MACROSCOPIC URBAN NETWORK DYNAMICS: ESTIMATION AND APPLICATIONS

A Dissertation
Presented to
The Academic Faculty

By

Rafegh Aghamohammadi

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in
Civil and Environmental Engineering

Georgia Institute of Technology

December 2021

MACROSCOPIC URBAN NETWORK DYNAMICS: ESTIMATION AND APPLICATIONS

Approved by:

Dr. Jorge A. Laval, Advisor
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. David Goldsman
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Srinivas Peeta
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Guanghui Lan
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Patricia Mokhtarian
School of Civil and Environmental Engineering
Georgia Institute of Technology

Date approved: December 3, 2021

To my parents and sister,

for their

unconditional love,

endless support and

empowering faith in me!

ACKNOWLEDGMENTS

Now that my PhD journey has come to an end, I would like to take this opportunity to appreciate the support and efforts of those who made this possible. I have felt blessed and lucky since the day I got admitted to the PhD program at a fine institution like Georgia Tech. This is in large part due to my world-class advisor, Professor Jorge Laval. I could not wish for a better mentor, and mere words cannot express my gratitude to him for his kindness, generosity, intelligence, critical thinking, guidance and support during the ups and downs of this journey. Thank you so much Dr. Laval for giving me this opportunity and bearing with me for the past four and half years despite all my shortcomings and setbacks. I have learned a lot from you, not only in terms of being a proficient researcher, but also a great deal of life lessons on how to keep an efficient work-life balance. I am pleased to have a role model like you and will always be a proud alumni of your lab, count on your guidance and friendship both in research and life, and remember your words and carry them with myself into my future career, especially this one:

"Brevity is the ultimate sophistication!" -Jorge A. Laval

A sincere debt of gratitude is extended to my dissertation committee, Prof. Srinivas Peeta, Prof. Patricia Mokhtarian, Prof. David Goldsman and Prof. Gunaghui Lan. I am honored to have their superb guidance, help, and support in my dissertation. Furthermore, I would like to thank Prof. Ludovic Leclercq of Gustave Eiffel University, who generously hosted me as a visiting researcher for 3 months in Lyon, France. I have fully enjoyed my collaboration with him and stay in Europe making great friends and visiting fabulous places. *Merci beaucoup pour votre gentillesse, Ludo!*

I have made wonderful friends for life at Georgia Tech including my labmates Hyun Woong Cho, Tu Xu and Hao Zhou. I have fully enjoyed our discussions on research

and also all the time that we have spent together making delightful memories. I am sure our friendships will last for many years to come and we will always find ways to meet no matter where we finally end up. I further owe my personal well-being during my PhD to my incredible friends all over the globe including but not limited to Salar Fattahi, Pouria Kourehpaz, Sina Sojoudi, Milad Jahed, Sina Sadeghi Mehr and Sergio Batista.

Atlanta will always hold a special place in my heart as the first city that I landed and lived in the US. I have so many pleasant memories in this city thanks to my amazing friends, who became like family thousands of miles away from home: my roommate Ashkan Ojaghi, Soheil Soltani, Ehsan Hesar and Fereshteh Shahmiri, Kia Mostaan, Sepideh Saremi and their adorable baby Artin, and Niki Navaei.

With all that being said, the main source of motivation and inspiration during my PhD and entire life has been my lovely family, my mom Giti, my dad Houshang and my one and only sister, Narmin. I am so lucky to have you three in my life and nothing can make up for all the sacrifices you have made and the hardships you have gone through for me to reach here. Your unconditional love and selfless support are the main reasons that I have been able to come along this way. I will always rely on you and make every possible effort to be a worthy son and brother to you.

Finally yet importantly, I need to recognize the Pink Floyd, whose epic music has been my refuge during difficult times. I owe you a great deal Roger Waters, David Gilmour, Rick Wright and Nick Mason! All in all, I will always try my best not to be *"Another Brick in the Wall"*!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

During the past decade there has been significant research efforts in developing traffic control and management methods based on an aggregated representation of traffic networks. In fact, the traditional link-level network representation imposes prohibitive computational costs for typical large-scale urban networks. Thankfully, it has been observed that at a macroscopic level, the relationship between any pair of network-average traffic variables can be described by simple functions called macroscopic fundamental diagrams (MFD). However, current MFD estimation methods were mainly conceived for individual arterial corridors and their application to urban networks has not been validated using extensive empirical data.

This dissertation fills this gap by extending current MFD estimation methods to large-scale real-life networks, while using empirical data from 41 cities around the world for calibration and validation. This dissertation further investigates the efficient application of MFD in travelers' route choice using the dynamic traffic assignment (DTA) methods and sets forth the discrete- and continuum-space DTA approaches are intrinsically similar and can be seen as equivalents on different aggregation levels, although they previously seemed to be the two extreme ends of the macroscopic DTA spectrum. A novel continuum-space DTA modeling framework consistent with the MFD theory and assumptions has been developed and a semi-Lagrangian solution method has been proposed by splitting up the network into smaller zones, which can be implemented for minimizing either the travel times of individual users or the total travel time of all users in the network. Finally, the potentiality of implementing the MFD in microscopic vehicular emissions estimation models has been explored.

The major findings of this dissertation are as follows. The empirical MFD validation results identify the most important challenges in both analytical and empirical MFD estimation approaches as: (i) the distribution of loop detectors within the links,

(ii) the distribution of loop detectors across the network, and (iii) the treatment of unsignalized intersections and their impact on the block length. The numerical experiment results using the proposed DTA framework indicate that partitioning the network into a finer grid of zones can yield more accurate results with respect to the approximated analytical solutions without significant loss of efficiency, and demonstrate the potential of application of this framework for real-life networks with arbitrary network and zone shapes. The comparison between the results and runtimes of the emissions estimations conducted in 4 different aggregation levels: lane, link, corridor, and network, reveals that the efficiency can be significantly improved by utilizing more aggregated network representation under some considerations. This will make the MFD a powerful tool for real-time emissions estimation and control.

CHAPTER 1
INTRODUCTION

Urban congestion is an overgrowing issue in many cities all over the globe. The continuously-increasing mobility demand in developing communities is resulting in (i) excessive delays and economic losses, (ii) uncurbed fuel consumption exacerbating the depletion of non-renewable energy resources, and (iii) emissions from the vehicles provoking local (air pollution, acid rain, health issues) and global (GHG[1] effect) damages. According to the 2019 INRIX traffic scorecard (Inrix, 2020), the average American driver lost 99 hours a year in traffic in 2019 costing $\$1,377$, and summing up to a total of $88 billion lost in time to congestion. Furthermore, the transportation sector accounted for the largest portion (28%) of total U.S. GHG emissions in 2018 according to a recent fact sheet published by the United States Environmental Protection Agency (EPA, 2020b). Between 1990 and 2018, GHG emissions in the transportation sector has increased more in absolute terms than any other sector (i.e. electricity generation, industry, agriculture, residential, or commercial) due in large part to the increased demand for both passenger and freight transportation. Therefore, from a sustainable development perspective, such effects must be studied and controlled in order to prevent any hindrance to the development process.

The financial, economic and environmental costs associated with congestion have been vastly studied and control and management methods have been proposed to improve the level of service and mitigate these costs. Dynamic traffic assignment (DTA) models have become the mainstream method for urban planning and traffic control purposes, either under user equilibrium (UE) or system optimum (SO) objectives. However, the prohibitive computation times necessary to achieve equilibrium

[1] GreenHouse Gas

on large-scale networks is one of the main challenges faced by practitioners in the real-time implementation of DTA-based control policies (Ziliaskopoulos et al., 2004). The main objective of most traffic control strategies has been to alleviate congestion and to reduce the delay incurred by travelers and emission and fuel consumption control have only gained interest in the past few years. However, the optimal emission and fuel consumption strategy may not lead to minimal delay and congestion (Csikós and Varga, 2011). Therefore, the optimal control strategy must be a trade-off between optimum emission and optimum delay conditions.

Although most of the current traffic control and management methods rely on a link-level representation of the transportation networks, significant amount of research efforts has been devoted in the past decades to develop a better understanding of the nature and dynamics of urban congestion and the way it propagates through the urban street network. The idea of an aggregated relationship between average density and flow over the network dates back to 1960s (Smeed, 1967; Godfrey, 1969). Later, many researchers have elaborated on this idea and proposed various models to describe the relationship between the network-wide traffic variables (see e.g., Herman and Prigogine, 1979; Mahmassani et al., 1984; Ardekani and Herman, 1987).

Daganzo (2007) claims that the relationship between the average flow and density is an invariant property of the network regardless of the demand and OD patterns. The empirical verification of existence of such a network-level macroscopic fundamental diagram (MFD) for urban city street networks by Geroliminis and Daganzo (2008) has opened the doors for its application in various traffic control aspects. For example, DTA models based on an aggregated representation of the network have become popular in the past decade. These models typically partition the urban network into a number of disjoint regions, where the average regional flow, $\bar{Q}(n)$, is a function of the accumulation n. On the other hand, there is a vast body of literature dealing with DTA models in the continuum space consisting of a conservation law in

two dimensions, supplemented with a Hamilton-Jacobi equation for the route choice. These models rely on existence of a speed-density equilibrium relationship, which now can be thought of as the MFD although the connection is not well understood yet.

This dissertation work will study the connection between the MFD-based and continuum-space macroscopic DTA approaches, bring light to the shortcomings and gaps in the state of the art, investigate ways to address these shortcomings, and develop more efficient traffic control methods based on a simplified aggregated representation of the traffic dynamics inside the urban networks. Toward this end, Chapter 2 will critically review the literature on both discrete- and continuum-space DTA models and provide insights into the shortcomings in the literature. In light of these insights, Chapter 3 will develop a novel continuum-space DTA modeling framework, which is consistent with the MFD theory and assumptions.

This simplified multi-reservoir representation has the potential to be utilized for large-scale network planning and control purposes whether aiming to mitigate congestion and reduce the delays or network-wide emissions and fuel consumption estimation and control. Chapter 4 will provide a method to minimize the total travel cost in the networks by establishing semi-DSO conditions in the DTA framework proposed in the previous chapter. Furthermore, Chapter 5 will try to implement the multi-reservoir traffic representation into microscopic emissions estimation using the EPA's state-of-the-science emissions model, MOVES (MOtor Vehicle Emission Simulator) (EPA, 2020a).

All the previous chapters rely on the assumption that the MFDs are already known, while estimation of the MFD might not be straightforward for real-life networks. Chapter 6 will try to address this issue by extending an analytical MFD estimation method to estimate the network MFDs and estimating the corresponding network parameters by fitting the model to empirical network-wide traffic data from 41 cities around the world. The takeaways from the results presented in this chapter

will touch on the required consideration for both empirical and analytical MFD estimation methods. Last but not least, Chapter 7 will present the concluding remarks of this dissertation work and identify the future research directions.

CHAPTER 2
DYNAMIC TRAFFIC ASSIGNMENT USING THE MACROSCOPIC FUNDAMENTAL DIAGRAM

This chapter is modified from the following published paper:

Rafegh Aghamohammadi and Jorge A. Laval (2020b). "Dynamic traffic assignment using the macroscopic fundamental diagram: A Review of vehicular and pedestrian flow models". In: *Transportation Research Part B: Methodological* 137, pp. 99–118

2.1 Introduction

Traditional dynamic traffic assignment (DTA) models based on a link-level representation of the network are becoming the mainstream method for urban planning, either under user equilibrium (UE) or system optimum (SO) objectives. Despite the vast body of literature in this field, there are still some important drawbacks. Arguably, the main challenge faced by practitioners is the prohibitive computation times necessary to achieve equilibrium on large-scale networks, which typically have hundreds of thousands of links and nodes (Ziliaskopoulos et al., 2004). This makes calibration and validation of such large networks a daunting and intractable task in real time.

The DTA models generally follow one of the following objectives, either (i) *dynamic user equilibrium (DUE)*, where each user tries to minimize his/her own travel cost, or (ii) *dynamic system optimum (DSO)*, where users are guided in a way to minimize the total travel cost incurred by all users in the system in a cooperative manner. DUE objective can be further divided up into (i) reactive dynamic user equilibrium (RDUE), which minimizes *instantaneous* travel cost of users, and (ii) predictive dynamic user equilibrium (PDUE), where each user tries to minimize their

expected travel cost. Once the equilibrium condition is satisfied through either one of the objectives, the travel cost on any used or unused possible route for each user will be equal, i.e. no user can unilaterally reduce his/her travel cost by shifting to another route. For a detailed review of different DTA methods please refer to Peeta and Ziliaskopoulos (2001).

In the past 50 years, various theories have been proposed to describe the traffic flow of urban networks at an aggregate level (Smeed, 1967; Herman and Prigogine, 1979; Herman and Ardekani, 1984; Mahmassani et al., 1984). The recent empirical verification of the existence of a network-level Macroscopic Fundamental Diagram (MFD) on congested urban areas in 2008 opened up a new paradigm (Daganzo, 2007; Geroliminis and Daganzo, 2008). The average flow MFD

$$Q = Q(n), \qquad \text{(average flow MFD)} \tag{2.1}$$

gives the average flow Q on network as a function of the number of vehicles inside the network, n, arguably independently of trip origins and destinations, and route choice. This makes the MFD an invaluable tool to overcome the difficulties of traditional planning models. For example, on a single region one can construct a reservoir type model to approximate the evolution of the accumulation of vehicles inside the reservoir. We can then estimate any average traffic variable of interest, provided that we know the outflow MFD, namely the o-MFD, defined as

$$o(n) = Q(n)\frac{L}{\ell}, \qquad \text{(o-MFD)} \tag{2.2}$$

where L corresponds to the length of network (in total lane-miles) and ℓ is the trip length, typically assumed identical for all commuters. The o-MFD, $o(n)$, gives the number of trip completions per unit time, as a function of n. The simplest model for the traffic dynamics inside the network is the simple reservoir (or bathtub) model,

as shown in Fig. 2.1(a), which simply states the conservation of vehicles via the following ordinary differential equation (ODE):

$$\begin{cases} n'(t) = \lambda(t) - o(n), & \text{(reservoir dynamics)} \quad (2.3a) \\ n(0) = n_0, & \text{(initial conditions)} \quad (2.3b) \end{cases}$$

where $\lambda(t)$ is the demand inflow into the network at time t, primes denote differentiation and n_0 specify the initial conditions. Fig. 2.1(b) illustrates a typical o-MFD, where μ is the capacity of the reservoir in vehicles per hour and n_j is the maximum accumulation (jam accumulation) of the reservoir. Fig. 2.1(c) shows the resulting average speed vs. accumulation diagram for the MFD in part (b) of the figure, where u is the free-flow speed over the network.

The shape of MFD depends on network topology and control parameters such as block length, existence of turn-only lanes, and traffic light settings. Up until recently, estimating the MFD was not an easy task because the method of cuts in Daganzo and Geroliminis (2008) becomes intractable for real-life networks and one needs to resort to simulation methods, which defeats the purpose of macroscopic modeling. However, Laval and Castrillón (2015) shows that (the probability distribution of) the MFD can be well approximated by a function of mainly two parameters: the density of traffic lights and the mean red to green ratio across the network. Since these are

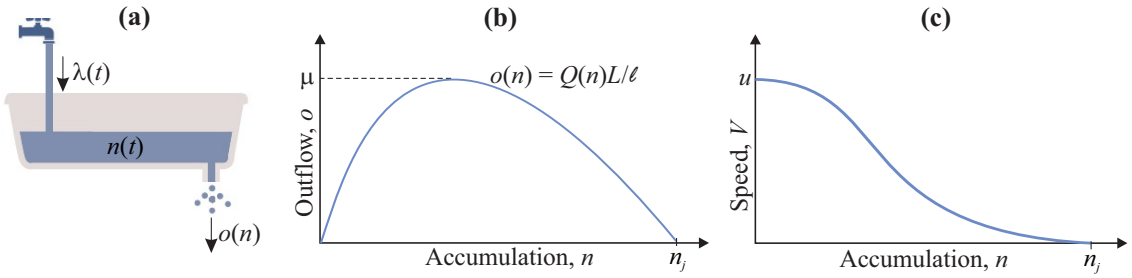

Figure 2.1: Illustration of: (a) reservoir model, (b) typical o-MFD diagram and (c) resulting Speed-Accumulation MFD.

observable parameters, we now have a simple method for estimating the MFD on arbitrary road networks.

We conclude that it is now possible to formulate DTA models based on a macroscopic representation of the network. This representation may be expressed in (i) discrete space, where the modeling region is divided in a finite number of zones with well-defined MFDs and the traffic dynamics given by the conservation ODEs, Eq. (2.3); in (ii) continuum space, where there is a MFD at each (x, y)-point.

It turns out that there is a large body of related literature prior to 2008, when the MFD was verified empirically, that dealt with similar problems. In particular, we have DTA models in continuum space that consist of a conservation law in two dimensions, supplemented with a Hamilton-Jacobi equation for the route choice. Mathematically, these models are identical to the continuum-space pedestrian flow models in the literature, which is why we have included the pedestrian literature here. All of these models rely on the existence of a speed-density (speed-accumulation) equilibrium relationship, which could be thought of as the MFD as mentioned in some of the studies in the literature. We will delve into the rationale of considering the speed-density function in the continuum-space models as MFD in more detail in the pertinent section.

Our purpose here is to review the literature in light of the aforementioned considerations. In particular, we would like to identify the current challenges to bringing to practice efficient macroscopic DTA models for cities.

It is worth mentioning that in order to keep the consistency of the notation through this chapter, we have changed the notation of some of the studies reviewed here. Moreover, a **bold-typed reference** in this chapter indicates the outset of its review.

2.2 Discrete-space Models

In discrete-space models, the modeling region is divided into a finite number of zones describing a tessellation of the (x, y)-plane, where each zone has a well-defined MFD with traffic dynamics given by the conservation ODE, Eq. (2.3). Several clustering methods for the heterogeneous networks have been proposed in the literature (see e.g., Ji and Geroliminis, 2012; Saeedmanesh and Geroliminis, 2017; Lopez et al., 2017). Here, our focus will be on the studies dealing with the DTA problem using MFD in the discrete space. A standard assumption in the literature is that in congestion, demand should be restricted by the congested branch of the o-MFD, $S(n)$:

$$\lambda(t) \leqslant S(n(t)). \qquad \text{(supply constraint)} \tag{2.4}$$

To reflect that, in congestion, the inflow cannot exceed the outflow. This is consistent with the idea that all links in the region, including those on its perimeter and those where vehicles enter the network, have the same (congested) flow and therefore if demand exceeds it, a queue accumulates outside the region.

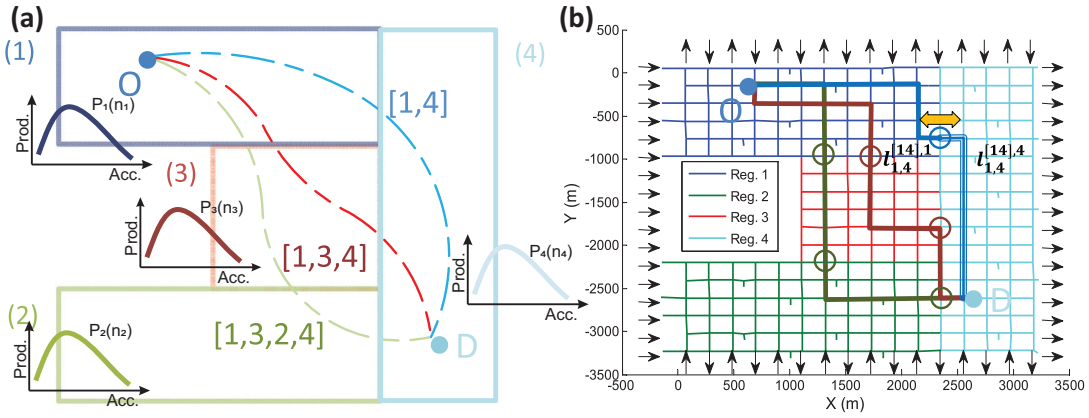

Figure 2.2: Various paths from region 1 to region 4 shown in: (a) regional and (b) link-level representations (Yildirimoglu and Geroliminis, 2014).

As the first explicit discrete-space DTA model, **Yildirimoglu and Geroliminis (2014)** incorporates MFD into a DTA model in a heterogeneous urban network divided into homogeneous regions with low-scatter MFDs and finds the dynamic stochastic user equilibrium (DSUE) numerically. Each path is defined as the sequence of regions from the origin to the destination. Fig. 2.2(a) shows three possible paths from region 1 to region 4 as sequences of different regions. Since the trip lengths are time-dependent and there may be various link sequences complying with the path definition, stochastic network loading (SNL) is implemented to address variable trip lengths within and between regions.

In order to determine the shortest path, the algorithm developed by Chabini (1998) is implemented to determine the sequence of links leading to the minimum experienced travel time inside a sequence of regions using the link-level representation of the network, as seen in Fig. 2.2(b), while the speed on all links in a region is given by the average speed of the MFD, based on the accumulation level. Then, the probability of using each path is calculated using a multinomial logit discrete choice model based on the calculated minimum travel times. Traffic equilibrium is formulated as a fixed-point problem that is solved in an iterative manner using method of successive averages (MSA), where the step size is determined a priori.

The proposed model is tested on an arbitrary network divided into 4 regions, where all the origins are in regions one and two and all the destinations are in region four, see Fig. 2.2. The results are compared to those of a micro-simulator and a one-shot (non-iterative) assignment model with the aggregate dynamics, where the path choice is updated periodically in a way that the vehicles departing at each period choose their path based on instantaneous travel times measured at the previous period. It is demonstrated that the one-shot assignment model does not satisfy equilibrium conditions and overstate the network congestion, whereas the proposed model results in DSUE condition and less congestion.

Yildirimoglu et al. (2015) extends the work in Yildirimoglu and Geroliminis (2014) to a route guidance system based on dynamic system optimum conditions, where each region is divided into a number of sub-regions with well-defined MFDs and routes are redefined as sequences of sub-regions instead of sequences of links in Yildirimoglu and Geroliminis (2014). Two macroscopic traffic models are developed in region-level and subregion-level, where the latter deploys a more detailed approach. The transfer of variables from the subregion-based model to the region-based model has also been described.

In this study, DTA has been achieved using three different approaches: (i) by establishing PDUE conditions in subregion-level, (ii) by establishing DSO conditions in subregion-level, and (iii) by providing travelers with route guidance (RG) information based on DSO conditions in region-level, where the route guidance commands are applied in subregion-level. DSO conditions require equal and minimal marginal travel times on alternative routes at the same departure time. Several numerical studies have been conducted in order to assess different DTA approaches. The results reveal that RG approach outperforms DSO and PDUE approaches, except in very high demands which might be due to the inconsistency between subregion-level and region-level accumulation values near to the jam accumulation.

