel Travel Time Estimation Techniques Based on Probe Vehicle Data: Kriging Versus Nonlinear Programming-Based Approaches", Presented at 91st Annual Meeting of the Transportation Research Board, Washington, D.C., 2012.

14. **Yildirimoglu, M.**, and Ozbay, K., "Comparative Evaluation of Probe-Based Travel Time Prediction Techniques Under Varying Traffic Conditions", Presented at 91st Annual Meeting of the Transportation Research Board, Washington, D.C., 2012.

15. Ozbay, K., and **Yildirimoglu, M.**, "Comparison of Real-time Travel Time Estimation Using Two Distinct Approaches: Universal Kriging and Mathematical Programming", Presented at Intelligent Transportation Systems (ITSC), 14th International IEEE Conference, Washington, DC, 2011.

## Research Projects

| | |
|---|---|
| 2015 | **METAFERW: Modeling and controlling traffic congestion and propagation in large-scale urban multimodal networks** <br> *Sponsor: European Research Council, Starting Grant* |
| 2013-2015 | **Route Choice and Perimeter Control for Heterogeneous Congested Networks with parsimonious models** <br> *Sponsor: Swiss National Science Foundation* |
| 2013-2014 | **Development of perimeter traffic signal control for the city of Sydney** <br> *Sponsor: Roads and Maritime Services, New South Wales, Australia (collaboration with UNSW (Australia))* |
| 2012-2014 | **Coordinated ramp metering control with variable speed limits for Swiss freeways** <br> *Sponsor: Swiss Federal Roads Office* |
| 2011-2014 | **OPTIMUM - Optimized ITS-based tools for intelligent urban mobility** <br> *Sponsor: European Commission (collaboration with INRETS (France) and QUT (Australia))* |
| 2009-2011 | **Work zone safety analysis** <br> *Sponsor: New Jersey Department of Transportation (NJDOT)* |

## Academic Services

| | |
|---|---|
| Reviewer | Transportation Research Part B <br> Transportation Research Part C <br> Transportmetrica B <br> Journal of Intelligent Transportation Systems <br> EURO Journal on Transportation and Logistics <br> Transportation Research Board <br> IEEE Conference on Intelligent Transport Systems |
| Organizer | Member of local organizing committee of LATSIS Symposium 2012 - EPFL, Switzerland <br> Co-organizer of Swiss Transportation Research Conference (STRC) 2015 |

## Training Courses

| | |
|---|---|
| 2014 | **Advanced modeling and simulation of transportation networks, Italy**<br>*Instructors: Prof. Moshe Ben-Akiva, Prof. Ennio Cascetta* |
| 2012 | **Short course on dynamic traffic flow modeling and control, Greece**<br>*Instructor: Prof. Markos Papageorgiou* |

## Skills

| | |
|---|---|
| Language | Turkish (Native), English (Fluent), French (Fluent) |
| Computer | MATLAB, LaTeX, Aimsun, C++, Visual Basic, MS Office, |