Stalling problem is the most significant potential drawback of the models proposed in Yildirimoglu and Geroliminis (2014) and Yildirimoglu et al. (2015), which can happen under intensive demand when the number of vehicles inside a (sub)region reaches the jam accumulation. In this case, all the vehicles in the (sub)region will have zero speed based on the average-speed MFD. Therefore, all the vehicles inside the (sub)region will stall, no vehicle will be able to move out of the (sub)region to resolve the jam, and the model will result in an insoluble unrealistic gridlock, which will spread to the neighboring (sub)regions.

Haddad et al. (2013) tries to deal with the macroscopic traffic modeling and control of a large-scale mixed transportation network consisting of an urban network and a freeway. The urban network is divided into two regions, where in each region the dynamics of traffic flow are represented by a well-defined MFD, and the traffic dynamics of the freeway are modeled according to a simplified version of Daganzo's Cell Transmission Model (CTM; Daganzo, 1994a; Daganzo, 1995), called Asymmetric Cell Transmission Model (ACTM; Gomes and Horowitz, 2006). They investigate and compare the effect of implementing RDUE and DSO route choice strategies on network performance using several control policies. The results show that integrating DSO within the centralized control scheme can increase the network performance by 12 percent compared to the RDUE route choice.

Sossoe and Lebacque (2017) develops an RDUE model for bi-directional traffic flow based on the traffic flow model proposed in Saumtally et al. (2013) and Sossoe et al. (2015). The urban network is decomposed into zones, where each zone is meshed into quadrangular cells, as shown in Fig. 2.3, and four inflow and outflow directions

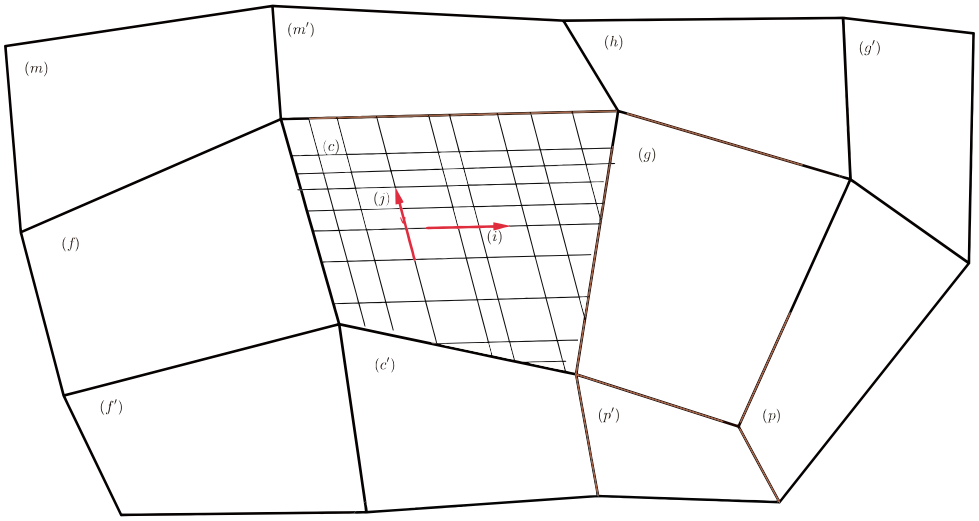

Figure 2.3: Network disaggregated into zones with quadrangular mesh of cells (Sossoe and Lebacque, 2017).

are considered for each cell. Cell flows and exit speeds in each direction are calculated by cell densities using intersection traffic flow model rules.

Based on the position of the target cell regarding the origin cell, only two possible outflow directions are assumed from the origin cell toward the designated target cell. Two possible paths are considered based on each of the possible directions and the turning proportions in each cell are computed using a weighted logit model, where the cost of each path is assumed to be the instantaneous travel time.

2.2.1 Numerical Solutions

In the case of discrete-space models, numerical methods are needed to solve (i) the single-MFD loading problem, and (ii) the network loading problem and equilibrium.

For the solution of (i), the main numerical component is the conservation ODE (2.3), which is typically solved with Euler's method, which is a first-order finite differences scheme:

$$n(t + \Delta t) = n(t) + \Delta t(\lambda(t) - o(n)) \tag{2.5}$$

which converges linearly to the true solution as $\Delta t \to 0$.

Laval et al. (2017) derived analytical solutions for this problem for a family of demand curves well suited to model rush-hour periods, under the Greenshields' (parabolic) MFD. These include second degree polynomial, exponential, and logistic functions. For a general MFD, they found analytical solutions for the autonomous case, i.e. when the demand is time-independent.

In order to solve (ii), we have seen that the methods used are similar to the ones used in traditional DTA models with link exit-flow functions (Merchant and Nemhauser, 1978; Friesz et al., 1989; Carey and McCartney, 2004). Although the literature review on this vast subject is out of the scope of this manuscript, it is

important to note that the lessons learned from those models can be directly applied to these simplified macroscopic models.

2.3 Continuum-space Models

The continuum-space models are useful when the network is dense such that both the distance between road sections and their longitude is small compared to the size of the region (see e.g., Newell, 1980; Romero and Benitez, 2010). In this case, the network can be considered as a continuum, where the users (vehicles or pedestrians) can circulate at any point $\mathbf{x} = (x, y)$ of the Euclidean, two-dimensional domain $\Omega \subset \mathbb{R}^2$. In this section, we review the large body of literature on continuum-space models and try to investigate the connection between these models and aforedescribed discrete-space DTA models.

Continuum-space vehicular DTA and pedestrian flow models are mathematically similar, as mentioned earlier. The first formulation of this type of models was proposed in the context of pedestrian dynamics by **Hughes (2002)**. In this study, a two-dimensional walking infrastructure is represented as a continuum with domain $\Omega \subset \mathbb{R}^2$, as seen in Fig. 2.4. Γ_o denotes the outer spatial boundary, Γ_h is the hard boundary of any obstruction at which no traveler is allowed to enter or exit the walking facility, and Γ_d represents the boundary of the destination area. In more general cases, where there are more than one destination area in the study, Γ_d^i will represent the boundary of each destination area, i.

Let $\rho(x, y, t)$ be the time-varying density of travelers at location (x, y) at time t, $\mathbf{u} = (\mathbf{u_1}(\mathbf{x}, \mathbf{y}, \mathbf{t}),\ \mathbf{u_2}(\mathbf{x}, \mathbf{y}, \mathbf{t}))$ be the velocity vector, and $\mathbf{f}(x, y, t) = (f_1(x, y, t),\ f_2(x, y, t)) \equiv \rho\mathbf{u}$ be the flux vector. Similar to the conservation of mass law in fluid dynamics, the conservation of travelers yields

$$\rho_t + \nabla \cdot \mathbf{f} = 0, \qquad \text{(2D conservation law)} \tag{2.6}$$

where $\nabla \cdot \mathbf{f} = \partial f_1/\partial x + \partial f_2/\partial y$ is the (spatial) divergence of $\mathbf{f}$. This equation follows simply by equating the net flow of pedestrians into a small region to the time rate of accumulation of travelers in the region, and letting the area of the region shrink to zero (Hughes, 2002). Further derivation of this equation can be found in texts on fluid mechanics under the name of *the continuity equation*. In order to model the nature of the pedestrian flow, Hughes makes three hypotheses. The first hypothesis asserts that the speed of pedestrians at each point is an isotropic function of the density at that point and time

$$\|\mathbf{u}\| = V(x, y, \rho), \tag{2.7}$$

where $\|\mathbf{u}\| \equiv \sqrt{u_1^2 + u_2^2}$ is the norm of the velocity vector and $V(x, y, \rho)$ represents the *location-specific* speed-density relationship. The general form of this relationship in the literature, especially the continuum-space vehicular DTA models, is

$$V(x, y, \rho) = U_f(x, y)e^{-\beta\rho^2}, \qquad \text{(isotropic speed-density relationship)} \tag{2.8}$$

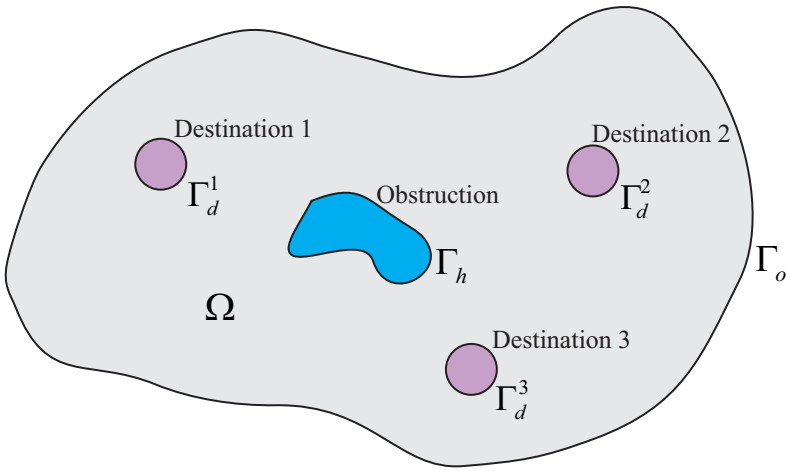

Figure 2.4: Illustration of continuum space with three destination areas and one obstruction (lake).

where $U_f(x, y)$ gives the free-flow speed as an increasing function of the distance to the destination and β is a positive parameter influenced by the road condition and etc. After the empirical evidence in 2008 in support of the existence of a MFD on urban networks, a number of recent studies have interpreted the speed-density relationship as the MFD (see e.g., Du et al., 2013; Du et al., 2015; Long et al., 2017), but no attempts exist to verify the assumptions of MFD theory. We will further investigate this claim in the discussion section of this chapter.

In the second hypothesis, Hughes (2002) defines a potential function, $\phi(x, y, t)$, that represents the cost of reaching the destination starting from (x, y, t) along the minimum cost path. Therefore, the motion of any pedestrian is in the direction with maximum potential reduction, i.e. in the direction perpendicular to the isopotential curves. Since the gradient vector of the potential function, $\nabla\phi \equiv (\frac{\partial\phi}{\partial x}, \frac{\partial\phi}{\partial y})$, is perpendicular to these isopotential curves in the direction of maximum increase, it follows that the flux vector, and as a result the velocity vector, are parallel to the gradient of the cost potential and in the *opposite* direction, i.e.:

$$\mathbf{u} \;//\; \mathbf{f} \;//\; -\nabla\phi. \qquad \text{(optimum movement direction)} \tag{2.9}$$

where $//$ means "is parallel to". Thus, the velocity vector is given by

$$\mathbf{u} = -\frac{\nabla\phi}{\|\nabla\phi\|}\, V, \qquad \text{(speed vector parallel to streamlines)} \tag{2.10}$$

where $\nabla\phi/\|\nabla\phi\|$ is the unit gradient vector of the potential function. If the cost potential is the travel time, the cost per unit distance is the pace, and we have :

$$\|\nabla\phi\| = 1/\|\mathbf{u}\| = 1/V(\rho) \tag{2.11}$$

which is an Eikonal PDE. The third hypothesis defines a discomfort function, $g(\rho)$, to account for the behavior of the pedestrians to avoid higher densities, which satisfies

$$gV = 1/||\nabla\phi||, \qquad \text{(discomfort function } g\text{)} \tag{2.12}$$

where $g \geqslant 1, \partial g/\partial\rho \geqslant 0$. However, Huang et al. (2009b) explain that assuming the discomfort factor as an increasing function of density makes pedestrians perceive faster movement at higher densities, which is counterintuitive. Instead, it proposes $\partial g/\partial\rho \leqslant 0$.

With all, Hughes' (2002) model can be expressed as

$$\text{(Hughes' model)} \quad \begin{cases} \rho_t - \nabla \cdot \left(\rho\, g\, V^2\, \nabla\phi\right) = 0, & \text{(2.13a)} \\ gV = 1/||\nabla\phi||, & \text{(2.13b)} \end{cases}$$

which is not a hyperbolic system, as customary in the 1-D traffic flow literature.

Furthermore, Hughes extended his model to incorporate multiple pedestrian types representing different walking characteristics and destinations. Segregated potential functions are defined for each pedestrian type and the same hypotheses are developed for different pedestrian types. The most important point in this extension is that the speed of a pedestrian type at each point and time is determined by the total density of all pedestrian types rather than the density of a single pedestrian type.

Although Hughes' study does not clarify the physical interpretation of the potential function and the route choice strategy, it has been the seminal work in the area of dynamic traffic assignment using continuum space, for both pedestrian and vehicular traffic. The following subsections will review the existing literature in this field.

2.3.1 Reactive Dynamic User Equilibrium Models

In the Reactive Dynamic User Equilibrium (RDUE) problem, travelers choose the route that minimizes the instantaneous travel cost and change their choice en route as a result (Boyce et al., 1995). **Huang et al. (2009b)** revisited Hughes' (2002) model and demonstrated that the route choice strategy in the model satisfies the RDUE principle. To see this, let $c(x, y, t)$ be the local cost per unit distance of movement,

$$c = \|\nabla\phi\|, \tag{2.14}$$

which is an Eikonal equation. It shows that to minimize the instantaneous walking cost of a traveler, the route choice strategy should satisfy

$$c\,\frac{\mathbf{f}}{\|\mathbf{f}\|} + \nabla\phi = 0. \qquad \text{(reactive route-choice strategy)} \tag{2.15}$$

The authors prove that if ϕ is set to zero for $(x, y) \in \Gamma_d$, the potential function can be interpreted as the minimum instantaneous travel cost to the destination. A general modeling structure consisting of a set of partial differential equations (PDE) is proposed as

$$\begin{cases} \rho_t + \nabla \cdot \mathbf{f} = 0, & \text{(2.16a)} \\ \|\mathbf{f}\| = \rho V, & \text{(2.16b)} \\ c\dfrac{\mathbf{f}}{\|\mathbf{f}\|} + \nabla\phi = 0, & \text{(2.16c)} \end{cases}$$

subject to appropriate initial and boundary conditions, typically:

$$\begin{cases} \rho(x,y,0) = \rho_0(x,y), & \forall (x,y) \in \Omega, \quad (2.17a) \\ \mathbf{f} \cdot \hat{\mathbf{n}}(x,y) = q(x,y,t), & \forall (x,y) \in \Gamma_o, \quad (2.17b) \\ \phi = 0, & \forall (x,y) \in \Gamma_d, \quad (2.17c) \end{cases}$$

where $\hat{\mathbf{n}}(x,y)$ is the unit normal vector toward Ω on Γ_o, $q(x,y,t)$ denotes the demand flow crossing Γ_o and $\rho_0(x,y)$ is the initial density inside Ω. **Jiang and Zhang (2011)** introduces a non-decreasing discomfort function and finds the direction of movement of pedestrians, $\hat{d}$, using

$$\hat{d} = \frac{\mathbf{u}}{\|\mathbf{u}\|} = \frac{\mathbf{f}}{\|\mathbf{f}\|} = \frac{\mathbf{w_1} + \epsilon \mathbf{w_2}}{\|\mathbf{w_1} + \epsilon \mathbf{w_2}\|}, \qquad \epsilon \geqslant 0, \quad (2.18)$$

where vectors $\mathbf{w_1} = -\frac{\nabla \phi}{\|\nabla \phi\|}$ and $\mathbf{w_2} = -\frac{\nabla g}{\sqrt{1+\|\nabla g\|^2}}$ are directed along the dynamic paths of the shortest instantaneous travel time and the higher comfort level, respectively, and ϵ is a positive constant representing the psychological influence of avoiding higher densities. The authors reconstruct the pedestrian flow model, Eq. (2.16), by replacing Eq. (2.16c) with Eq. (2.18) subject to the initial and boundary conditions of Eq. (2.17) and adding

$$\mathbf{f} \cdot \hat{\mathbf{n}} = 0, \qquad \forall (x,y) \in \Gamma_h, \quad (2.19)$$

to reflect that no traveler enters or exits the obstructions inside the continuum domain.

The concept of RDUE in continuum space has been adapted for the dynamic traffic assignment of vehicles inside an urban network by **Jiang and Zhou (2014)** for the first time. The urban network is considered as a 2-D continuum with a configuration similar to Fig. 2.4, where the destination areas are the central business districts (CBD) inside the urban network. The local travel cost per unit distance is

reformulated as

$$c(x,y,t) = \kappa\left(\frac{1}{U} + \pi_0\rho^2\right), \qquad \text{(general local cost function)} \quad (2.20)$$

where κ is the value of time and π_0 is a nonnegative constant interpreting other costs, such as avoiding higher densities. The resulting model is the following conservation law (CL):

$$CL: \begin{cases} \rho_t + \nabla\cdot\mathbf{f} = q, & \forall(x,y)\in\Omega, \quad (2.21a)\\ \mathbf{f} = -\|\mathbf{f}\|\dfrac{\nabla\phi}{\|\nabla\phi\|}, & \forall(x,y)\in\Omega, \quad (2.21b)\\ c = \|\nabla\phi\|, & \forall(x,y)\in\Omega, \quad (2.21c) \end{cases}$$

subject to initial and boundary conditions similar to Eq. (2.17), where Eq. (2.17b) is replaced by Eq. (2.19).

Bi-directional pedestrian flow models

In parallel, a number of studies have been conducted to investigate the interaction of bi-directional pedestrian flows. **Jiang et al. (2009)** develops the first bi-directional pedestrian flow model, where two pedestrian groups, a and b, have local walking speeds U^a and U^b dependent on the densities and direction of movement of both pedestrian groups, i.e.:

$$\begin{cases} U^a(\rho^a,\rho^b,\Psi) = U_f\ e^{-\alpha\left(\rho^a+\rho^b\right)^2}\ e^{-\beta(1-\cos(\Psi))(\rho^b)^2}, & (2.22a)\\ U^b(\rho^a,\rho^b,\Psi) = U_f\ e^{-\alpha\left(\rho^a+\rho^b\right)^2}\ e^{-\beta(1-\cos(\Psi))(\rho^a)^2}, & (2.22b) \end{cases}$$

where $\rho^a(x,y,t)$ and $\rho^b(x,y,t)$ are densities of each group, $\Psi(x,y,t)$ is the intersecting angle between two pedestrian groups, U_f is the free-flow walking speed of pedestrians, and α and β are the model parameters. This study also assumes that the walking

cost is equal to the travel time. Using these definitions and assumption, the authors propose segregated systems of PDEs and initial and boundary conditions similar to Eqs. (2.16) and (2.17), respectively, for each pedestrian group. The authors observed that the pedestrians walking to the same destination form stripes, which shows the cooperative strategy of pedestrians moving in different directions, as discussed in Daamen and Hoogendoorn (2003). All later bi-directional pedestrian flow models in the literature are based on the model proposed in this study and suggest improvements to the model or its numerical solution.

Huang et al. (2009a) extends the bi-directional pedestrian flow model developed by Jiang et al. (2009) by introducing look-ahead behavior in order to induce a viscosity effect on movement patterns of pedestrians. The authors first define the look-ahead displacement as

$$(\delta x^i, \delta y^i) = \epsilon \, \frac{\mathbf{f^i}}{\|\mathbf{f^i}\|}, \qquad\qquad i \in \{a, b\}, \tag{2.23}$$

where $\epsilon > 0$ is a small look-ahead distance of a pedestrian following the crowd in front of him corresponding to the traffic condition ahead at location $(x+\delta x^i, y+\delta y^i)$. The walking speed function with look-ahead effect for each group of pedestrians is given by

$$\begin{cases} \ddot{U}^a(x,y,t) = U^a\left(\rho^a(x+\delta x^a, y+\delta y^a, t), \rho^b(x,y,t), \Psi\right), & \text{(2.24a)} \\ \ddot{U}^b(x,y,t) = U^b\left(\rho^a(x,y,t), \rho^b(x+\delta x^b, y+\delta y^b, t), \Psi\right), & \text{(2.24b)} \end{cases}$$

where U^a and U^b are as defined in Eq. (2.22). The rest of the model is the same as the model developed in Jiang et al. (2009). **Xiong et al. (2011)** proposes a high-order computational scheme for the numerical solution of the bi-directional pedestrian flow model proposed in Jiang et al. (2009). This study found that the high-order scheme

can obtain the same resolution on much coarser meshes with less computation time, compared with the lower-order schemes.

Jiang et al. (2012) revisits the bi-directional pedestrian flow model proposed in Jiang et al. (2009) and introduces a more general local cost function, rather than assuming it to be equal to the travel time, as

$$c^i(x,y,t) = \frac{1}{U^i(\rho^a,\rho^b,\Psi)} + \gamma\, g^i(\rho^a,\rho^b,x,y,t), \qquad i \in \{a,b\}, \tag{2.25}$$

where γ is a weighted factor and g^i represents the discomfort function for pedestrian group i. Then, the authors reconstruct their previous model by substituting the local cost function with Eq. (2.25). They claim that numerical results of the model conform well to the experimental data and show that the model is rational and efficient.

2.3.2 Predictive Route Choice

In the Predictive Dynamic User Equilibrium (PDUE) problem travelers choose their routes to minimize the actual travel cost they will experience, in contrast to the RDUE problem, where travelers seek to minimize their instantaneous travel cost. Thus, the cost potential function $\phi(x,y,t)$ in the continuum-space PDUE models represents the *actual* travel cost (to the destination incurred by a traveler who leaves location (x,y) at time t toward the destination). The first continuum-space pedestrian PDUE model is developed by **Hoogendoorn and Bovy (2004)** for multiple pedestrian groups, where each group has a different destination, as

$$\begin{cases} \rho_t^i + \nabla\cdot\mathbf{f^i} = q^i(x,y,t), & \text{(CL equation)} \quad (2.26a) \\ -\phi_t^i = \min\limits_{\tilde{\mathbf{u}}}\{\mathcal{L}(x,y,t,\tilde{\mathbf{u}}) + \tilde{\mathbf{u}}\cdot\nabla\phi^i\}, & \text{(HJB equation)} \quad (2.26b) \\ \mathbf{u^i} = arg\min\{\mathcal{L}(x,y,t,\tilde{\mathbf{u}}) + \tilde{\mathbf{u}}\cdot\nabla\phi^i\}, & (2.26c) \end{cases}$$

where, $\mathcal{L}$ is the running cost, $\tilde{\mathbf{u}}$ is any possible velocity vector and $\mathbf{u^i}$ is the optimal velocity vector of group i. This model is comprised of two coupled PDEs; the first PDE is a CL equation and the second PDE is a Hamilton-Jacobi-Bellman (HJB) equation. In order to solve the model numerically, the authors presented a heuristic iterative approach, where the HJB and the CL equations are calculated in each iteration, respectively. Thus, the minimum value problem has to be solved twice in each iteration.

Jiang et al. (2011) proposes a PDUE model for vehicular traffic in an urban network with only one CBD based on the seminal work of Hughes (2002). The local cost per unit distance is redefined as

$$c(\rho) = \kappa \left(\frac{1}{V(\rho)} + \pi\left(\rho\right) \right), \qquad \text{(general local cost function)} \tag{2.27}$$

where $\pi(\rho)$ is a function representing other density-dependent costs. The authors assert that the PDUE principle is satisfied, if the following conditions are met

$$\begin{cases} (u_1, u_2, 1) \parallel -\bar{\nabla}\phi, & \text{(2.28a)} \\ \|\bar{\nabla}\phi\| = \dfrac{cV}{\sqrt{V^2+1}}, & \text{(2.28b)} \end{cases}$$

where $\overline{\nabla}\phi = \left(\frac{\partial\phi}{\partial x}, \frac{\partial\phi}{\partial y}, \frac{\partial\phi}{\partial t}\right)$ is the potential gradient including the time component. Then, they develop a vehicular PDUE model based on these definitions and assumptions. Nevertheless, **Du et al. (2013)** has revisited this model and claimed that the path-choice strategy is not appropriate, e.g., when the cost function is time-independent, $\phi_t = 0$, the condition stated by Eq. (2.28a) cannot be satisfied. This study also indicates that there is an inconsistency in the units of one the PDE equations in the model and they prove that this equation is redundant.

To resolve these issues, a special case is considered, where the travel cost is equal to the time, and later a general PDUE model is proposed, whereas the local travel cost per unit distance is given by Eq. (2.27). The final model consists of a conservation law (CL) part similar to Eq. (2.21) and a Hamilton-Jacobi (HJ) part, governing $\phi(x, y, t)$:

$$HJ: \begin{cases} \frac{1}{V}\phi_t - \|\nabla\phi\| = -c, & \forall (x,y) \in \Omega, \quad (2.29a) \\ \phi(x, y, t) = \phi_{CBD}, & \forall (x,y) \in \Gamma_d, \quad (2.29b) \\ \phi(x, y, t_{end}) = \phi_0(x, y), & \forall (x,y) \in \Omega. \quad (2.29c) \end{cases}$$

The authors interpreted ϕ_{CBD} as the cost of entering to the CBD and $\phi_0(x, y)$ as the instantaneous travel cost from any point at time $t = t_{end}$ to the CBD, when there is no traffic in the city and the travel cost is only related to the travel time, computed by an Eikonal equation as:

$$\begin{cases} \|\nabla\phi_0\| = c(x, y, t_{end}), & \forall (x,y). \in \Omega \quad (2.30a) \\ \phi_0 = \phi_{CBD}, & \forall (x,y) \in \Gamma_d. \quad (2.30b) \end{cases}$$

Notice that the initial time in the CL portion is $t = 0$, whereas in the HJ portion the initial time is $t = t_{end}$. **Du et al. (2015)** revisits the model proposed in Hoogendoorn and Bovy (2004) for vehicular traffic for the sake of simplifying the numerical solution. In this study, the authors have reduced the HJB equation to a HJ equation by removing the minimum value problem from the solution procedure and computing the actual travel cost when the velocity vector is known. Hence, the minimum value problem is only solved once in each iteration in the CL portion of the model. They

first reconsider the speed function as an anisotropic function as

$$U^i = U_f \, h\left(\psi^i(x,y,t)\right) \; g\left(\sum_i \rho^i\right), \quad \text{(anisotropic speed-density relationship)} \tag{2.31}$$

where U_f is the free-flow speed, ψ^i is the angle between the direction of movement of group i and the x-axis, $h(\psi^i)$ denotes the adjustment to the free-flow speed due to the direction of movement, and $g(\sum \rho^i)$ is a monotonically decreasing discomfort function. The local travel cost per unit distance for each group is defined as Eq. (2.27). The proposed model consists of two systems of PDEs: the CL portion and the HJ portion. The HJ portion of the model for each group is the same as Eq. (2.29) and the CL PDEs and their initial and boundary conditions are expressed as:

$$\begin{cases} \rho_t^i + \nabla \cdot \mathbf{f^i} = q^i, & \forall (x,y) \in \Omega, & \text{(2.32a)} \\ U^i = U_f \, h \, g, & \forall (x,y) \in \Omega, & \text{(2.32b)} \\ \psi^i = arg\min\limits_{\tilde{\psi}} p^i(x,y,t,\tilde{\psi}), & \forall (x,y) \in \Omega, & \text{(2.32c)} \\ \mathbf{f^i} \cdot \hat{\mathbf{n}} = 0, & \forall (x,y) \in \Gamma_o \cup (\cup_{m \neq i}) \Gamma_d^m, & \text{(2.32d)} \\ \rho^i(x,y,0) = \rho_0, & \forall (x,y) \in \Omega, & \text{(2.32e)} \end{cases}$$

where the CBDs other than the destination CBD of group i are viewed as obstructions for the travelers of group i and p^i is an auxiliary function defined as

$$p^i(x,y,t,\tilde{\psi}) \equiv \nabla \phi^i(x,y,t) \cdot \tilde{\mathbf{u}}^{\mathbf{i}}(x,y,t,\tilde{\psi}) + \tilde{c}^i(\tilde{\psi}) \tilde{U}^i(\tilde{\psi}). \tag{2.33}$$

Lin et al. (2017) extends the model proposed in Du et al. (2013) to a polycentric urban city, where the speed and local cost per unit distance are computed separately for each group of vehicles based on Eqs. (2.8) and (2.27), respectively, using the total density instead of the individual density of each group.

2.3.3 Combination of Reactive and Predictive Route Choice

We observed that the reactive and predictive route choice behaviors have been combined in some papers in order to obtain more realistic models. **Xia et al. (2009)** proposes a pedestrian DUE model founded on the hypothesis that "the pedestrians seek to minimize their estimated travel cost based on memory but temper this behavior to avoid high densities". In this hypothesis, the memory effect is the predictive or global portion and the behavior to avoid higher densities is the reactive or local portion of the route choice strategy. The route choice strategy is stated as

$$\mathbf{f}(x, y, t)// - \nabla\phi(x, y) - \omega\nabla c\left(\rho\right), \tag{2.34}$$

where $\phi(x, y)$ is the time-independent minimum travel cost based on memory, $c(\rho)$ denotes the density-dependent costs per unit distance of movement at time t, and ω is a positive constant representing the psychological influence. The function $\phi(x, y)$ is given by an Eikonal equation and the function $c(\rho)$ is given by

$$c(\rho) = \frac{1}{V(\rho)} + \beta\, g(\rho), \tag{2.35}$$

where $V(\rho) = U_f\left(1 - \frac{\rho}{\rho_{max}}\right)$ is the walking speed, $g(\rho) = \rho^2$ is the discomfort function, β reflects the sensitivity of the pedestrians' route choice to discomfort, and ρ_{max} is the jam density. **Hoogendoorn et al. (2015)** suggests another model for combining the global and local route choice behaviors, where the authors consider multiple groups of pedestrians with different destinations. The composite cost function for each group, i, is defined as

$$\omega_i = \phi_i + \varphi_i, \tag{2.36}$$

where ϕ_i is the global cost function representing the minimum cost of getting to the destination, Γ_d^i, from (x, y, t), and φ_i is the local cost function reflecting additional walking costs due to local (unforeseen) fluctuations in the density, which is composed of crowdedness and delay components, as

$$\varphi_i = \varphi_i^{crowdedness} + \alpha_i \varphi_i^{delay}, \tag{2.37}$$

where α_i is the weight representing the relative importance of the two factors for each pedestrian group. The route choice strategy gives the direction of movement as

$$\mathbf{f_i}// - \nabla\omega_i = -\nabla\phi_i - \nabla\varphi_i. \tag{2.38}$$

The crowdedness term in local cost function describes the tendency of the pedestrians to avoid areas with higher densities and the delay term reflects the expected increase in delay caused by local densities. The authors also did some simulation experiments in order to investigate the impact of different factors introduced in their new model.

2.3.4 Higher-order Macroscopic Models

The first-order models that have been discussed earlier are based on the assumption that the traffic flow is always in the equilibrium state and hence, they are not capable of describing non-equilibrium phenomena such as stop-and-go waves and formation of congestions. **Jiang et al. (2010)** introduces a second-order pedestrian flow model incorporating an equilibrium of linear momentum equation in addition to the conservation of mass Eq. (2.6), as:

$$\mathbf{u_t} + (\mathbf{u}\cdot\nabla)\mathbf{u} = \frac{V(\rho)\,\hat{d} - \mathbf{u}}{\tau} - R'(\rho)\,\frac{\nabla\rho}{\rho}, \quad \text{(second-order conservation law)} \tag{2.39}$$

where $\mathbf{u}(x, y, t)$ is the average pedestrian velocity vector, $\hat{d} = (d_1, d_2)$ denotes the unit optimal movement direction, τ is the relaxation time of $\mathbf{u}$ toward the optimal velocity (taken as 0.5 s by the authors), $R(\rho)$ denotes the traffic pressure, which describes the response of pedestrians to compression, and $R'(\rho) = c_s^2(\rho)$, where c_s is the sonic speed at which small disturbances propagate in a crowd. The first term at the right side of Eq. (2.39) is called the relaxation term, which represents the relaxation to equilibrium, and the second term is called the anticipation term, which reflects the reaction to the surrounding traffic condition. Equation (2.39) can be rewritten more explicitly in the components form as

$$\begin{cases} \dfrac{\partial u_1}{\partial t} + u_1 \dfrac{\partial u_1}{\partial x} + u_2 \dfrac{\partial u_1}{\partial y} = \dfrac{V\, d_1 - u_1}{\tau} - \dfrac{c_s^2}{\rho} \dfrac{\partial \rho}{\partial x}, & \text{(2.40a)} \\ \dfrac{\partial u_2}{\partial t} + u_1 \dfrac{\partial u_2}{\partial x} + u_2 \dfrac{\partial u_2}{\partial y} = \dfrac{V\, d_2 - u_2}{\tau} - \dfrac{c_s^2}{\rho} \dfrac{\partial \rho}{\partial y}. & \text{(2.40b)} \end{cases}$$

As usual, the velocity vector is assumed tangential to the gradient of $-\phi$, i.e. satisfying the Eikonal PDE (2.14). The final RDUE model proposed in Jiang et al. (2010) is a system of PDEs comprised of the CL Eq. (2.6), the equilibrium of linear momentum Eq. (2.39), and the Eikonal Eq. (2.14). Using numerical experiments, the authors found that the traffic becomes more unstable as the anticipation factor decreases, which is explained more explicitly later in Jiang et al. (2016a).

Jiang et al. (2016b) develops a second-order PDUE pedestrian flow model using Eqs. (2.6) and (2.39) for the CL part and Eq. (2.29) for the HJ part. They model is validated using experimental pedestrian flow data collected under non-congested conditions. For congested pedestrian flow conditions, using numerical experiments, this study found that increasing the anticipation factor results in significant reduction of the density near bottlenecks.

Jiang et al. (2016a) compares the RDUE and PDUE models developed by Jiang et al. (2010) and Jiang et al. (2016b), respectively, and evaluates their behavior by

numerical experiments. In order to conduct these numerical experiments, the authors define the equivalent traffic sonic speed as

$$c_s(\rho) = \sigma\rho, \tag{2.41}$$

where σ is the anticipation degree representing the anticipation behavior of pedestrians to compression. The RDUE and PDUE models are tested using different values for the anticipation degree and they found that if the pedestrians have strong anticipation consciousness, i.e. $\sigma > 0.33$, traffic flow is always in a stable state, less congestion occurs, and there are no significant difference between two models. On the contrary, if the anticipation degree falls below a critical value, e.g. $\sigma < 0.33$, traffic instability occurs, density increases at certain points resulting in congestion, and the RDUE model exhibits more uniform density distribution compared to the PDUE model.

2.3.5 Dynamic System Optimum

In dynamic system optimum (DSO) assignment problems the goal is minimizing the total travel cost for all of the travelers in the system. Solving DSO problem using macroscopic fundamental diagram is yet an underdeveloped area, with **Tao et al. (2014)** being the only DSO model using the continuum space for vehicular traffic. This model is based on the assumption that the intelligent transportation system (ITS) and the advanced traveler information system (ATIS) have perfect information about the time-varying traffic status and the travelers choose and change their route, if needed, entirely on the information given by the ITS and ATIS. The local cost per unit distance is defined as

$$c(\rho) = \frac{\kappa}{V(\rho)}, \tag{2.42}$$

where κ is the value of time and $V(\rho(x,y,t)) \equiv U_f(x,y)e^{-\beta\rho}$. The system cost is expressed as:

$$\Phi(\rho,\theta) = \kappa \iiint_{\Omega\times T} \rho \; dx\, dy\, dt, \qquad \text{(system cost)} \tag{2.43}$$

where $\theta(x,y,t)$ is the directed angle of speed, representing travelers' route choice. The DSO model is formulated as an optimization problem with a feasible region in the function space as:

$$min \;\; \Phi(\rho,\theta), \tag{2.44}$$

subject to the constraints

$$\begin{cases} \rho_t + \nabla\cdot\mathbf{f} = q, & \forall(x,y)\in\Omega, & \text{(2.45a)} \\ V = \|\mathbf{u}\|, & \forall(x,y)\in\Omega, & \text{(2.45b)} \\ \rho(x,y,0) = \rho(x,y,t_{end}) = 0, & \forall(x,y)\in\Omega, & \text{(2.45c)} \\ \rho(x,y,t) = 0, & \forall(x,y)\in\Gamma_o\cup\Gamma_h. & \text{(2.45d)} \end{cases}$$

Since the analytic properties of the proposed model were not studied by the authors, they were not able to develop a global optimal solution for the model. Instead, they derive a locally optimal solution with low computation cost.

2.3.6 CTM-based Models

The aforedescribed continuum-space models are solved numerically using standard PDE solution methods for discretization of continuum space and time. **Hänseler et al. (2014)** argues that the numerical discretization schemes of continuum space are relatively costly, which make those models infeasible for large-scale applications. The authors propose a discrete-time discrete-space pedestrian flow model, named

PedCTM, which utilizes a discretization scheme similar to Daganzo's Cell Transmission Model (CTM; Daganzo, 1994a; Daganzo, 1995). We have included this model in the continuum-space models category since PedCTM can be viewed as a numerical solution method of a conservation law in 2-D, similar to the CTM for one spatial dimension.

The period of analysis is discretized into intervals, τ, with uniform length Δt. The 2-D walking domain is partitioned into cells, $\varkappa$, with uniform shape, which are assumed to be partitioned orthogonally with spacing size of ΔL. The pedestrians are assumed to be distributed homogeneously within a cell and their movements are not modeled explicitly.

A pedestrian group, i, is defined by a route η_i, a departure time interval τ_i, and a size X_i. A route can be defined as a sequence of areas, $\eta = (r_o, r_1, ..., r_d)$, where each area r is a set of cells and r_o and r_d denote the origin and destination areas, respectively. Inside each route, several sequences of cells can connect the origin area to the destination area, where each sequence of cells is called a path and the en-route path choice is made by computing turning proportions at each cell.

In order to describe the internal dynamics of the model, a normalized PedCTM model is developed, which is independent of the absolute values of free-flow speed and jam density. This model utilizes an isotropic normalized cell-based version of the speed-density relationship proposed by Weidmann (1993) as

$$\upsilon(\bar{\rho}_{\varkappa,\tau}) = 1 - \exp\left(-\omega(\frac{1}{\bar{\rho}_{\varkappa,\tau}} - 1)\right), \quad \text{(normalized speed-density relationship)} \quad (2.46)$$

where υ is the normalized prevailing walking speed function, $\bar{\rho}_{\varkappa,\tau}$ represents the normalized density of cell $\varkappa$ at time interval τ and ω is a dimensionless parameter. The route choice behavior of pedestrians is described using route-specific cellular potential

fields $P^{\eta}_{\varkappa,\tau}$, given by

$$P^{\eta}_{\varkappa,\tau} = \alpha F^{\eta}_{\varkappa} - \beta D_{\varkappa,\tau}, \qquad \text{(potential field)} \tag{2.47}$$

where $F^{\eta}_{\varkappa}$ is the static floor field, which is assumed to be the minimum distance (number of cells) to be traversed from cell $\varkappa$ toward the destination area along route η, $D_{\varkappa,\tau}$ represents the dynamic floor field, which is taken as $\upsilon(k_{\varkappa,\tau})$ and α and β are positive weights. If the set of all adjacent cells to $\varkappa$ that are part of route η is denoted by $\Theta^{\eta}_{\varkappa}$, the turning proportion corresponding to link $\varkappa \to \chi \in \Theta^{\eta}_{\varkappa}$ is expressed as

$$\delta^{\eta}_{\varkappa\to\chi,\tau} = \frac{\exp\left(-P^{\eta}_{\chi,\tau}\right)}{\sum_{\chi'\in\Theta^{\eta}_{\varkappa}} \exp\left(-P^{\eta}_{\chi',\tau}\right)}. \qquad \text{(turning proportion)} \tag{2.48}$$

In this manner, the route-specific potential field can be interpreted as disutility of each cell in the route toward the destination area and the pedestrians try to decrease their disutility along the route in a reactive manner. The results of two real case studies reveal a realistic behavior of the framework under different conditions and a slightly better performance compared to social force model (SFM; Helbing and Molnár, 1995).

Hänseler et al. (2017) proposes an anisotropic discrete-time discrete-space pedestrian flow model based on a stream-based pedestrian fundamental diagram (SbFD). Unlike Hänseler et al. (2014) where areas were comprised of cells, in Hänseler et al. (2017), each area, ζ, contains a number of streams, s, and the set of streams associated with area ζ is denoted by Λ_{ζ}. Each route, η, consists of a pair of origin and destination nodes and a set of streams ,Λ_{η}, connecting them. The walking speed of stream s in area ζ is computed by

$$U_s = U_f \exp\left(-\vartheta(\frac{N_{\zeta}}{A_{\zeta}})^2\right) \prod_{s'\in\Lambda_{\zeta}} \exp\left(-\beta(1-\cos\theta_{s,s'})\frac{M_{s'}}{A_{\zeta}}\right), \tag{2.49}$$

where N_ζ and A_ζ are the accumulation and surface size of area ζ, $M_{s'}$ denotes the number of pedestrians in stream λ', $\theta_{s,s'}$ is the intersection angle between streams s and s', and ϑ and β are the model parameters. At each node Υ, where the streams intersect, a potential value $P^{\eta}_{\Upsilon^{d}_{s},\tau}$ is defined as the remaining walking time to destination d using stream s in route η that can be calculated using any shortest path algorithm, e.g., Dijkstra (1959). A path choice strategy similar to Hänseler et al. (2014) is proposed, which incorporates a weighted logit-type model at each node.

The performance of the model is compared to a few isotropic specifications (Wong et al., 2010; Weidmann, 1993) at the example of two case studies. The analysis reveals that the consideration of anisotropy improves the accuracy of the proposed model compared to the tested isotropic specifications. Overall, the model proposed in Hänseler et al. (2017) can be seen as an analogy for the PedCTM model (Hänseler et al., 2014), where the cell-based fundamental diagram, potential fields and path choice are replaced by stream-based ones and anisotropy is also taken into account.

2.3.7 One-dimensional Spatial Model

Unlike previous models, which are in two spatial dimensions, **Laval et al. (2017)** incorporates a single spatial dimension to describe the changes in the area covered by the MFD. This study investigates DUE conditions on a single origin-destination pair with two alternative routes, a freeway with a fixed capacity μ_0 and the surrounding city-streets (CS) network, described by an o-MFD with capacity μ_1. The authors first describe the single MFD dynamics and find analytical solution for special cases. Subsequently, they add a freeway alternative under two scenarios for the CS network: (i) constant network length, and (ii) variable network length to account for the spatial extent of congestion. In the first scenario, the combined system dynamics are given

by the following ODE

$$\begin{cases} k'(t) = \dfrac{b\varpi(t) - kv}{1 - bmv'/v^2}, & \text{(reservoir dynamics)}, \quad (2.50a) \\ k(0) = k_0, & \text{(initial conditions)}, \quad (2.50b) \end{cases}$$

where k is the occupancy, t is measured in units of free-flow travel time inside the MFD, τ^*, b is a positive constant defining the shape of the MFD, and $v(k)$ is a dimensionless version of the speed-occupancy MFD. The dimensionless parameters m and ϖ are given by

$$m = \frac{\mu_0}{\mu_1}, \quad \text{(Freeway to CS capacity ratio)}, \quad (2.51a)$$

$$\varpi(t) \equiv \frac{\lambda(t) - \mu_0}{\mu_1} = \frac{\lambda(t)}{\mu_1} - m, \quad \text{(MFD demand intensity)}. \quad (2.51b)$$

The authors were able to solve the ODE (2.50) analytically for the autonomous case, where the demand is constant over time and the evolution of the system depends only on the occupancy. For scenario (ii) the authors incorporate continuum approximation (CA), proposed by Laval (2009), for off-ramps into their proposed formulation, as seen in Fig. 2.5. The discrete off-ramps are considered as a continuum, where vehicles can exit the freeway to CS at any location upstream the bottleneck along $0 \leq x \leq \xi(t)$, where $x = 0$ is the bottleneck location and $\xi(t)$ is a characteristic of DUE solution, named information wave, marking the most upstream location where vehicles divert from the freeway to the CS.

The implementation of CA in the proposed model makes it able to cope with time-varying network and trip lengths for the MFD. Notwithstanding the fact that analytical solutions become impossible for variable network and trip lengths, the numerical solutions strongly suggest that the proposed system have many analogies to the constant length model. The main difference is that gridlock does not happen in

the proposed model and the steady-state solution is independent of surface network parameters, when time is expressed in units of MFD free-flow travel time.

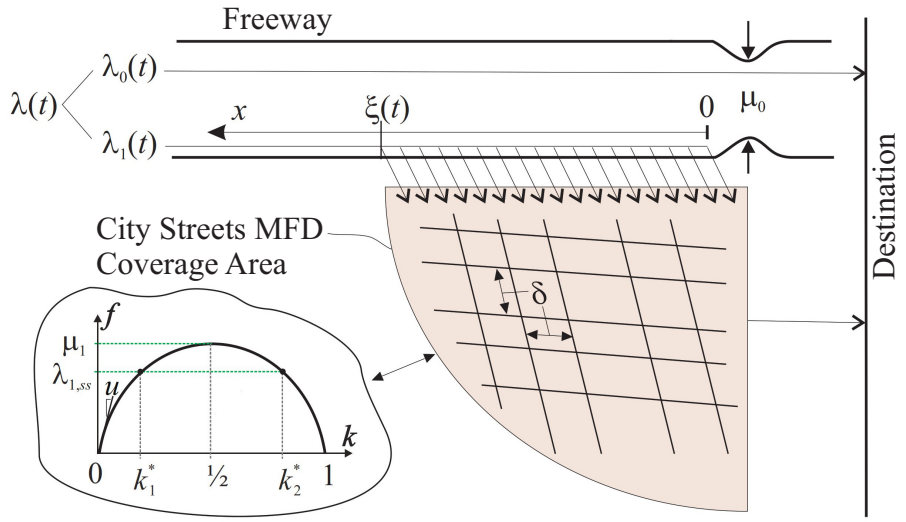

Figure 2.5: Freeway vs. City streets network with continuum approximation for off-ramps (Laval et al., 2017).

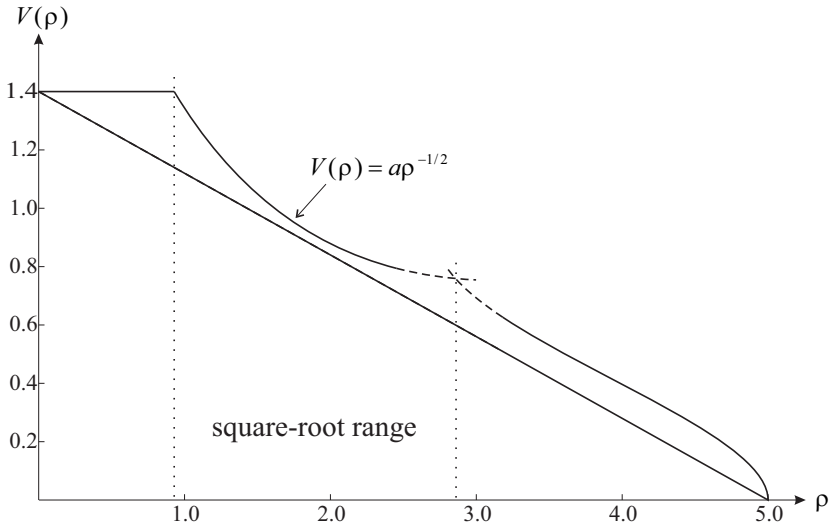

Figure 2.6: Square-root range in speed-density relationship (Hughes, 2002).

2.4 Solution Methods for Continuum-Space models

2.4.1 Analytical Solutions for Hughes' Model

Hughes (2002) shows that when the speed-density relation is proportional to $\rho^{-1/2}$, i.e.

$$V(\rho) = a\rho^{-1/2}, \tag{2.52}$$

where a is a positive constant, then his formulation has a symmetry under conformal mappings, i.e. Eq. (2.6) looks identical before and after the transformation. These maps make it possible to obtain analytical solutions to problems with obstructions of complicated shape. There are map dictionaries where one can identify which particular mapping will transform the shape of the obstructions to lines or circles, for which the solution is known.

Since the square root approximation, Eq. (2.52), is not a good representation of speed for all densities, Hughes assumes that it holds inside a density range, which we can call the "square-root" range; see Fig. 2.6. Unfortunately, when the density falls outside of this range, the symmetry is lost and the solution method becomes convoluted.

Later in this chapter, we will show that the conformal mapping symmetry is valid for all density ranges and for any speed-density relationship. In fact, Eq. (2.52) is needed only to simplify Hughes' model to Laplace's equation,

$$\nabla^2\phi = 0, \qquad \text{(Laplace's equation)} \tag{2.53}$$

whose solution is well understood. To see this, notice that the conservation law Eq. (2.13a) can be expressed as

$$\rho_t - \left(2\rho V' \nabla\rho \cdot \nabla\phi + (\rho\nabla^2\phi + \nabla\rho \cdot \nabla\phi)V\right) V = 0, \tag{2.54}$$

by expanding the divergence term and assuming the discomfort function $g(\rho) = 1$ without loss of generality. The reader can verify that the two terms of involving $\nabla\rho \cdot \nabla\phi$ cancel out under Eq. (2.52) and since $\rho_t = 0$ in steady-state, we end up with Laplace equation as sought.

Hughes also identifies two limit cases and proposes simpler equations that the solution satisfies in steady state:

1. at low free-flow densities $\|\nabla\rho\| >> \|\nabla\phi\|$, therefore the CL Eq. (2.13a) becomes

$$\nabla\rho \cdot \nabla\phi \approx 0 \tag{2.55}$$

   which means that lines of constant density are almost perpendicular to lines of constant potential.

2. at high congested densities $\|\nabla\rho\| << \|\nabla\phi\|$, thus the CL becomes

$$\nabla^2\phi \approx 0 \tag{2.56}$$

Although not shown in the original study, we can obtain these results using Eq. (2.54) and letting $\rho_t = 0$ and $\nabla^2\phi = 0$ for the first case, and $\|\nabla\rho\| \approx 0$ for the second case.

2.4.2 Numerical Solutions

In order to solve the continuum-space models, various numerical solutions are proposed in the literature. Most of the solution procedures use a combination of standard solution methods for PDEs in order to discretize the continuous space and time. Conservation law and Hamilton-Jacobi equations are often solved by methods such as finite volume methods (FVM), finite difference methods (FDM), including Lax-Friedrichs (LF) and weighted essentially non-oscillatory (WENO) schemes, and finite element methods (FEM), including discontinuous Galerkin method (DGM).

Fast sweeping method (FSM) and fast marching method (FMM) are mostly utilized to solve the Eikonal equation. Most of the studies use total variation diminishing Runge-Kutta (TVDRK) for time discretization and/or integration. Some of the PDUE studies have also viewed the model as a fixed-point problem and have solved the CL and HJ parts simultaneously using self-adaptive method of successive averages (SA-MSA). Table 2.1 presents the different solution methods used in each study.

2.5 Discussion and Outlook

In this section, we discuss our main findings in terms of the observed trends in the literature, and identify issues that deserve further research. Table 2.1 presents an overview of the studied papers and their specifications. The lines between some rows delimit the studies reviewed in different sections and subsections of this chapter, where the bold-faced characteristic at each row specify the common feature of the models in the pertinent subsection.

2.5.1 The Missing Connection Between Continuum-space and Discrete-space Models

As mentioned earlier in the third section, a few studies in the continuum-space DTA literature have briefly mentioned MFD as a justification for the speed-density rela-

tionship traditionally used within this modeling approach (Du et al., 2013; Du et al., 2015; Long et al., 2017). However, the speed-density relationships used in most of the studies are similar to Eq. (2.8), which might be convenient for mathematical purposes but is unrealistic (infinite jam density) and does not explicitly incorporate the effects of network topology and signal timing.

Nevertheless, we agree that the speed-density relationship should be the MFD (as opposed to a traditional fundamental diagram, FD), mainly because in a well-defined continuum-space DTA model, the speed-density relationship has to incorporate the influence of intersections. To illustrate the point more clearly, Fig. 2.7 demonstrates a simple case where there is a signalized intersection downstream of a link. Fig. 2.7(a) shows the uninterrupted link FD (without the intersection), where u is the free-flow speed, w is the wave speed and the blue line indicates the intersection capacity, which is a function of the ratio of green time over the cycle length. Combining the effect of intersection with the supposed uninterrupted FD results in Fig. 2.7(b), the MFD of the link upstream of the intersection. We think that it is preferable to call the

Table 2.1: Overview of the macroscopic DTA (discrete-/continuum-space) studies

Source	Granularity	Pedestrian or Vehicular	Route Choice	Multi-Destination	Directionality	Solution Methods*	Main Contribution
Yildirimoglu and Geroliminis (2014)	**Discrete**	Vehicular	Predictive	Yes	Isotropic	SNL, MSA, Logit	Divides heterogeneous urban network into homogeneous regions and defines the route as the sequence of regions
Yildirimoglu et al. (2015)	**Discrete**	Vehicular	Predictive & Marginal	Yes	Isotropic	MSA	Divides regions into sub-regions and redefines the route-choice as the sequence of sub-regions
Haddad et al. (2013)	**Discrete** & Freeway	Vehicular	Reactive & DSO	Yes	Isotropic	MPC	Shows that integrating DSO within centralized control scheme increases network performance compared to RDUE
Sossoe and Lebacque (2017)	**Discrete**	Vehicular	Reactive	Yes	Bi-directional	Logit	Decomposes the urban network into zones, which are meshed into cells, and computes the turning proportion in each cell using logit
Hughes (2002)	Continuum	Pedestrian	—	Yes	**Isotropic**	Conformal maps	Develops the framework for continuum modeling of DTA
Huang et al. (2009b)	Continuum	Pedestrian	**Reactive**	No	**Isotropic**	WENO, FSM, TVDRK	Proves that Hughes (2002) satisfies the RDUE conditions
Jiang and Zhang (2011)	Continuum	Pedestrian	**Reactive**	No	**Isotropic**	DGM, TVDRK	Introduces a new route-choice strategy
Jiang and Zhou (2014)	Continuum	Vehicular	**Reactive**	No	**Isotropic**	FVM, FSM, TVDRK	Extends the concept of RDUE to the vehicular traffic
Jiang et al. (2009)	Continuum	Pedestrian	**Reactive**	Yes, 2	**Bi-directional**	FVM, FMM, TVDRK	Extends the RDUE concept to bi-directional pedestrian flow
Huang et al. (2009a)	Continuum	Pedestrian	**Reactive**	Yes, 2	**Bi-directional**	FDM, LF	Extends the model developed in Jiang et al. (2009) by introducing look-ahead behavior
Xiong et al. (2011)	Continuum	Pedestrian	**Reactive**	Yes, 2	**Bi-directional**	WENO, LF, FSM, TVDRK	Proposes a different numerical solution procedure for the model developed in Jiang et al. (2009)
Jiang et al. (2012)	Continuum	Pedestrian	**Reactive**	Yes, 2	**Bi-directional**	FVM, FSM, TVDRK	Revisits the model developed in Jiang et al. (2009) and introduces a more general local travel cost
Hoogendoorn and Bovy (2004)	Continuum	Pedestrian	**Predictive**	Yes	Isotropic	FDM	Develops the first pedestrian PDUE model comprised of CL and HJB parts
Jiang et al. (2011)	Continuum	Vehicular	**Predictive**	No	Isotropic	FVM, FEM, LF, TVDRK	Develops the first continuum-space PDUE model for vehicular traffic
Du et al. (2013)	Continuum	Vehicular	**Predictive**	No	Isotropic	LF, FSM/ SA-MSA	Revisits the route-choice strategy proposed in Jiang et al. (2011)
Du et al. (2015)	Continuum	Vehicular	**Predictive**	Yes	Anisotropic	LF/ SA-MSA	Revisits the model developed in Hoogendoorn and Bovy (2004) for vehicular traffic and reduces the HJB to HJ
Lin et al. (2017)	Continuum	Vehicular	**Predictive**	Yes	Isotropic	FVM/ SA-MSA	Extends the model developed in Du et al. (2013) to polycentric urban city
Xia et al. (2009)	Continuum	Pedestrian	**Predictive & Reactive**	No	Isotropic	DGM, FSM, TVDRK	Proposes a route-choice strategy including the global and local route-choice behaviors
Hoogendoorn et al. (2015)	Continuum	Pedestrian	**Predictive & Reactive**	Yes	Isotropic	FDM	Proposes a more detailed local route-choice formulation compared to Xia et al. (2009)
Jiang et al. (2010)	Continuum	Pedestrian	Reactive	No	Isotropic	FVM, FSM, TVDRK	Develops a **second-order** RDUE model incorporating equilibrium of linear momentum
Jiang et al. (2016b)	Continuum	Pedestrian	Predictive	No	Isotropic	FVM, FSM/ SA-MSA	Develops a **second-order** PDUE model similar to Jiang et al. (2010)
Jiang et al. (2016a)	Continuum	Pedestrian	Predictive & Reactive	No	Isotropic	FVM, FSM/ SA-MSA	Compares the **second-order** RDUE (Jiang et al., 2010) and PDUE (Jiang et al., 2016b) models
Tao et al. (2014)	Continuum	Vehicular	**DSO**	No	Isotropic	FVM	Finds the DSO solution instead of the DUE solution
Hänseler et al. (2014)	Cells	Pedestrian	Reactive	Yes	Isotropic	Logit	Develops PedCTM model and incorporates **cell-based** logit route-choice
Hänseler et al. (2017)	Areas	Pedestrian	Reactive	Yes	Anisotropic	Logit	Replaces the cell-based route-choice in Hänseler et al. (2014) to **link-based** route-choice using logit
Laval et al. (2017)	**Continuum & Freeway**	Vehicular	—	Single O-D	Isotropic	**Analytical Solution**	Investigates the DUE conditions for CS network with an alternative Freeway

* **DGM**: Discontinuous Galerkin Method, **FDM**: Finite Difference Method, **FEM**: Finite Element Method, **FMM**: Fast Marching Method, **FSM**: Fast Sweeping Method, **FVM**: Finite Volume Method, **LF**: Lax-Friedrichs method, **MPC**: Model Predictive Control, **MSA**: Method of Successive Averages, **SA-MSA**: Self-Adaptive Method of Successive Averages, **SNL**: Stochastic Network Loading, **TVDRK**: Total Variation Diminishing Runge-Kutta, **WENO**: Weighted Essentially Non-Oscillatory

resultant diagram MFD, aka Network Fundamental Diagram, since the effects of network are also included in it, although this interpretation may not fit the common belief regarding MFD that it has to be defined over a large network. The same thing can be done at the corridor level, and also at the network level. This is equivalent to the well-known method of cuts used to obtain MFD (Daganzo and Geroliminis, 2008).

Much in the same way as in 1-D continuum traffic flow theory, where the discrete vehicle trajectories are "liquefied" in its vicinity, in 2-D continuum models it is the network, in particular the effects of intersections, that are spread out within a vicinity. In this regard, any equation describing the fundamental relationships of traffic flow in any portion of such dense network has to take into account the network effects, and thus it is a MFD and not a FD. Hence, a crucial direction for future research is the formulation of speed-density relationship in continuum-space DTA framework based on the assumptions and methods of MFD theory.

Regarding the connection with the MFD in discrete-space models, recall that in these models the spatial dimension is omitted after assuming that the network is more or less homogeneous in a given region. This means that if the spatial dimension were to be re-introduced, then at each (x, y) point in the region the continuum MFD should

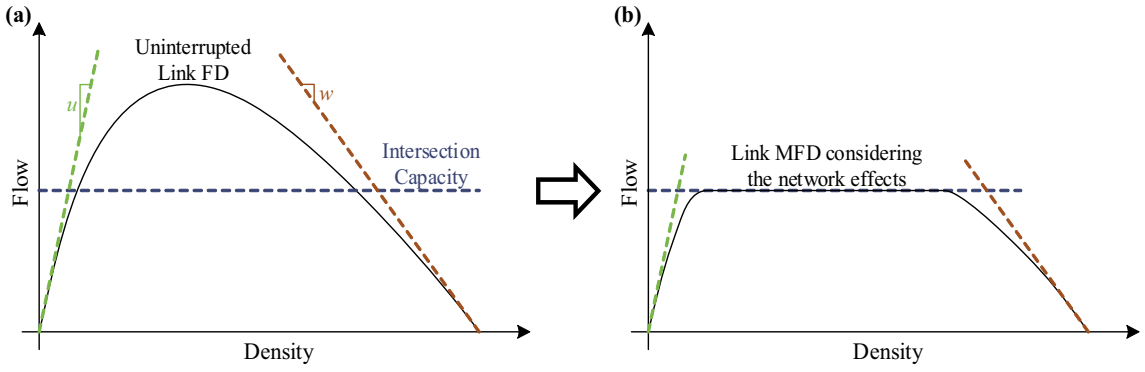

Figure 2.7: Simple case of obtaining MFD of a link upstream of a signalized intersection.

be a scaled version of the discrete-space MFD. This interpretation is consistent with the premise of a well-defined connection between discrete and continuum models: *a well-defined discrete-space model can be interpreted as the numerical solution of a continuum model; i.e., when the size of sub-regions tends to zero while keeping the total area constant, the discrete model tends to the continuum model.*

2.5.2 From MFD to the Conservation Law

No attempts can be found in the continuum-space literature to verify whether or not the assumptions of MFD theory are met. This is perhaps because in existing models, travelers do not exit within the region except at discrete destinations. But in the case of origins and destinations distributed randomly in the region, we can postulate the conservation law with source term, i.e. $\rho_t + \nabla \cdot \mathbf{f}(\rho) = (\lambda - o(\rho))/L$, where $\lambda(x, y, t)$ is the demand flow, which may or may not be constrained by Eq. (2.4) as discussed earlier[1]. In the isotropic case, we have that the flux vector satisfies $\|\mathbf{f}\| = V(\rho)\,\rho$, and since by definition (2) we have $o(\rho) = V(\rho)\,\rho\,\frac{L}{\ell}$, it follows that the corresponding conservation law becomes

$$\rho_t + \nabla \cdot \mathbf{f}(\rho) = \lambda/L - V(\rho)\,\rho/\ell. \tag{2.57}$$

Notice that the network length L and the trip length ℓ appear explicitly in this formulation, unlike existent literature. Research is needed to fully understand the solution of this equation in the context of the different cost functions proposed in the literature, and to generalize it for the anisotropic case and variable trip length.

[1]Notice that in this subsection the density ρ is defined per unit distance, and not per unit area as in the rest of this review.

2.5.3 Differences Between the Continuum-space Pedestrian Flow and Vehicular DTA Models

We mentioned at the outset that continuum-space pedestrian flow and vehicular DTA models are mathematically very similar, and therefore can be analyzed within the same framework. From a modeling perspective, however, there are some differences. A quick look at Table 2.1 reveals that most of the studies investigating pedestrian flow incorporate reactive route choice component, whereas the studies regarding vehicular flow subsume predictive route choice component. This seems rational since it means that pedestrians usually do not have complete information about the conditions toward their destination and seek to reduce their instantaneous travel cost by choosing the shortest path and also avoiding higher densities, which cause discomfort and excessive delay. On the other hand, drivers are increasingly having better information about the traffic conditions in different routes, using their own experience or traffic applications, and try to minimize their actual travel cost.

Another obvious difference between continuum-space pedestrian flow and vehicular DTA models is that in the pedestrian flow models the demand is exogenous, i.e. set to zero inside the continuum domain and set to q on outer boundary, while it is endogenous in vehicular DTA models, i.e. set to q inside the continuum domain and set to zero on outer boundary.

2.5.4 Departure Time Choice

Departure time choice modeling has not been incorporated in the literature. There is a clear need to concentrate in departure time choice within the macroscopic DTA framework, especially after the compelling results by Fosgerau (2015), which shows that under "regular sorting" (shorter trips depart later and arrive earlier compared to longer trips) the problem becomes simplified very significantly in the case of a single MFD, to the point that reservoir dynamics do not need to be computed explicitly.

Contrary to the results of Fosgerau (2015), Lamotte and Geroliminis (2017) shows that if there is only a single peak in the morning commute, a First-in, First-out (FIFO) sorting pattern emerges within early and late user families having identical $\alpha - \beta - \gamma$ scheduling preferences and heterogeneous trip lengths. The α, β and γ parameters are, respectively, the per-time-unit costs of travel/queuing, earliness, and lateness, as defined first in Vickrey (1969). Empirical measurements are needed to resolve these discrepancies.

2.5.5 System Optimum Assignment

Only a few studies, Haddad et al. (2013) and Yildirimoglu et al. (2015) in discrete space, and Tao et al. (2014) in continuum space have tried to establish the DSO conditions, but the proposed solution methods are numerical. Since it is hard to extract insights from numerical solution methods, future research should focus on analytical solutions, possibly of simplified systems.

2.5.6 Directionality

Most of the studies propose isotropic models and only a few investigate bi-directional or anisotropic conditions. Although solving anisotropic models seems to be more costly and sometimes unattainable, it is clear that more effort is needed on investigating anisotropic conditions and the interaction of intersecting flows.

One possibility could be estimating separate MFDs for each cardinal direction, with the method in Laval and Castrillón (2015) for example, and incorporate a model to manage the interactions between the flows in different directions.

2.5.7 Capacity Constraint

We saw that a standard assumption in the literature is Eq. (2.4), i.e. that in congestion, the inflow cannot exceed the outflow. We argue that this constraint can be

relaxed to account for more realistic operations. For example, (i) transient surges in demand could be allowed temporarily, and (ii) distance traveled within the reservoir may increase with congestion as travelers find longer routes. Research is needed to identify how the capacity constraint should be modified to accommodate these and other modeling improvements.

2.5.8 Solution Methods

Numerical solution of continuum-space models

For continuum-space models, we have seen that the proposed numerical solution methods correspond to standard methods in numerical PDEs. These methods converge to exact solutions only when the mesh size of the numerical grid tends to zero; otherwise, significant numerical viscosity can be introduced. Additionally, it is not clear from the literature how computation times compare with traditional DTA methods; it is possible that for acceptable accuracy, the computational times of continuum-space models might be even greater.

Further research is needed to extend the recent advances in *exact* numerical solution methods for one-dimensional kinematic wave (LWR) model (Lighthill and Whitham, 1955; Richards, 1956) to the two-dimensional models in order to develop more efficient and less diffusive numerical methods. These recent advances exploit (i) a shear symmetry property of conservation laws (Laval and Chilukuri, 2016), and (ii) the link between conservation laws and the Hamilton-Jacobi equation. In the one-dimensional case, the viscosity solution of the Hamilton–Jacobi equation, $\varpi_t + H(\varpi_x) = 0$, is the primitive of the unique entropy solution of the corresponding conservation law, $\rho_t + H(\rho)_x = 0$, where $\rho = \varpi_x$. However, in the multi-dimensional case, this one-to-one correspondence no longer exists, but the gradient $\nabla\varpi$ satisfies a system of conservation laws (see Kurganov et al., 2001). Of particular interest

would be a cellular automaton type solution method, which could potentially reduce computational times by orders of magnitude.

Analytical solution of continuum-space models

As mentioned in section 2.4.1, it turns out that the conformal mapping symmetry is valid for all density ranges and for any MFD. The implications are profound, as it implies that one can choose the conformal mapping that would simplify boundary conditions as much as possible, thereby simplifying the overall solution method. Research is needed to see if this symmetry is also valid for the other continuum models in the literature.

To show that the conformal mapping symmetry is valid for all density ranges and for any MFD, let the discomfort function $g(\rho) = 1$ in Hughes' model without loss of generality, which becomes

$$\begin{cases} \rho_t - \nabla \cdot \left(\rho V^2 \nabla \phi\right) = 0, & \text{(2.58a)} \\ ||\nabla \phi|| = 1/V, & \text{(2.58b)} \end{cases}$$

with appropriate boundary conditions. We now study Hughes (2002) equations under a general transformation and then conclude that conformal mappings are the only invariant ones. To see this, let $\mathbf{z} = (x, y)$ represent the original coordinate and let

$$\mathbf{Z}(x, y) = (X(x, y), Y(x, y)) \tag{2.59}$$

be a spatial transformation to the new coordinate system $\mathbf{Z} = (X, Y)$. The Jacobian of this transformation, $\mathbf{J} = \begin{pmatrix} \partial X/\partial x & \partial X/\partial y \\ \partial Y/\partial x & \partial Y/\partial y \end{pmatrix}$ can be written more concisely as $\mathbf{J} = \nabla_z \mathbf{Z}$ where we have introduced the coordinate system as a subscript to the gradient operators. It

can be shown that

$$\nabla_z \phi(X(x,y),Y(x,y),t) = \nabla_Z \phi(X,Y,t) \cdot \mathbf{J}, \tag{2.60a}$$

$$\nabla_z \cdot \mathbf{f}(X(x,y),Y(x,y),t) = \nabla_Z \cdot (\mathbf{J} \cdot \mathbf{f}(X,Y,t)). \tag{2.60b}$$

Replacing Eq. (2.60) in Eq. (2.58a) gives

$$\rho_t - \nabla_Z \cdot \left(\rho V^2 \mathbf{J} \cdot \mathbf{J}^T \cdot \nabla_Z \phi\right) = 0.$$

Since $\rho(x,y,t) = \rho(X,Y,t)|\mathbf{J}|$ and $V(\rho(x,y,t)) = V(\rho(X,Y,t))|\mathbf{J}|^{-1/2}$ (because transformed small areas equal the original area scaled by the determinant $|J|$), we have

$$|\mathbf{J}|\rho_t - \nabla_Z \cdot \left(\rho V^2 \mathbf{J} \cdot \mathbf{J}^T \cdot \nabla_Z \phi\right) = 0. \tag{2.61}$$

Conformal maps have the unique property that $\mathbf{J} \cdot \mathbf{J}^T = |\mathbf{J}| \cdot \mathbf{I}$, so that Eq. (2.61) becomes

$$\rho_t - \nabla_Z \cdot \left(\rho V^2 \nabla_Z \phi\right) = 0. \tag{2.62}$$

which reads exactly as Eq. (2.58a). Therefore, conformal mappings are the only ones where Hughes' model is invariant.

2.6 Chapter Summary

This chapter has presented a detailed and comprehensive literature review of dynamic traffic assignment (DTA) models that implement the macroscopic fundamental diagram (MFD). These models can be categorized into (i) discrete-space, and (ii) continuum-space models depending on the treatment of the spatial dimensions.

Discrete-space DTA models have been recently given attention by researchers and the literature in this area is yet underdeveloped. In these models, a MFD at each subregion determines the speed of all vehicles present in that subregion. One potential problem of these models can be encountered in highly-heterogeneous networks, where the homogeneous subregions are relatively small. Under excessive demand, if the accumulation of a subregion reaches its maximum accumulation, the average speed of vehicles over the subregion will become zero. Since the speed of all vehicles in a sub-region is assumed to equal to the average speed computed by MFD, no vehicles can move out of the subregion and the subregion will get clogged forever. Ongoing research by the authors is trying to address this drawback.

On the other hand, the literature on continuum-space vehicular DTA and pedestrian flow models, which dates back to 2002, is more developed in various aspects. These models are usually solved numerically using standard PDE solution methods, that discretize the space and time. One of the significant gaps in the literature is the missing link between discrete-space and continuum-space models, which we have tried to elucidate in this review. Although it seems to be vague in the literature, we have tried to clarify why some of the recent studies in the literature have interpreted the speed-density relationship used in these models as the MFD in subsection 2.5.1.

Based on the discussion in subsection 2.5.1, the numerical solution of continuum-space models can be regarded as a discrete-space model. Unfortunately, we have seen that the numerical solution methods used in the literature correspond to standard methods in numerical PDEs, which prevents an intuitive interpretation of the resulting discrete model. The numerical solution methods can become costly and intractable for large-scale and more general applications that are not observed in the literature. It seems promising to extend the latest solution methods proposed for the one-dimensional traffic flow models based on Godunov's method (such as the cell transmission model), in which case the interpretation is very intuitive and therefore

could be worked with independently of the continuum model. A step in this direction is the models in Hänseler et al. (2014) and Hänseler et al. (2017), which have recently made some effort to using CTM-based space discretization schemes.

Other major shortcomings of the literature have been presented in subsection 2.5 and recommendations for future research are also highlighted based on these drawbacks.

CHAPTER 3
CONTINUUM-SPACE DYNAMIC TRAFFIC ASSIGNMENT FRAMEWORK CONSISTENT WITH THE MFD THEORY

This chapter is modified from the following published paper:

Rafegh Aghamohammadi and Jorge A. Laval (2020a). "A continuum model for cities based on the macroscopic fundamental diagram: A semi-Lagrangian solution method". In: *Transportation Research Part B: Methodological* 132, pp. 101–116

3.1 Introduction

The previous chapter has presented an extensive literature review on the dynamic traffic assignment (DTA) models using an aggregated representation of the network and pointed out the current gaps and shortcomings in the literature. These models are categorized into 2 approaches: **(i) the discrete-space approach**, where the network is partitioned into a number of zones and the traffic dynamics inside each zone is given by its MFD, and **(ii) continuum-space approach**, where the network is too dense such that it can be approximated as a continuum space that the users can circulate at any point $\mathbf{x} = (x, y)$ of the Euclidean, two-dimensional domain $\Omega \subset \mathbb{R}^2$ and the traffic dynamics is given by a 2-dimensional conservation law Eq. (2.6).

Nevertheless, the connection between these two modeling approaches is not well understood. To the best of our knowledge, there are only three references in the continuum literature that briefly mention the MFD to justify the use of a speed-density relationship, $V(\rho)$, in the conservation law (Du et al., 2013; Du et al., 2015; Long et al., 2017), and no references in the MFD literature citing continuum models. Chapter 2 delves into this gap in the literature and argues that the speed-density relationships in the continuum models need to consider the network effects and hence,

it is valid to interpret them as MFD. However, it is still not clear how the flux function for the conservation law $Q(\rho) = \rho V(\rho)$ and $\bar{Q}(n)$ are related, or how the average trip length affects the continuum model. It is the purpose of this chapter to fill this gap by reformulating the theory such that both modeling approaches are *consistent*, in the sense that taking the limit of the reservoir size going to zero leads to the continuum model, and vice versa, the spatial integration of the continuum model leads to the reservoir model.

As pointed out in Chapter 2, the numerical solution methods that have been proposed to solve both the conservation law and route choice component of continuum models correspond to the standard numerical solution methods for PDEs. These methods converge to exact solutions only when the mesh size of the numerical grid tends to zero, which could be problematic if fast computations are required; otherwise, significant numerical viscosity can be introduced even in the case of a single commodity (i.e. one type of users, all going to the same destination). For multi-commodity problems, the formulation becomes cumbersome and the numerical solution methods introduce additional errors that can lead to first-in-first-out (FIFO) violations (Jin and Jayakrishnan, 2005).

We argue that these challenges in current continuum models arise because they were formulated and solved numerically in Eulerian coordinates, which are attached to the infrastructure. A more natural representation of this problem is achieved in Lagrangian coordinates which are attached to the vehicle particles. In this representation, origin-destination information (and any vehicle-specific characteristic) is propagated naturally in the traffic stream (Leclercq et al., 2007; Laval and Leclercq, 2013). However, in a fully Lagrangian representation, it is cumbersome to keep track of variables that are fixed in space such as demand inflow or changes in the infrastructure, or the flux transfers between neighboring reservoirs, in particular, which are fixed in space.

In this chapter, we will develop a semi-Lagrangian scheme that uses Eulerian and Lagrangian representations where it is most convenient. These schemes use an Eulerian computation grid where the motion of particles is tracked. Using these trajectories, at every time step we estimate the continuum variables at the grid points. This general idea has been applied copiously in the field of atmospheric sciences (Staniforth and Cote, 1991). In traffic flow, there have been efforts to combine these two coordinate systems in one spatial dimension for the kinematic wave model (Leclercq, 2007) using a microscopic car-following model for the road segments, and a macroscopic network junction model for merges. Here, we adapt Leclercq's (2007) framework for the two-dimensional case and modify the microscopic components in the spirit of the MFD; i.e. instead of a car-following model where each vehicle can have a different speed, we impose that all vehicles traveling from one reservoir to an adjacent one have the same speed. The calculation of this speed is not trivial. We found that the conventional approach to assign the MFD speed $V(\rho)$, as in Yildirimoglu and Geroliminis (2014), Yildirimoglu et al. (2015), and Leclercq et al. (2017a), might lead to stalling in the multi-reservoir setting. To avoid stalling, and in the spirit of the kinematic wave model, the speed is taken directly from the network junction model.

While the relevant literature contains no efforts to achieve a consistent formulation both in discrete and continuum space, the Eulerian components of our solution method are closely related to the framework proposed in Sossoe and Lebacque (2017), Hänseler et al. (2014), and Hänseler et al. (2017). Sossoe and Lebacque (2017) proposes a multi-reservoir reactive DTA model where the flows between reservoirs are obtained with a network junction model (see e.g., Jin, 2012; Jin and Zhang, 2003, and references therein). This model considers only two possible paths in each cell for the route choice based on a logit model, where the cost of each path is assumed to be the instantaneous travel time. There is no mention of the MFD or of the corresponding continuum model, however. Hänseler et al. (2014) develops a discrete-time discrete-

space pedestrian flow model, named PedCTM, which extends Daganzo's Cell Transmission Model (CTM; Daganzo, 1994a; Daganzo, 1995) to 2 dimensions. The model in Hänseler et al. (2017) is analogous to the PedCTM model, where the cell-based fundamental diagram, potential fields and path choice are replaced by stream-based ones and anisotropy is also taken into account.

3.2 Continuum-space MFD Formulation

In this section, we will formulate a continuum-space model consistent with the assumptions behind the MFD theory. We will show that this consistent formulation is capable of handling randomly distributed destinations over the network. Later, the properties of solutions to this model in both isotopic and anisotropic cases will be discussed. Consistency in this section means that the spatial integration of the continuum model should lead to the reservoir model. Notice that this formulation pertains only to the conservation law component of the model; the route choice component is arbitrary.

Our purpose here is to find the connection between the continuum-space conservation law Eq. (2.6) and the reservoir dynamics ODE Eq. (2.3). Toward this end, we first need to reformulate the conservation law to incorporate the outflow of the vehicles from the continuum space. The key for this addition is to assume point destinations instead of closed area destinations. Based on this assumption, we let $\mu(x, y, t)$ represent the trip completion rate at point (x, y) at time t, in units of flow per unit area. It follows that the conservation law becomes:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{f} = \lambda - \mu. \qquad \text{(conservation law consistent with MFD theory)} \quad (3.1)$$

To see the relationship between the variables in the discrete- and continuum-space model formulations, let Ω be the region for which the reservoir model has been defined.

Then,

$$n(t) = \int_{\Omega} \rho(x, y, t)\ dx\ dy, \qquad \text{(accumulation and density)} \tag{3.2a}$$

$$\Lambda(\Omega, t) = \int_{\Omega} \lambda(x, y, t)\ dx\ dy, \qquad \text{(inflow and demand rate)} \tag{3.2b}$$

$$\mathrm{M}(\Omega, t) = \int_{\Omega} \mu(x, y, t)\ dx\ dy, \qquad \text{(outflow and trip completion rate)} \tag{3.2c}$$

where n is the accumulation (number of vehicles inside the reservoir), Λ denotes the total inflow to the reservoir at any time t, and M is the total outflow from the reservoir at time t. This formulation is consistent with MFD theory because the reservoir model Eq. (2.3) can be obtained by integrating Eq. (3.1) over the region Ω and replacing the variables from Eqs. (3.2). Note that from the divergence theorem

$$\int_{\Omega} \nabla \cdot \mathbf{f}\ dx\ dy = \int_{\Gamma_o} \mathbf{f} \cdot \hat{\mathbf{n}}\ ds = 0, \qquad \text{(2D divergence theorem)} \tag{3.3}$$

where the last equality is due to the boundary condition Eq. (2.17b). Hence, the proposed conservation law Eq. (3.1) will lead to a new realization of the continuum space as shown in Fig. 3.1, where both the origin and destination point can be randomly distributed over the continuum space.

Recall assumption (2.2) of MFD theory stating that the trip completion rate on a physical link is proportional to the flow on that link. In the continuum, this amounts to assuming that the trip completion rate is proportional to the norm of the flux, $\mu = \|\mathbf{f}\|/\ell$, which in the isotropic case is

$$\mu = Q(\rho)/\ell, \tag{3.4}$$

which can be integrated over the region A to obtain

$$\mathrm{M}(A,t) = \int_A Q(\rho)/\ell \; dx \; dy, \tag{3.5a}$$

$$\approx Q(\rho) \; |A|/\ell, \tag{3.5b}$$

$$= \bar{Q}(n) \; L/\ell. \tag{3.5c}$$

where approximation (3.5b) follows when both the flux MFD and the density are approximately constant in A. Equality (3.5c) follows because the production in the continuum framework, $Q(\rho) \; |A|$, and in the discrete framework, $\bar{Q}(n) \; L$, are equal. To see this, one can use the fundamental traffic flow relationship in both frameworks to obtain $\rho V(\rho) \; |A|$ and $V(n)n$, respectively. Using the identity $n = \rho|A|$ gives $V(\rho) = V(n)$. Therefore, we have shown that the spatial integration of the continuum model leads to the reservoir model, as sought.

The introduction of the term μ in conservation law Eq. (3.1) suggests that the vehicles can complete their trip and exit the network at any point. Therefore, the proposed formulation is intrinsically capable of handling multi-commodity flows and multiple destinations, which is more compatible with the nature of the real-life net-

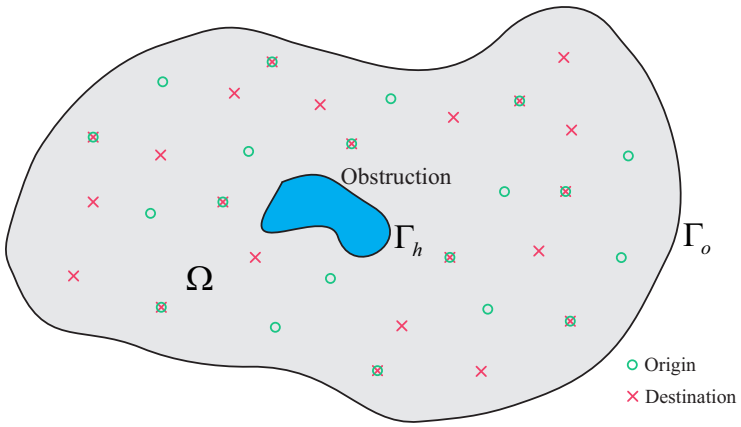

Figure 3.1: Randomly distributed origins and destination over the continuum space in the proposed consistent model.

works compared to the traditional continuum-space models, where there is only one or a few destination areas. When more than one destination is considered, the standard approach is to breakdown the total density into commodities, one per destination (see e.g., Hughes, 2002; Saumtally et al., 2013; Lin et al., 2017). As shown next, although some general insight might be derived from this disaggregation method, the general solution using this method can be cumbersome and intractable. Subsequently, a new disaggregation method based on travel direction is introduced in Section 3.2.2, which is amenable for the semi-Lagrangian solution proposed later in this chapter.

3.2.1 Vehicles Disaggregated by Destination (Isotropic Case)

In the proposed formulation, any point on the continuum space can be a destination; see Fig. 3.1. To understand how each destination contributes to the conservation law of global quantities in Eq. (3.1), we let group $i = 1, 2, \ldots$ denote the vehicles bound towards destination point $\mathbf{x_i} = (x_i, y_i)$, and subscript i to denote group-specific quantities, e.g. ρ_i, $\mathbf{f_i}$, ϕ_i, etc. Clearly,

$$\rho = \sum_i \rho_i, \qquad \mathbf{f} = \sum_i \mathbf{f_i} = \sum_i \rho_i \mathbf{u_i}, \qquad \text{and} \qquad \lambda = \sum_i \lambda_i. \tag{3.6}$$

The speed vector of each group is now given by

$$\mathbf{u_i} = -\frac{\nabla \phi_i}{||\nabla \phi_i||} V(\rho), \tag{3.7}$$

in the isotropic case, and notice that:

$$\|\nabla \phi_i\| = 1/\|\mathbf{u_i}\| = 1/V(\rho), \tag{3.8}$$

as in Eq. (2.11). It follows that

$$\mathbf{f_i} = -V(\rho)^2 \rho_i \, \nabla \phi_i. \tag{3.9}$$

The disaggregated outflow μ_i is less obvious because it should be identically zero everywhere except at the group's destination $\mathbf{x_i}$. Introducing the Dirac delta function $\delta_{\mathbf{x_i}}(\mathbf{x})$ defined as:

$$\delta_{\mathbf{x_i}}(\mathbf{x}) = \begin{cases} 1, & \text{if } \mathbf{x} = \mathbf{x_i}, \quad (3.10a) \\ 0, & \text{otherwise}, \quad (3.10b) \end{cases}$$

we have:

$$\sum_i \mu_i \; \delta_{\mathbf{x_i}}(\mathbf{x}) = \mu = V(\rho)\rho/\ell, \tag{3.11}$$

as required, and where the last equality follows from Eq. (3.4) and the fundamental traffic flow relationship $Q(\rho) = V(\rho)\rho$. With all, the conservation law of each group can be expressed as $\partial \rho_i / \partial t - \nabla \cdot \mathbf{f_i} = \lambda_i - \mu_i \; \delta_{\mathbf{x_i}}(\mathbf{x})$, where summation over all the groups will recover the global conservation law Eq. (3.1), as expected. Using Eq. (3.9) the conservation law for each group can also be expressed as:

$$\frac{\partial \rho_i}{\partial t} - \nabla \cdot \left(V(\rho)^2 \rho_i \, \nabla \phi_i \right) = \lambda_i - \mu_i \; \delta_{\mathbf{x_i}}(\mathbf{x}). \tag{3.12}$$

Following Hughes (2002), we examine two limit cases of Eq. (3.12), light and heavy traffic, under steady-state conditions. Expanding the divergence term in Eq. (3.12), rearranging and considering locations $\mathbf{x} \neq \mathbf{x_i}$, we have:

$$- \left(V(\rho) \nabla \rho_i \cdot \nabla \phi_i \; + \rho_i \left(2V'(\rho) \nabla \rho \cdot \nabla \phi_i + V(\rho) \nabla^2 \phi_i \right) \right) V(\rho) = \lambda_i. \tag{3.13}$$

At low free-flow densities, we can assume $\nabla^2\phi_i \approx 0$, $\rho_i \approx \rho \approx 0$ since speed is constant along the optimal path and consequently $\|\nabla\phi_i\|$ will also be constant. Thus, Eq. (3.13) becomes

$$\nabla\rho_i \cdot \nabla\phi_i \approx -\lambda_i/V^2(0), \tag{3.14}$$

which means that lines of constant density (or flow) are no longer perpendicular to lines of constant potential, as in Hughes model, unless the inflow λ_i is zero. Instead, the angle between the two increases with $\lambda_i/V^2(0)$ (since the dot product is proportional to the cosine of the angle between two vectors). We can say that a local increase in the inflow of group i increases the cost of the optimal route by imposing a detour.

At high congested densities, we have $||\nabla\rho|| << ||\nabla\phi||$, so that both $\nabla\rho \cdot \nabla\phi$ and $\nabla\rho_i \cdot \nabla\phi_i \approx 0$ and Eq. (3.13) becomes

$$\nabla^2\phi_i \approx -\lambda_i/(V^2(\rho)\ \rho_i), \tag{3.15}$$

known as the Poisson equation, where the right-hand side is negative and probably large in magnitude (because speeds are low), which means that the cost potential is increasing very rapidly and nonlinearly away from the destination, possibly inducing gridlock near the destination. However, if the inflow is to be restricted by the outflow, it is likely that the the right-hand side would be ≈ 0 and the steady-state solution should be close to Laplace's equation $\nabla^2\phi_i = 0$ as in Hughes model with zero inflow.

3.2.2 Vehicles Disaggregated by Direction (Anisotropic Case)

The multi-commodity formulation in the previous section is not easy to solve in Eulerian coordinates, especially when the number of destinations is large as in our case. Recognizing that in the vehicular networks there are only a few possible movement

directions at any given point (typically 4, one per each cardinal direction), here we propose:

$$\frac{\partial \rho_a}{\partial t} + \nabla \cdot \mathbf{f_a} = \lambda_a - \mu_a, \tag{3.16}$$

where $a = 1, 2, \ldots, m$ is an index for the direction of travel, and subscript a is used to denote group-specific quantities. Naturally, summation of Eq. (3.16) over all directions yields the global conservation law Eq. (3.1), as expected Notably, since the flux $\mathbf{f_a}$ is in direction a only, it suffices to consider a regular scalar function f_a, and therefore its divergence reduces to the scalar partial derivative:

$$\nabla \cdot \mathbf{f_a} = \frac{\partial f_a}{\partial x_a}, \tag{3.17}$$

where x_a is the scalar distance in direction a. Replacing Eq. (3.17) into Eq. (3.16) gives the 1-D kinematic wave model with sink and source terms:

$$\frac{\partial \rho_a}{\partial t} + \frac{\partial f_a}{\partial x_a} = \lambda_a - \mu_a. \tag{3.18}$$

The main distinction is that in the current model at a given point the speed and flux in each direction, v_a and f_a, are not only functions of ρ_a, but also of the densities in other directions as well. The next section will describe a semi-Lagrangian solution method to this model.

3.3 Semi-Lagrangian Multi-reservoir Solution Method

Here, we propose using a semi-Lagrangian solution scheme for the anisotropic case discussed earlier exploiting Eulerian or Lagrangian representations where it is most convenient. These schemes use an Eulerian computational grid (cells) over which the motion of particles is tracked. Using these trajectories, we estimate the continuum

variables at the grid points. This general idea has been applied copiously in the field of atmospheric sciences (Staniforth and Cote, 1991). In our case, we propose to use the Eulerian representation to compute the flux transfers between cells, and the Lagrangian representation to move discrete particles and compute the corresponding cell densities of each commodity; see Fig. 3.2. The following subsections will formulate each of the components in this figure.

3.3.1 The Computational Grid and Network

In the proposed method, an Eulerian uniform computational grid is defined by spatial cells (reservoirs) of rectangular shape of sides $\Delta x, \Delta y > 0$ together with the time-step Δt, which is assumed to be small enough so as to satisfy the Courant-Friedrich-Lewy (CFL) condition:

$$\max_{\rho} \left(\frac{\max |f_1'(\rho)|}{\Delta x}, \frac{\max |f_2'(\rho)|}{\Delta y} \right) \Delta t \leq 1, \tag{3.19}$$

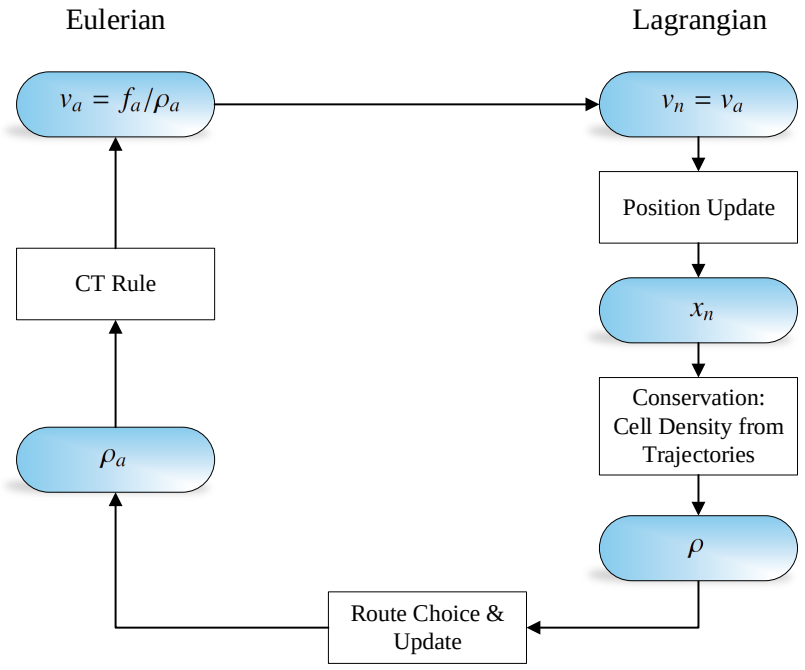

Figure 3.2: Flowchart of the proposed semi-Lagrangian solution framework.

which implies that a vehicle will not be able to traverse a cell in less than one time step. Assuming $\Delta y = \Delta x$ for simplicity, the numerical grid is then

$$x_i = i\Delta x, \qquad y_j = j\Delta x, \qquad t_k = k\Delta t, \qquad i, j, k = 1, 2 \ldots \tag{3.20}$$

It will be convenient to denote a cell by the *object* c, and allow the dot notation $c.i$ and $c.j$ to represent the discrete spatial coordinates of the cell's center. In this way, the numerical scheme will produce densities ρ_c^k that approximate $\rho(x_{c.i}, y_{c.j}, t_k)$. *For clarity in notation, from now on we drop the time index k as much as possible.*

On top of the Eulerian grid, we superimpose a Lagrangian network consisting of the directed links connecting every pair of adjacent cells; see Fig. 3.3. Each link, a, represents one of the four directions of travel, and contains the vehicles in the source cell, $a.s$, for which the minimum path to their destination passes by the adjacent target cell, $a.t$. Links do not have spatial dimensions, and the length of each link, ℓ_a, is simply an attribute to represent the average trip length from source to target cells. At each time-step k, a link has speed v_a^k, flux f_a^k and density ρ_a^k in units of vehicles per unit area. The speed on link a is computed using the fundamental traffic flow relationship:

$$v_a = f_a/\rho_a, \tag{3.21}$$

and all vehicles on link a move at this speed. The key to our solution method is that fluxes f_a also represent the flux at the boundary between source cell $a.s$ and target cell $a.t$, and are computed using Godunov's method, as explained momentarily.

3.3.2 The Route Choice Component

The role of the route choice component is the calculation of ρ_a, the density in cell c *willing* to use outgoing link a. A popular method in the literature to calculate

the ρ_a is to use partial flows to represent each origin-destination commodity and specify turning proportions at each cell, typically with a discrete choice model of the logit type (Hänseler et al., 2014; Yildirimoglu and Geroliminis, 2014; Hänseler et al., 2017). But as pointed out earlier, the solution to this problem in Eulerian coordinates is cumbersome, and here we adopt a Lagrangian approach instead.

We saw previously that to achieve RDUE conditions, vehicular motion should be in the direction with maximum potential reduction, i.e. in the direction of $-\nabla\phi/\|\nabla\phi\|$, which boils down to solving the Eikonal PDE Eq. (2.11). In the literature, this PDE is solved numerically in Eulerian coordinates with methods that are intimately connected with shortest path algorithms. This connection is not a coincidence since in a Lagrangian representation, we only need the direction of motion at each cell, which can be obtained with a standard shortest path routine on our Lagrangian network.

Once all shortest path trees have been computed, vehicles in cell c are assigned to the outgoing link a, which is the first link of their shortest path. This allows the computation of ρ_a, simply by counting how many vehicles are assigned to each link

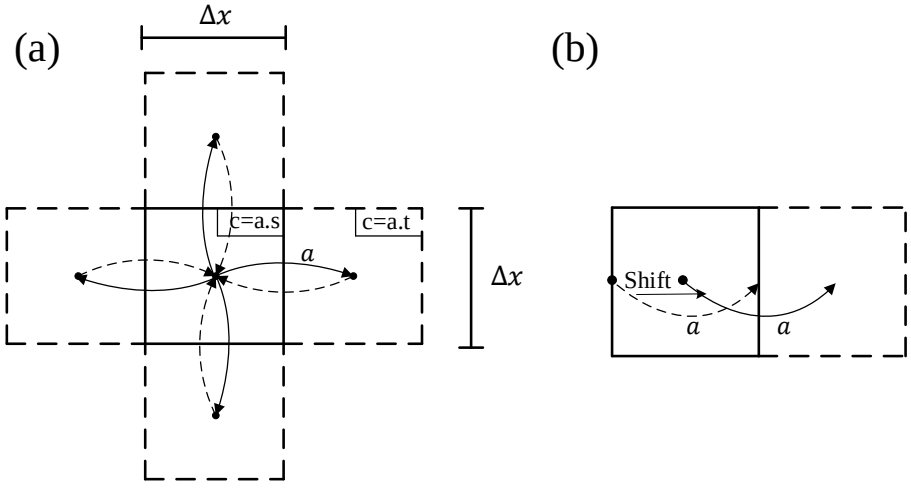

Figure 3.3: (a) Eulerian cells and Lagrangian links. Note that a link represents the distance traveled in the source cell only, as shown by the dashed arrow in (b). But we shift the beginning of the link to the centroid to avoid multiple overlaps.

(and dividing by Δx^2). The total density in cell c is therefore:

$$\rho_c = \sum_{a \in \mathcal{O}_c} \rho_a, \tag{3.22}$$

where $\mathcal{O}_c$ is the set of directed links from c to all its adjacent cells. The speed of each of the outgoing links is calculated by Eq. (3.21), where the flux on the link is determined by the Cell Transmission (CT) rule explained next.

3.3.3 The Cell Transmission Rule

As customary in the 1-D traffic flow literature, Godunov scheme is a preferred method due to its simplicity and accuracy compared to the traditional methods. Godunov's method was first used in traffic flow in the well-known Cell Transmission (CT) model (Daganzo, 1994b) to solve the 1-D traffic flow problem, i.e. a scalar conservation law in one spatial dimension with no source term. Here, we need a method in 2-D and with source term. Laval et al. (2016) proposed a Godunov-type scheme based on variational theory (Daganzo, 2005a) to incorporate source terms in 1-D problems, and also showed that the CT scheme is able to cope with source terms, albeit not as accurately. For simplicity, our starting point here is the CT scheme, which can be expanded to two spatial dimensions as shown below.

The key to Godunov's method is the computation of the fluxes at the boundary between adjacent cells, which are obtained by finding the entropy solution to the corresponding Riemann problems. To illustrate, let us assume for a moment that all vehicles in cell want to go to one of the adjacent cells, i.e. there is a single user class or *commodity*. The Riemann problem, in this case, is to obtain the maximum flux at the boundary between these two cells, given their initial densities. If these two cells are connected by link a, then the CT rule is precisely the solution to the Riemann

problem:

$$f_a = \min\{D(\rho_{a.s}), S(\rho_{a.t})\}, \qquad \text{(CT rule, single commodity)} \tag{3.23}$$

where $D(\cdot)$ and $S(\cdot)$ are the well-known demand and supply flux functions in traffic flow, respectively:

$$D(\rho) = \begin{cases} Q(\rho), & \rho \leq \rho^* \\ Q^*, & \rho > \rho^* \end{cases} \qquad \text{and} \qquad S(\rho) = \begin{cases} Q(\rho), & \rho \geq \rho^* \\ Q^*, & \rho < \rho^* \end{cases} \tag{3.24}$$

where ρ^* and Q^* are the critical density and capacity of the flux MFD, $Q(\rho)$, respectively; see Fig. 3.4.

Nevertheless, flows are multi-commodity in two dimensions, one commodity per each origin-destination pair in the network. At the cell level, however, there are only a few commodities that matter, one per outgoing link a. The mapping from the origin-destination commodities to the outgoing link commodities is a result of the route choice model, which gives us the density willing to use each outgoing link ρ_a. The problem of deciding the corresponding flux through that link, f_a, given all the

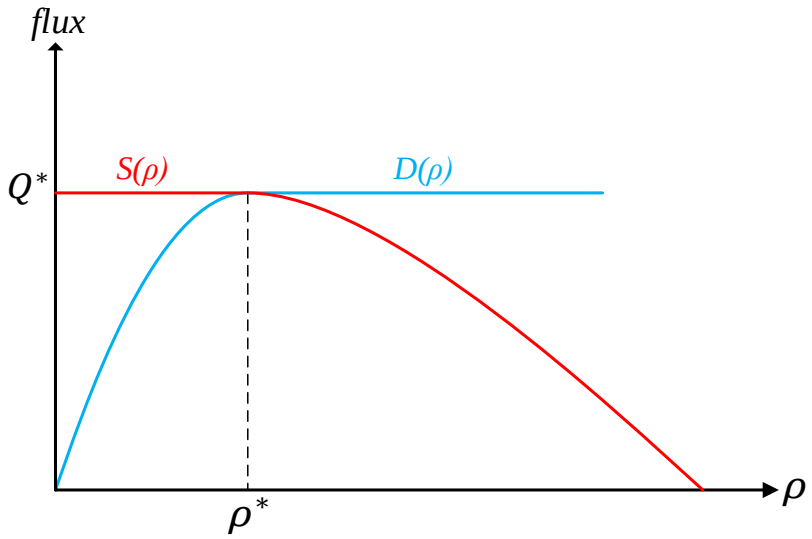

Figure 3.4: Demand, $D(\rho)$, and supply, $S(\rho)$, functions

competing densities willing to travel to cell $a.t$ is identical to the network junction problem in the literature (see e.g., Jin, 2012, and references therein). The solution to this problem can be expressed as the following multi-commodity CT rule:

$$f_a = \min\{D_a, S_a\}, \qquad \text{(CT rule, multiple commodities)} \tag{3.25}$$

and the speed on link v_a can then be computed with (3.21). There has been multiple formulations of these commodity-specific demand and supply functions in the literature, in the context of intersection modeling (see e.g. Jin, 2017, and references therein). Here, we use:

$$D_a = D(\rho_{a.s}) \cdot \frac{\rho_a}{\rho_{a.s}}, \qquad \text{and} \qquad S_a = S(\rho_{a.t}) \cdot IT(b \in \mathcal{I}_{a.t}), \tag{3.26}$$

where $\mathcal{I}_c$ is the set of directed links from c's adjacent cells into c. Eq. (3.26) describes that demand and supply of link a are a fraction of the total demand in source cell $a.s$ and supply in the target cell $a.t$, respectively. The demand fraction is proportional to the number of vehicles willing to use link a. For the supply fraction, we use the Incremental Transfer (IT) principle (Daganzo et al., 1997), where the total supply $S(\rho_{a.t})$ is viewed as a reservoir being filled gradually from links coming from each of the four adjacent cells, i.e, links $b \in \mathcal{I}_{a.t}$. In this way, the IT principle states that each reservoir will send fluid downstream until upstream reservoirs are empty or the receiving reservoir is full.

The output of this component is the link speeds v_a, which are updated using Eqs. (3.21), (3.25) and (3.26). Next, we update vehicle positions accordingly.

3.3.4 Vehicle Position Update

Let v_n^k and x_n^k be vehicle n's speed and distance traveled along the current link a at time-step k, respectively. Assuming that the speed remains constant during a time

step, the position of this vehicle in the next time step will be:

$$x_n^k = x_n^{k-1} + \Delta t \; v_n^{k-1}. \tag{3.27}$$

When at a time step t^k, the distance traveled by the vehicle n on its current link becomes greater than the link length ℓ_a, the vehicle has crossed the cell boundary (i.e. entered cell $a.t$ and one of its outgoing links based on the shortest path of the vehicle) at some time t_0, and all relevant variables are updated accordingly. In order to reduce the numerical error, we assume that the vehicle travels the remainder of time step after t_0 in the source cell, such as its position becomes:

$$x_n^k = (t^k - t_0) \; v_n^{k-1}, \tag{3.28}$$

where x_n^k is now the distance traveled in cell $a.t$ and v_n^{k-1} is the speed of the new link in cell $a.t$ at the previous time step.

A crucial point is how to compute the speeds v_n^k. In the spirit of the MFD one might postulate $v_n^k = V(\rho_{a.s}^k)$, which is a common assumption in the literature (Yildirimoglu and Geroliminis, 2014; Yildirimoglu et al., 2015; Leclercq et al., 2017a), but can lead to *stalling* (complete gridlock) on the multi-reservoir configurations. To see this, notice that when a cell reaches jam density, both the speed and the outflow will be zero even when neighboring cells might be freely flowing. Instead, here we use the CT rule Eq. (3.25) to compute fluxes f_a and then use the fundamental traffic flow relationship to compute the speed on link, v_a, per Eq. (3.21) and we assign this speed to all vehicles in the link, i.e.:

$$v_n = v_a. \tag{3.29}$$

This formulation avoids unrealistic stalling because if a cell is at jam density its flow will be determined by the supply on the downstream cell, which is zero only if that cell is also full.

3.3.5 Conservation of Vehicles

After all vehicles' positions and current cell have been updated, total cell densities ρ_c are updated simply by (i) counting the number of vehicles in each cell (ii) add the number of vehicles entering the network at cell c, Λ_c, (iii) subtract the number of vehicles whose destination is cell c, and (iv) divide by the cell area Δx^2. Notice that it is customary to cap the inflow at the supply of the origin cell; i.e.,

$$\Lambda_c = \min\{S(\rho_c)\ ,\ \lambda(x_{c.i}, y_{c.j}, t_k)\Delta x\}\ \Delta t \Delta x, \tag{3.30}$$

since otherwise unrealistic gridlock can occur. At this point, we recalculate the minimum paths with the updated link speeds from Eq. (3.21).

3.4 Numerical Experiments

In this section, we have used the aforedescribed numerical solution method to simulate a square-shaped city of side b with parabolic homogeneous flux-MFD given by $Q(\rho) = \rho u \cdot (1 - \rho/\rho_j)$, where u is the free-flow speed and ρ_j is the jam density (in units of vehicles per unit area). We eliminate these two parameters by measuring distance in units of $\rho_j^{-1/2}$ and time in units of $\rho_j^{-1/2} u^{-1}$, which gives:

$$Q(\rho) = \rho \cdot (1 - \rho). \tag{3.31}$$

Total arrivals are given by a Poisson process of constant rate λ veh/hr-km^2, while origins are uniformly distributed across the region. The average trip length inside each cell to all adjacent cells is assumed as the side length of the cell, $\ell_a = \Delta x$. Two

sets of experiments are performed differing only in the treatment of destinations: (i) a single region-destination, e.g. CBD, and (ii) multiple point-destinations.

3.4.1 A Single-destination Region

In this experiment, vehicles travel toward a single destination located at the center of the square region starting their trips from uniformly distributed origins. Different simulation runs have been done using different cell sizes, while the size of the destination area is kept constant across simulation runs in order to have the same maximum outflow rate, which makes us able to study the convergence of the model. This convergence is measured in terms of the delay with respect to the analytical solution, which is simple for this problem since the destination will serve at constant capacity when congested. Therefore, we can find the theoretical cumulative departure curve and compare it with results of the experiment. Other assumptions for this experiment are summarized in Table 3.1.

The same origin-destination points are used for all simulations. Fig. 3.5 shows the cumulative arrivals, entrances and departures for a 10 by 10 and a 40 by 40 partitioning of the network area. The analytical cumulative departure curve is plotted by shifting the constant departure rate curve by the average free flow travel time of vehicles to the right. Although in the experiment with coarser partitioning we observe slight variations in the experimental departure curve whereas its slope is faintly lower

Table 3.1: Assumptions of the single-destination experiment

Parameter	**Value**
Side of the square region, b	5000 meters
Perimeter of the destination area	500 meters
Jam accumulation of the region	20,000 vehicles
Loading duration	10 minutes
Arrival rate, Λ	5 veh/s
Maximum completion rate, M	1 veh/s
Free-flow speed, u	10 m/s

than the expected maximum departure rate, the departure curve for the experiment with finer partitioning fits the analytical departure curve impeccably.

The distinction in the arrival and entrance curves arises due to the supply cap Eq. (3.30), i.e. the demand at the origin cell is equal or greater than its supply, therefore the vehicle has to wait for adequate supply to begin its trip. This phenomenon mostly arises at the cells near to the destination area, since these cells are utilizing the maximum supply they get, most of the time. As seen in Fig. 3.5, the entrance delay, i.e. the time that vehicles have to wait for sufficient supply to enter the network, is considerably lower in the experiment with finer partitioning.

Furthermore, Fig. 3.6 shows the average delay per vehicle for different number of cells versus the analytical approximation of average delay. It can be seen that the average delay converges to the analytical approximation with increasing the number of cells, i.e. decreasing the size of cells.

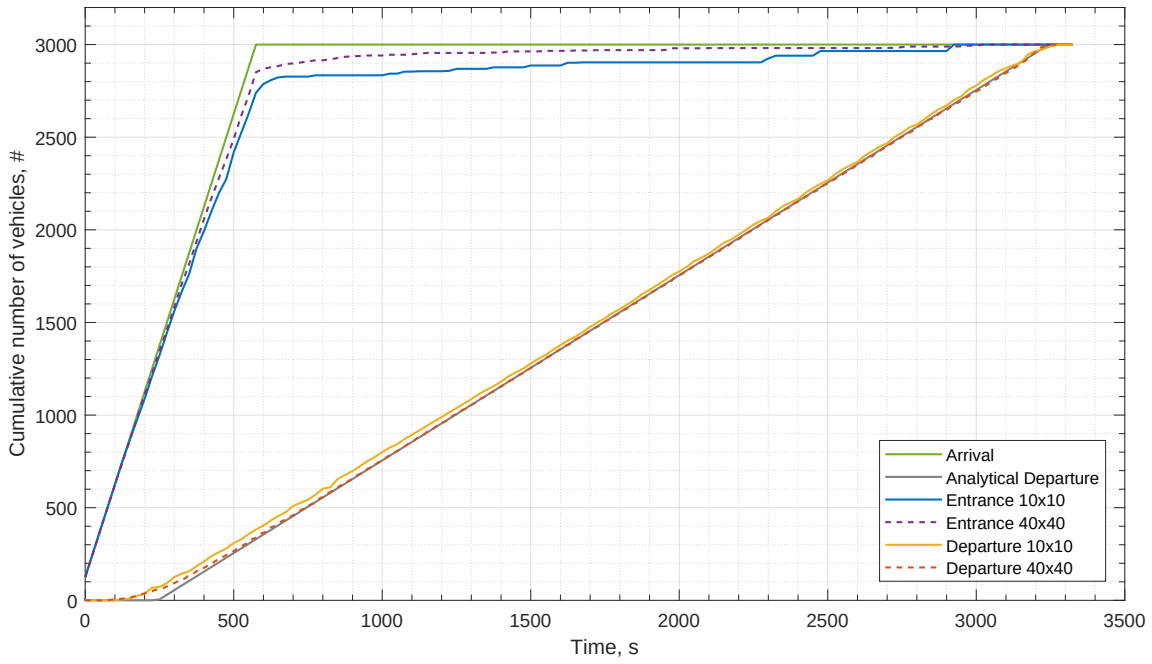

Figure 3.5: Cumulative arrival, entrance, and departure curves for two different experiments

The results suggest that the overall performance of the model improves with partitioning the region into smaller cells. However, by decreasing the cell size, the jam accumulation of each cell also decreases, which could result in stalling problem under congested conditions if the speed was directly taken from the MFD as in the discrete-space literature, which is not the case in the proposed model since the flux and speed are calculated using the supply and demand functions.To get a better insight into the performance of the model under congested conditions, in another experiment, the network is divided into 125-meter-long square cells, i.e. 40 cells in each direction, and it is loaded for 20 minutes with an average arrival rate of 5 vehicles per second, while the maximum outflow rate of the destination area is one vehicle per second. Fig. 3.7 shows the cell occupancies at time $t = 1200$ s , the end of loading duration, while the network is the most congested.

The interesting point in Fig. 3.7 is that the peak of density is not immediately on the borders of the destination area, rather it is observed at a distance from the destination area. An approximate steady-state analytical solution to this problem without inflow supply cap, can be found in polar coordinates (r, θ), assuming that

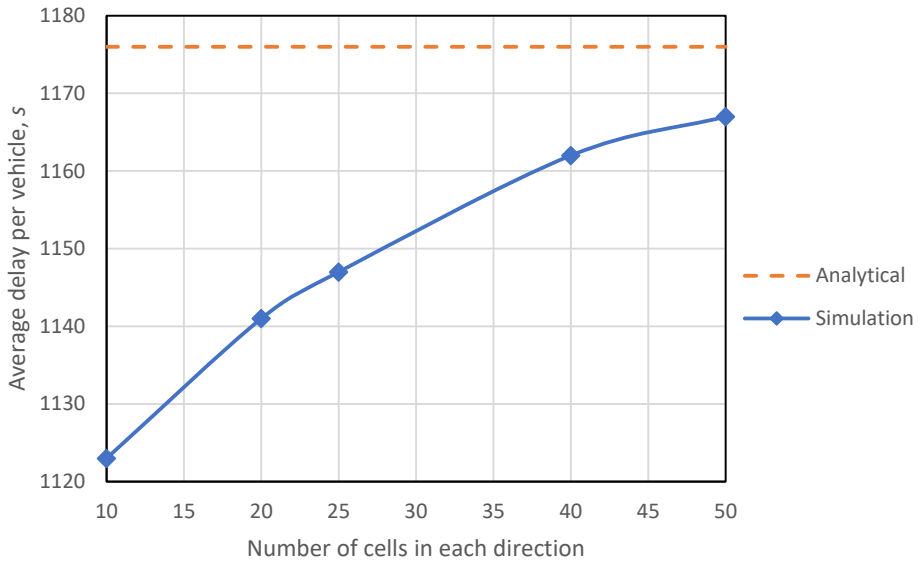

Figure 3.6: Average delay per vehicle: Analytical vs. Simulation

the destination is a circle of radius r_0. As shown in Newell (1980, p. 130), fluxes should be in the radial direction and rotationally symmetric, and therefore the conservation law in steady-state can be expressed as:

$$\frac{\partial}{\partial r}(rQ(\rho)) = r\lambda. \tag{3.32}$$

By imposing that far away from the destination the inflow vanishes, Newell shows that the flux decays with distance as $1/r$, independently of the shape of the flux function. Here, we use the parabolic flux-MFD Eq. (3.31) and solve Eq. (3.32) to find:

$$\rho(r) = \frac{1}{2} \pm \sqrt{2f_0\ r_0 + r/2 - \lambda(r^2 - r_0^2)}, \tag{3.33}$$

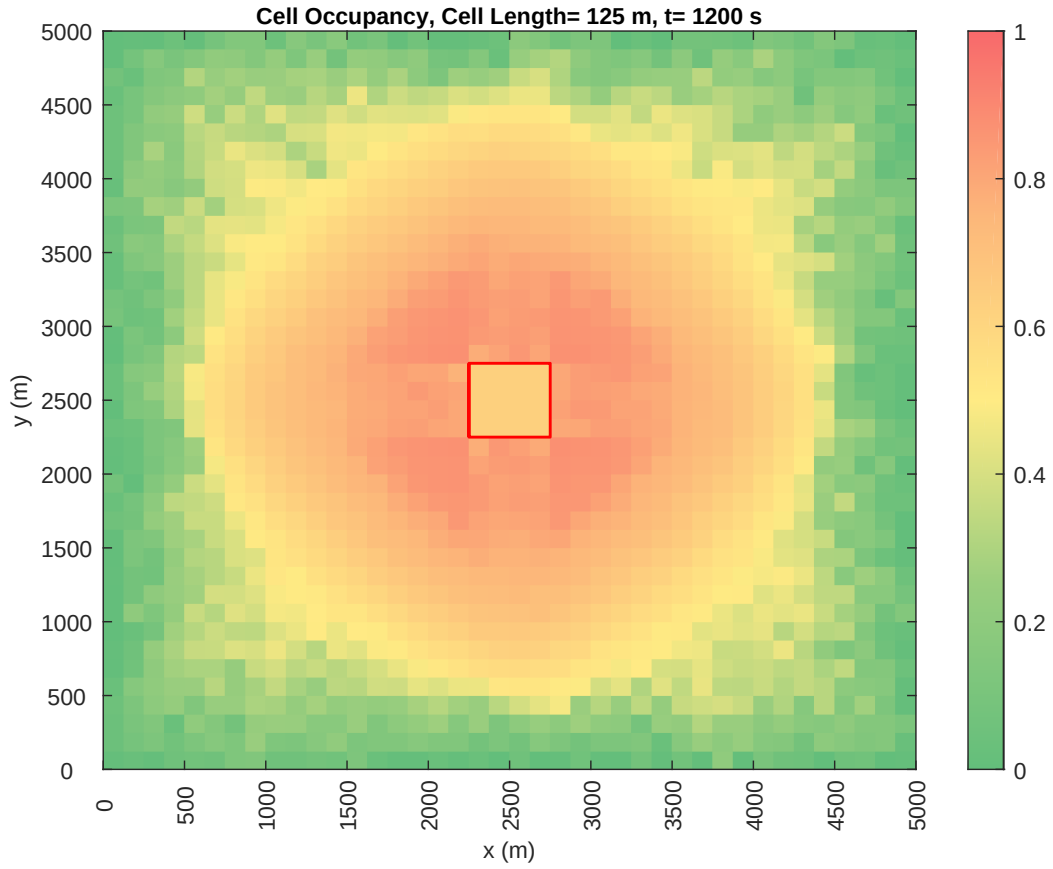

Figure 3.7: Distribution of vehicles over the region partitioned into a 40 by 40 mesh

where $\rho_0 = \rho(r_0)$ is the boundary condition at the destination's perimeter, $f_0 = Q(\rho_0)$ is the initial flux and $\pm$ is the sign of $(\rho_0 - 1/2)$. Fig. 3.8 shows a cross-section of the density obtained with the simulation superimposed to the analytical approximation Eq. (3.33). It can be seen that despite some small variations, simulation results accord well with the analytical solution.

The flux as a function of distance can be readily obtained using Eq. (3.33):

$$Q(\rho(r)) = \frac{\lambda(r^2 - r_0^2)}{2r} + \frac{f_0\ r_0}{r}, \tag{3.34}$$

where it can be seen that when the inflow vanishes ($\lambda = 0$) the flux indeed decays as $1/r$ as predicted by Newell, but if it is nonzero then eventually the flux will start increasing linearly with r; the minimum flux happens at location r^* given by:

$$r^* = \sqrt{r_0\ (2f_0/\lambda - r_0)}. \tag{3.35}$$

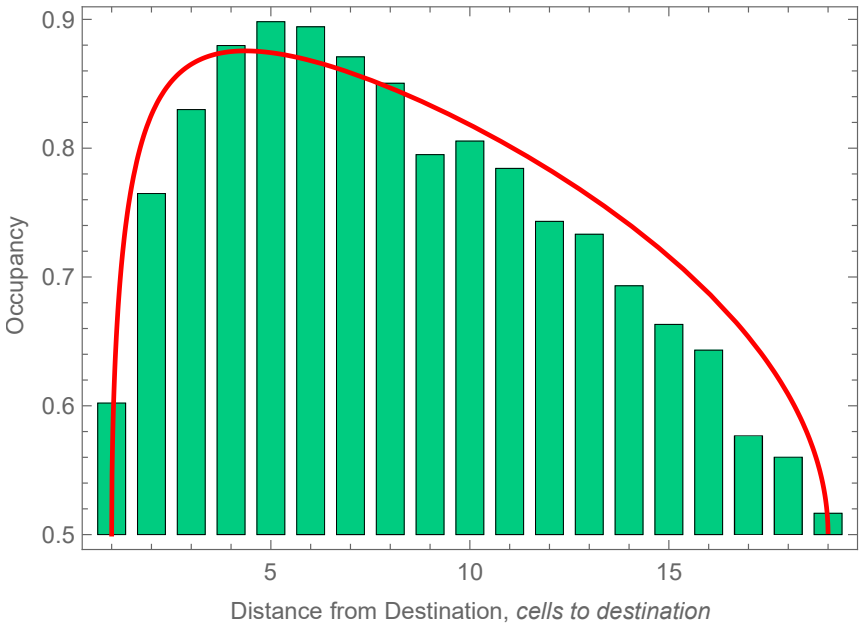

Figure 3.8: Density Cross-section: Analytical, Eq. (3.33) vs. Simulation

This result indicates that congestion is most severe at a distance r^* from the destination, and that this distance increases as the square root of f_0/λ.

3.4.2 A Multi-destination Region

In this experiment, the arrival rate and other assumptions are similar to the previous experiment but now both origins and destinations are uniformly distributed across the region. In steady-state, this problem is characterized by the fundamental traffic flow relationship:

$$\rho = \frac{f}{v} = \frac{\lambda E[D]}{v}, \tag{3.36}$$

where v is the (deterministic) steady-state speed in the network, and $E[D]$ gives the expected value of the distance traveled by a driver, D. Notice that the flux $f = \lambda E[D]$ is a constant of this problem, and can be used to obtain the speed v using the flux-speed MFD provided we know the traffic regime. In our case we have:

$$v(f) = \frac{1}{2}\left(1 \pm \sqrt{1-4f}\right), \tag{3.37}$$

where $\pm$ represents addition when in free-flow and subtraction in congestion.

In our case the Manhattan distance traveled by a driver is $D = |x_1 - x_2| + |y_1 - y_2|$, where subscripts 1 and 2 refer to the coordinates of the origin and destination, respectively. For origins and destinations uniformly distributed on a square region of side b, one can show that the probability density function (pdf) of D, $p_D(d)$, is:

$$p_D(d) = \frac{1}{b}\begin{cases} -\frac{2}{3}(d/b-2)^3, & 1 \le d/b \le 2, \\ \frac{2}{3}(d/b)\left((d/b)^2 - 6d/b + 6\right), & 0 < d/b < 1, \\ 0, & \text{otherwise}, \end{cases} \qquad \text{(pdf of } D\text{)} \tag{3.38}$$

where one can verify that $E[D] = 2b/3$; therefore, we can express the speed as a function of the demand

$$v(\lambda) = \frac{1}{2}\left(1 - \sqrt{1 - 8\lambda b/3}\right). \tag{3.39}$$

The travel time random variable is simply $T = D/v$, and its pdf given by:

$$p_T(t) = v p_D(vt). \qquad \text{(travel time pdf)} \tag{3.40}$$

Fig. 3.9 shows travel time histograms from the simulation for different values of the demand parameter λ. The thick lines correspond to the analytical pdf Eq. (3.40), where v is obtained using Eq. (3.39). It can be seen that the agreements between analytical and simulated histograms is apparent. Although not surprising for the free-flow histogram, the good fit with the congested histogram indicates that Eq. (3.40) captures the correct shape of these density functions without the need of an extra parameter.

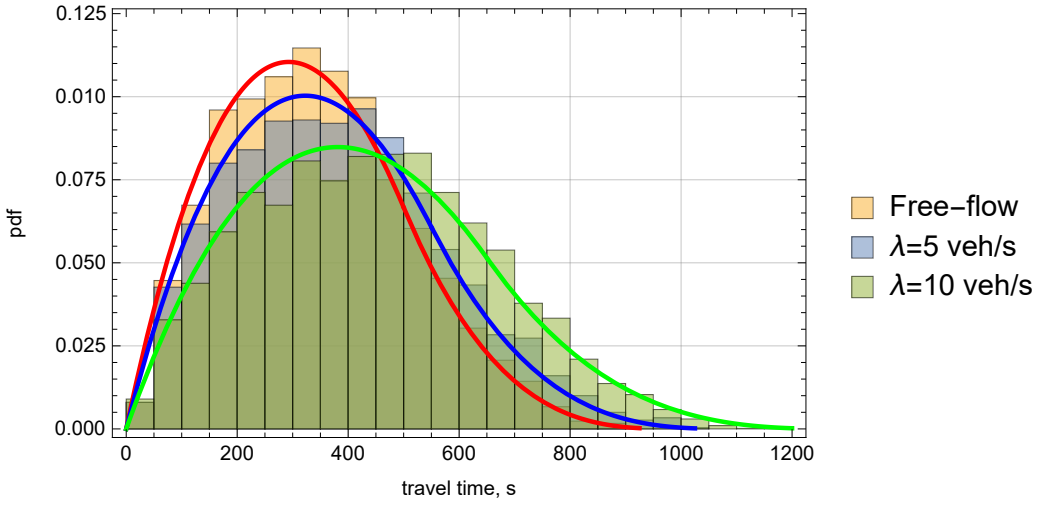

Figure 3.9: Simulated and analytical (red) travel time histograms

3.5 Chapter Summary

Existing continuum-space models for cities do not consider the completion rate term μ in Eq. (3.1), perhaps because they are based on pedestrian models where it is customary to assume that travelers exit the network only at discrete destinations represented as small regions, as in Fig. 2.4. We argue that region-destinations are problematic, and should be avoided in general. The size of these regions is arbitrary and are devoid of flow. This means that the spatial integration will not lead to a consistent model unless some additional considerations are included. But even then one could argue that the size of the destinations might affect the shape of the MFD, which complicates matters even further. Additionally, we saw that in the isotropic case the missing term is $\mu = Q(\rho)/\ell$, and therefore in current formulations, the average trip length is also ignored. Incorporating destinations continuously distributed across the network would be a daunting task in the existing framework, but a very simple one if the MFD concepts are incorporated as in the formulation proposed here.

Section 3.2 formulates a *consistent* continuum MFD model, in the sense that as the size of the subregions tends to zero, the limit of the discrete model corresponds to the continuum model. The emergence of trip completion rate term, μ, in the consistent conservation law Eq. (3.1) suggests that the proposed model is capable of incorporating randomly distributed destinations all over the continuum space, which is in more conformance with the nature of real-life networks. Our motivation is that existing multi-reservoir MFD models might not be consistent, which would imply that its predictions might change drastically upon changes in the partition of the network. The simplest way to accomplish consistency is to start with a numerical scheme to solve our problem and interpret each spatial cell as a subregion of the multi-reservoir model.

The numerical solution method presented here is based on the assumption that there are only a limited number of possible movement directions at each point in the vehicular networks, unlike the pedestrian flow models, where the pedestrians can move at any desired direction. By choosing rectangular cells, there are only four possible movement directions from each cell to each of the neighboring cells. Although the MFD is the same in all four cardinal directions, this does not mean that the speed v_a in all directions is the same, necessarily. In fact, our commodity-specific demand and supply functions in Eq. (3.26) imply that the speed on all outgoing links are different in congestion, and determined by the downstream reservoir of each link. However, in congestion all incoming links will have the same speed since the IT principle splits the supply of the receiving cell evenly among incoming links. Of course, different priorities in the IT principle could be assigned to each cardinal direction to reflect uneven infrastructure or signal timing in each direction. This solution appears superior to imposing a different MFD for each direction altogether because extra conditions should be added to govern any correlation or interactions between the traffic states in all directions.

The proposed framework can be extended in a number of directions in a rather straightforward way, including reservoirs of arbitrary shapes, different cost functions, and inhomogeneous cities. Extensions such as users with different car-following behavior e.g. trucks, buses or automated vehicles, are more challenging because the different fleet compositions should change the MFD in ways that are not well understood yet.

CHAPTER 4
DYNAMIC SYSTEM OPTIMUM TRAFFIC ASSIGNMENT IN CONTINUUM SPACE

This chapter is modified from the following presented conference paper:

Rafegh Aghamohammadi, Ludovic Leclercq, and Jorge A. Laval (2020). "A Semi-Lagrangian Dynamic System Optimum Traffic Assignment model in Continuum Space". In: 99^{th} *Annual meeting of Transportation Research Board*. Washington, D.C.

4.1 Introduction

The literature review on the macroscopic DTA models in Chapter 2 of this dissertation highlights the gaps in the literature and states that a well-defined discrete-space DTA model can be interpreted as the numerical solution of a continuum-space DTA model; i.e., when the size of regions tends to zero while keeping the total area constant, the discrete-space model tends to the continuum-space model. To fill this gap between two modelling approaches, Chapter 3 develops a novel continuum-space DTA framework which is consistent with the MFD theory and assumptions. The route choice component of this DTA framework is arbitrary and the semi-Lagrangian solution method presented in Section 3.3 achieves the reactive dynamic user equilibrium (RDUE) goal by minimizing the instantaneous travel cost of the individual users.

Most of the developed models in the macroscopic DTA literature propose dynamic user equilibrium (DUE) solutions, which do not necessarily result in the minimum total travel cost over the network and might not be favorable from a system point of view. Achieving dynamic system optimum (DSO) conditions by minimizing the total

travel time of the users in two-dimensional large-scale networks is cumbersome and to the best of our knowledge, there are only two studies in literature trying to develop DSO condition, one in discrete space (Yildirimoglu et al., 2015) and the other one in continuum space (Tao et al., 2014).

In this chapter, we revisit the solution method in Section 3.3 and try to achieve semi-DSO conditions using the instantaneous travel times for the sake of minimizing the total travel cost of all travelers using the proposed consistent DTA framework.

4.2 Background

The basic notion behind the continuum-space traffic models is that when the network is dense enough such that the distances between road segments are small compared to the size of network, the network can be approximated by a continuum space where the vehicles can circulate at any point $\mathbf{x} = (x, y)$ of a Euclidian two-dimensional domain $\Omega \subset \mathbb{R}^2$; see Fig. 4.1(i). The traffic dynamics inside the continuum space is governed by the following conservation law (CL) partial differential equation (PDE):

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{f} = \lambda, \qquad \text{(traditional 2D conservation law)} \tag{4.1}$$

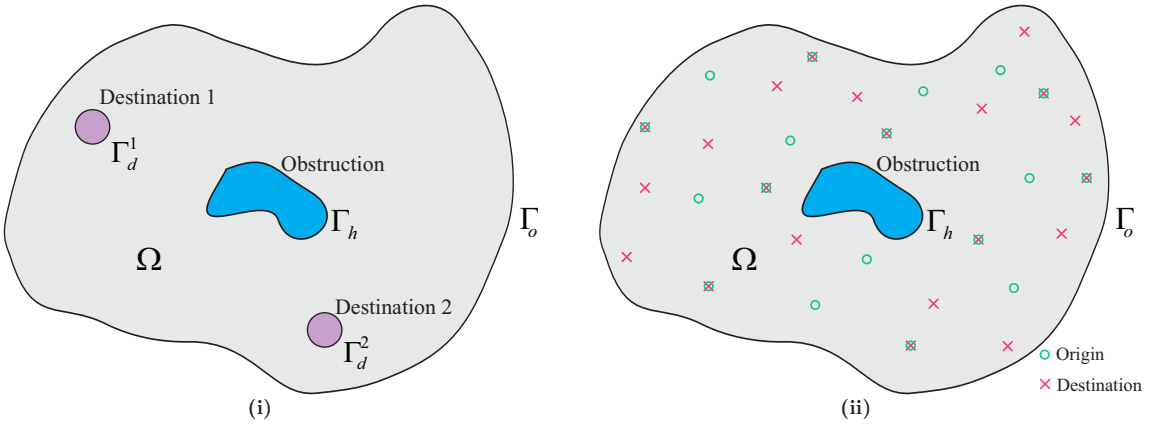

Figure 4.1: Illustration of the continuum-space models: (i) the traditional models based on (Hughes, 2002), (ii) the model proposed in Chapter 3 of this dissertation.

where boldface symbols indicate vector quantities, $\rho(x, y, t)$ represents the time-varying density of travelers at point (x, y) and time t (in units of # per unit area), $\mathbf{u} = (\mathbf{u_1}(\mathbf{x}, \mathbf{y}, \mathbf{t}), \ \mathbf{u_2}(\mathbf{x}, \mathbf{y}, \mathbf{t}))$ denotes the velocity vector, $\mathbf{f}(x, y, t) = (f_1(x, y, t), \ f_2(x, y, t)) \equiv \rho\mathbf{u}$ is the flux vector (in flow units per unit distance), $\nabla \cdot \mathbf{f} = \partial f_1/\partial x + \partial f_2/\partial y$ is the spatial divergence of $\mathbf{f}$ and $\lambda(x, y, t)$ represents the time-varying demand rate (in flow units per unit area).

The final model consists of the CL PDE Eq. (4.1) coupled with an Eikonal or Hamilton-Jacobi PDE for route choice, which is solved using the traditional numerical PDE solution methods. The solution methods in the literature mostly solve the model for RDUE conditions and Tao et al. (2014) presents the only continuum-space DTA model trying to find DSO solution. In this study, the system cost is expressed as:

$$\Phi(\rho, \theta) = \iiint_{\Omega \times T} \kappa\rho \ dx \, dy \, dt, \qquad \text{(system cost)} \qquad (4.2)$$

where $\theta(x, y, t)$ is the directed angle of speed representing travelers' route choice, κ is the value of time and T is the length of analysis period. The DSO model is formulated as an optimization problem with an objective function to minimize $\Phi(\rho, \theta)$ subject to proper constraints. The proposed numerical solution method first discretizes the continuum space using a finite volume method (FVM) and later optimizes a local form of the discretized objective function using a heuristic method. In the heuristic optimization method, the RDUE solution is used as the initial approximation for the DSO solution. Since the analytic properties of the proposed model were not fully studied by the authors, they were not able to develop a global optimal solution for the model. In order to reduce the computational cost, the computational grids are too coarse, which leads to a large degree of numerical error.

In the discrete space, Yildirimoglu et al. (2015) is the only study investigating DSO conditions, where each region is divided into a number of sub-regions with well-defined

MFDs and routes are redefined as sequences of sub-regions instead of sequences of links in Yildirimoglu and Geroliminis (2014). Two macroscopic traffic models are developed in region-level and subregion-level, where the speed of all vehicles inside a subregion is equal to sub-region's average speed taken from its speed MFD.

To overcome the limitations and drawbacks of the current MFD-based DTA models, Chapter 3 has proposed a new continuum-space DTA framework consistent with the MFD theory and assumptions. In this model, the conservation law PDE Eq. (4.1) has been revised and rewritten as:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{f} = \lambda - \mu. \qquad \text{(2D conservation law in the new framework)} \quad (4.3)$$

where $\mu(x, y, t)$ is the trip completion rate at point (x, y) at time t, which has been associated with the outflow MFD. The introduction of this term to the conversation law enables the model to incorporate destinations at any point over the network as seen in Fig. 4.1(ii).

A semi-Lagrangian solution method has also been developed for the proposed model, which is able to solve the RDUE problem in the two-dimensional space for both single-destination and multi-destination regions. This solution method utilizes an Eulerian grid with infinitesimal cells and superimposes a Lagrangian Network on top of that consisting of directed links connecting each pair of adjacent cells. While this model is consistent with both modeling approaches, the RDUE route choice component may not result in the minimum total cost in the network over the modeling period and it might not be favorable from a system perspective.

4.3 Methodology

4.3.1 Nature of the Solution

The nature of the DSO solution is well understood for simple problems (Muñoz and Laval, 2005; Laval et al., 2015). In essence, it requires that bottlenecks to be utilized at their capacity, while keeping their queues at zero for as long as possible. Once a particular bottleneck becomes operating at capacity, the extra inflow should be diverted to a longer route provided that the extra free-flow travel time on that route is less then the *marginal cost* of the original route at that time.

The difficulty resides in the computation of the marginal cost of a certain route, as it reflects the extra travel time imposed to all users in the network until the end of the rush, i.e. until congestion disappears everywhere in the network. Adding an additional driver to a congested route will cause an extra delay (given by the inverse of the bottleneck capacity on that route) to all subsequent drivers using this route, and possibly to the users of other intersecting routes. However, when the route is not congested, this marginal cost is identically zero. Once the marginal costs are known, it is well known in the literature that the DSO solution can be obtained in the same way as DUE solution but by minimizing the total marginal cost instead of the total user cost, typically the travel time.

Nevertheless, in two-dimensional networks, obtaining the optimal DSO solution becomes more complicated due to interdependence of the local directional flows and their impact on travel cost and route choice behavior of other travelers in the network. Here, we propose a *reactive* DSO solution method *aiming to minimize the instantaneous or perceived total travel time inside the network rather than the actual or experienced travel time*. This is achieved at each timestep by first finding the shortest paths all over the network using the current travel times, which are obtained from the solution of the previous timestep, and later assigning the vehicles at each

point to the possible paths with the objective of minimizing the local marginal travel cost considering the interactions between directional flows. Although the proposed solution method might not result in the global DSO solution minimizing the actual total travel cost, it is a step in this direction, and is based on an approximation of marginal costs that does not require knowing the actual marginal cost until the end of the congestion.

4.3.2 Proposed Solution Method

In the numerical scheme presented in Section 3.3, the two-dimensional continuum space is divided into a number of homogeneous rectangular zones, i.e. cells, of sides $\Delta x, \Delta y > 0$. The density of each cell computed by the numerical scheme at time step τ, ρ_c^τ, approximates the density of its centroid, $\mathbf{x_c} = (x_c, y_c)$, at time $t_\tau = \tau\ \Delta t$. The time-step Δt is assumed to be small enough to satisfy the Courant-Friedrich-Lewy (CFL) condition:

$$\max_{\rho} \left(\frac{\max |f_1'(\rho)|}{\Delta x}, \frac{\max |f_2'(\rho)|}{\Delta y} \right) \Delta t \leq 1, \qquad \text{(CFL condition)} \tag{4.4}$$

which implies that no vehicle will be able to traverse more than a cell during one time step. Each zone is connected to the adjacent zones via *links*. In order to compute the fluxes at the boundaries between adjacent connected zones, links, l, are defined as the set of all possible paths from the source zone, $l.s$, to the target zone, $l.t$. For example, Fig. 4.2(i) shows some of the possible paths consisting the link from zone a to zone b. As seen in this figure, all the travel on a link takes place in its source zone.

The total demand and supply of each zone are calculated using the well-known demand, $D(\cdot)$, and supply, $S(\cdot)$, functions as:

$$D(\rho) = \begin{cases} Q(\rho), & \rho \leq \rho^* \\ Q^*, & \rho > \rho^* \end{cases}, \qquad \text{and} \qquad S(\rho) = \begin{cases} Q(\rho), & \rho \geq \rho^* \\ Q^*, & \rho < \rho^* \end{cases}, \tag{4.5}$$

where ρ^* and Q^* are the critical density and capacity of the flux MFD, $Q(\rho)$, respectively. In two dimensions, flows are multi-commodity, i.e. one per each origin-destination pair, but at the cell level, there are only a few commodities that matter, one per each outgoing link a. The mapping from the origin-destination commodities to the outgoing link commodities is a result of the route choice model, which gives us the density willing to use each outgoing link, ρ^a. The problem of deciding the corresponding flux through that link, f^a, given all the competing densities willing to travel to cell $a.t$ is identical to the network junction problem in the literature (see e.g., Jin, 2012, and references therein). The solution to this problem can be expressed as the following multi-commodity CT rule:

$$f^a = \min\{D^a,\ S^a\}, \qquad \text{(CT rule, multiple commodities)} \qquad (4.6)$$

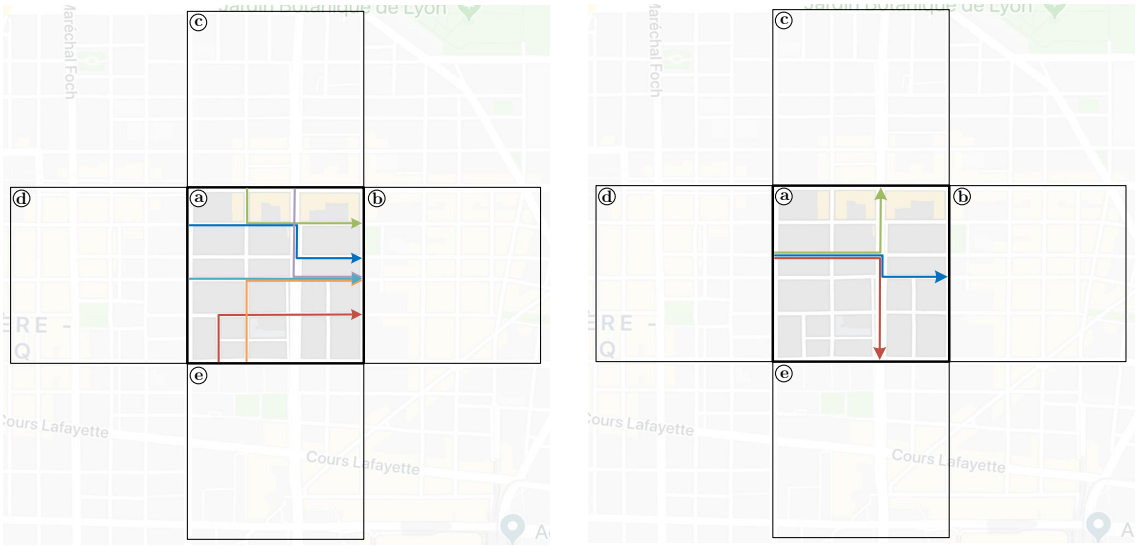

Figure 4.2: Illustration of zones and links on a real network: (i) some of the paths constituting the link from cell ⓐ to cell ⓑ, (ii) mutuality of the first portion of links from cell ⓐ to the neighboring cells (Background maps taken from Google Maps)

and the speed on link v^a can then be computed by:

$$v^a = f^a/\rho^a. \qquad \text{(link speed)} \tag{4.7}$$

There are multiple formulations of these commodity-specific demand and supply functions in the literature, in the context of intersection modeling (see e.g. Jin, 2017, and references therein). Here, we use:

$$D^a = D(\rho_{a.s}) \cdot \frac{\rho^a}{\rho_{a.s}}, \qquad \text{and} \qquad S^a = S(\rho_{a.t}) \cdot \frac{1}{N^i_{a.t}}, \tag{4.8}$$

where N^i_c is the number of incoming links from adjacent cells to cell c. Eq. (4.8) describes that demand and supply of link a are a fraction of the total demand in source cell $a.s$ and the total supply in the target cell $a.t$, respectively. The demand fraction is proportional to the number of vehicles willing to use link a, while the cell supply is assumed to be equally divided among the incoming links from the adjacent cells.

4.3.3 Solution Algorithm

Before describing the proposed two-dimensional DSO assignment method, it is necessary to state the further assumptions made in the solution method. These assumptions can be stated as follows:

a) The network is partitioned into square-shaped cells.

b) The travel distance on all links inside a cell is assumed to be identical and equal to the side length of the cell.

c) The first 50% of the links are assumed to be mutual, i.e. if a vehicle is in the first 50% of a link, it can shift to another link and move toward a different target cell, if required by the route choice strategy, as illustrated in Fig. 4.2(ii). However, to prevent erratic fluctuations, the vehicles are allowed to change their

route only once inside each cell. Note that this percentage is arbitrary and it can be modified based on the specific geometry of any cell.

As discussed earlier, at the cell level, there are only a few commodities that matter, one per each outgoing cell. Thus, our goal is to find the flow on each outgoing link in a way that satisfies the route choice strategy objective. The travel time from cell c to the destination through each outgoing link can be calculated as:

$$TT_c^{l,\tau+1}(n^{l,\tau+1}) = T^{l,\tau+1}(n^{l,\tau+1}) + TT_{l.t}^{*,\tau}, \qquad \forall\ l \in \mathcal{O}_c \tag{4.9}$$

where:

$n^{l,\tau+1}$: number of vehicles on link l at this time step $\tau + 1$,

$TT_c^{l,\tau+1}$: travel time from cell c to the destination via outgoing link l at time step $\tau + 1$,

$T^{l,\tau+1}$: travel time of link l at time step $\tau + 1$,

$TT_{l.t}^{*,\tau}$: the shortest path travel time from link l's target cell to the destination cell computed using the current instantaneous travel times, which is not a function of $n^{l,\tau+1}$,

$\mathcal{O}_c$: the set of all outgoing links of cell c.

For each link, the number of vehicles that have travelled less than 50% of their trip distance inside the current cell and have not changed their route inside this cell is denoted by $s^{l,\tau}$. These vehicles are the candidates of shifting to other outgoing links of their current cell if required by the route choice component. The number of the rest of the vehicles on the link, which will be on the same link during the next time step as well, is denoted by $m^{l,\tau}$.

The marginal cost of adding p vehicles to link l during time step $\tau+1$, $M^{l,\tau+1}(p)$, can be derived by differentiating Eq. (4.9) and expressed as:

$$M^{l,\tau+1}(p) = p\ TT_{l.s}^{l,\tau+1}(m^{l,\tau}+p) + m^{l,\tau}\ [T^{l,\tau+1}(m^{l,\tau}+p) - T^{l,\tau+1}(m^{l,\tau})], \quad (4.10)$$

where the first part of RHS is the sum of total travel cost for the added vehicles and the second part is the sum of additional travel cost incurred by the vehicles already present on the link as a result of adding p vehicles to the link.

At each simulation time step, the steps below will be followed to move the vehicles through the network toward their destination achieving the DSO conditions:

Step 1. Vehicles arriving to the network are assigned to the corresponding origin cell based on the location of their origin point and queued to enter the network during the route choice. The number of vehicles in the queue to enter cell c at time step τ is denoted by q_c^τ.

Step 2. Shortest paths from each cell to the destination cell are computed using the current link travel times.

Step 3. Number of vehicles inside each cell during time step $\tau+1$ can be found as:

$$n_c^{\tau+1} = \min\{n_c^\tau + q_c^\tau,\ n_c^{jam}\}, \quad (4.11)$$

where n_c^{jam} is the jam accumulation of cell c. Eq. (4.11) ensures that the number of vehicles inside a cell does not exceed the jam accumulation of the cell. Therefore, some of the vehicles willing to enter the network might need to wait outside the network until their origin cell gets enough capacity to let them in. The change in the accumulation of each cell, $e_n^{\tau+1}$, is expressed as:

$$e_c^{\tau+1} = n_c^{\tau+1} - n_c^\tau. \quad (4.12)$$

Step 4. Cell demands and supplies are computed based on $n_c^{\tau+1}$. Consequently, link supplies can be calculated using Eq. (4.8).

Step 5. Number of shift candidates from each link, $s^{l,\tau}$, and minimum number of vehicles staying on the same link, $m^{l,\tau}$, is calculated for all the links across the network. The total number of shift candidates inside a cell, s_c^τ, is found as:

$$s_c^\tau = \sum_{l \in \mathcal{O}_c} s^{l,\tau}. \tag{4.13}$$

Step 6. The aim of DSO assignment method is to assign $s_c^\tau + e_c^\tau$ vehicles to the outgoing links of the cell with the aim of minimizing the total travel cost for all vehicles. This can be achieved by minimizing the sum of marginal costs for all the outgoing links of each cell. At the cell level, the objective function can be written as:

$$\min \mathcal{Z}_c = \sum_{l \in \mathcal{O}_c} M^{l,\tau+1}(r^{l,\tau+1}), \tag{4.14}$$

subject to:

$$\sum_{l \in \mathcal{O}_c} r^{l,\tau+1} = s_c^\tau + e_c^\tau, \qquad \forall c, \tau \tag{4.15a}$$

$$r^{l,\tau+1} \geq 0, \qquad \forall\, l, \tau \tag{4.15b}$$

$$r^{l,\tau+1} \text{ integer}, \qquad \forall\, l, \tau \tag{4.15c}$$

where $r^{l,\tau+1}$ is the number of vehicles added to link l during time period $\tau + 1$ in addition to $m^{l,\tau}$ vehicles present on the link. Although the optimization problem may seem simple at the first glance, by incorporating Eqs. (4.10), (4.9), (4.8), (4.7) and (4.6) into the objective function one can realize that it is a mixed-integer nonlinear program (MINLP) problem.

There are several methods to solve this problem: (i) comparing the objective function at all candidate points in the set of possible solutions and finding the optimal point, (ii) using the well-known standard MINLP solvers, or (iii) relaxing the integer constraint, solving the resultant continuous sub-problem and rounding the optimum point to the closest integer point in the feasible region, which might not lead to the global optimum solution of the problem. The last two methods become significantly more efficient when feasible space is large. However, since in the present optimization problem there are only a few variables and the feasible region is relatively small, the first method can also efficiently reach the optimal solution in a very short time. Therefore, we have chosen to solve the optimization problem Eq. (4.14) using this basic method without resort to the complicated MINLP solvers.

Step 7. Solving the minimization problem Eq. (4.14) results in the optimum number of added vehicles to each link. The vehicles inside the set of shift candidates and arrivals are assigned to the outgoing links based on $r^{l,\tau+1}$ of each link. Later, vehicle positions are updated using the speed of their current link. Any vehicle completing its trip inside a cell either exits the region if it was in its destination cell or moves to the next cell on its path if the accumulation of the target cell is below its jam accumulation.

These steps are followed at each time step to achieve the DSO conditions at each cell. In the next section, the results of two simulation experiments using this algorithm will be presented and compared to the results of the RDUE assignment method developed in the previous chapter.

4.4 Numerical Experiments

This section will present and discuss the results of two simulation experiments using the proposed DSO framework. The same experiments are also done using the RDUE

solution method proposed previous chapter and the results from both approaches are compared to demonstrate the advantage of the DSO model over RDUE model in terms of congestion management.

Both experiments have been carried out on a rectangular network with a length of 5000 meters and width of 4500 meters, which is partitioned into 90 square cells with side length of 500 meters, as seen in Fig. 4.3(i). Without loss of generality and for the sake of simplicity, all cells are assumed to have the same parabolic (Greenshields') flux MFD with a jam cell accumulation of 200 vehicles. The maximum average speed inside the network is assumed to be 10 m/s (36 km/h), while the speed at the boundary of any adjacent pair of cells is computed using Eq. (4.7).

Although the model is capable of handling randomly distributed origins all across the network, to better see the performance of model under congested conditions, all the vehicles start their trips inside the network from a single origin point and travel toward a single destination point. The corresponding cells with the origin and destination locations are indicated by the green and red borders, respectively, in Fig. 4.3.

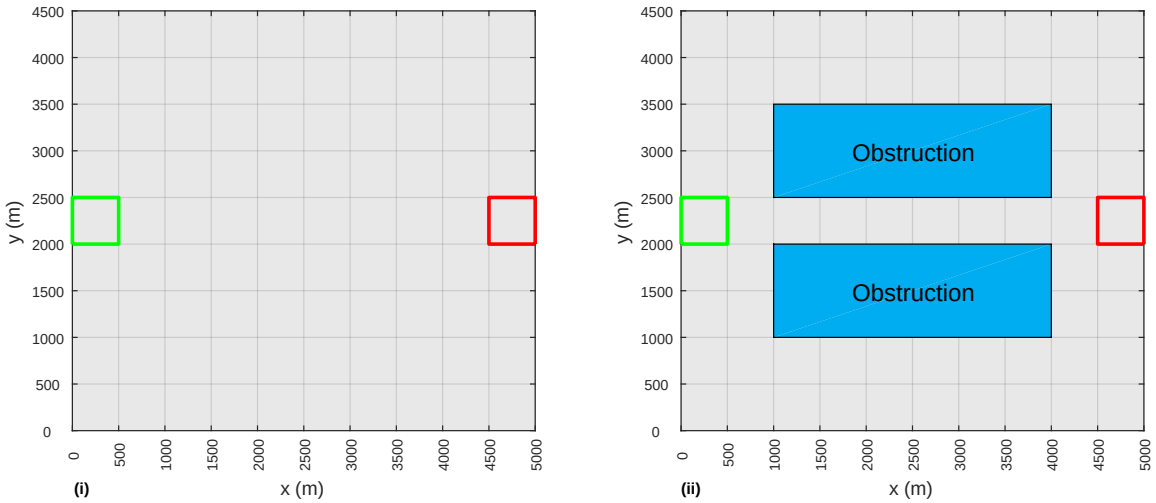

Figure 4.3: Network configurations: (i) Experiment 1, (ii) Experiment 2

4.4.1 Experiment 1: No Physical Bottlenecks

This experiment investigates the behavior of traffic in a two-dimensional domain without any physical bottlenecks, where the vehicles can circulate at any point on the network; see Fig. 4.3(i). The second experiment will analyze the performance of the proposed model on a similar network with obstructions creating a physical bottleneck in the network as seen in Fig. 4.3(ii). The network is loaded with a demand rate of 2 veh/s for 30 minutes and simulation runs till the last vehicle completes its trip inside the network by reaching the destination point. The maximum outflow rate from the region is equal to the capacity inflow rate of the destination cell, which can be computed using the parabolic flux MFD as 1 veh/s. Therefore, it is expected to see some congestion over the network, since the demand rate is higher than the maximum outflow rate.

This experiment has been also conducted in the RDUE model proposed in Chapter 3 using the same network and loading configurations. Fig. 4.4(i) shows the cumulative curves of demand, inflow to the region and outflow from the region for both DSO and RDUE models. The difference between the demand and inflow curves arises when the origin cell reaches the jam accumulation and does not have enough capacity to let the demand inside the network, as explained earlier in Step 3 of the solution algorithm.

We observe that on average the inflow rate of the RDUE model is lower than the DSO model. This is due to the fact the RDUE model overloads the initial link of the shortest path from the origin zone without considering its effect on the vehicles currently on that link that causes over-congestion of the link and consequently congestion in the link's target cell. Eventually, this results in reduction in the supplies of adjacent cells to the origin cell, reduction in the outflow rate from the origin cell and subsequently reduction in the inflow rate to the origin cell. The same phenomenon is also observed under the DSO conditions but comparing the inflow curves, we can

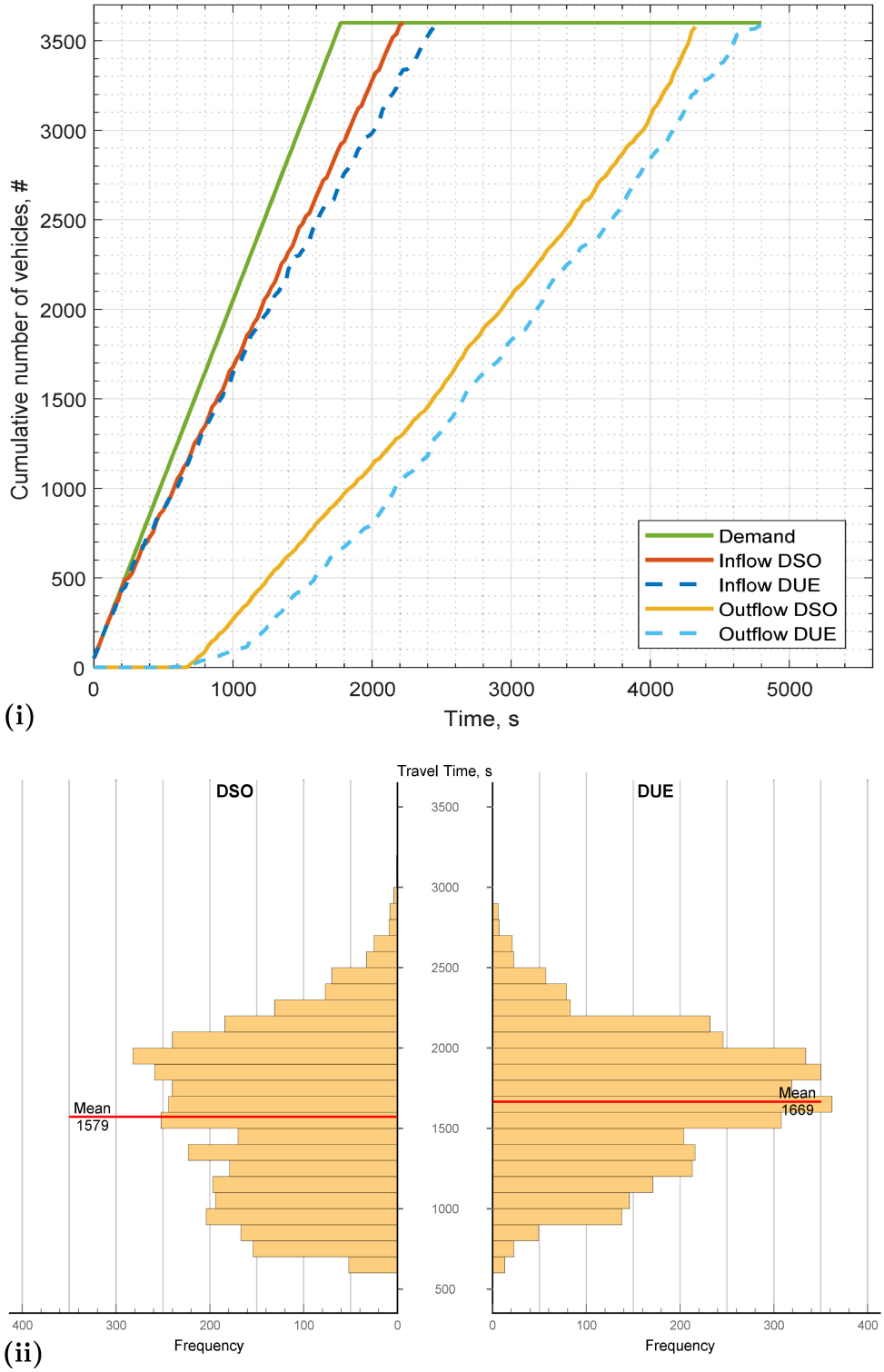

Figure 4.4: Experiment 1 results: (i) Cumulative curves, (ii) Travel time distributions

observe that the DSO model manages the over-congestion due to excessive demand better than the RDUE model.

The outflow from the region starts earlier and has a steadier rate in the DSO model compared to the RDUE model. Considering both inflow and outflow curves for two models, we can conclude that the DSO model results in less total travel time and total delay over the network compared to the RDUE model. This can be further confirmed by comparing the travel time histograms for these two models in Fig. 4.4(ii). The average travel time in the DSO model is 1579 seconds, while for the RDUE model it is 1669 seconds. Considering that the free-flow travel time from the origin to destination is 500 seconds (dividing the distance between origin and destination, 5000 meters, by the maximum average speed in the region, 10 m/s), we observe that the DSO model reduces the average delay and consequently, the total travel cost by 7.7% in this experiment.

The interesting observation in this figure is that the frequency of vehicles with shorter travel time is higher in the DSO model than the RDUE model as opposed to what we were expecting to observe based on the fact that under RDUE conditions vehicles greedily try to minimize their travel time and as a result there would be more vehicles with lower travel times compared to the DSO conditions. In contrary, the greedy behavior of drivers under the RDUE conditions to minimize their own travel time based on the perceived travel times results in the over-congestion of the shortest paths and eventually higher experienced travel times.

Fig. 4.5 exhibits the distribution of cell occupancy, $o_c = \rho_c/\rho_c^{max}$, at 3 different times for both DSO and RDUE models. The shortest path over-congestion phenomena under the RDUE conditions can be observed clearly in Fig. 4.5(b), where the adjacent cells to the origin cell are more congested compared to the DSO method at the same time. This issue imposes excessive delay on the vehicles choosing the shortest path and hinders their trip to the destination. On the other hand, the DSO model

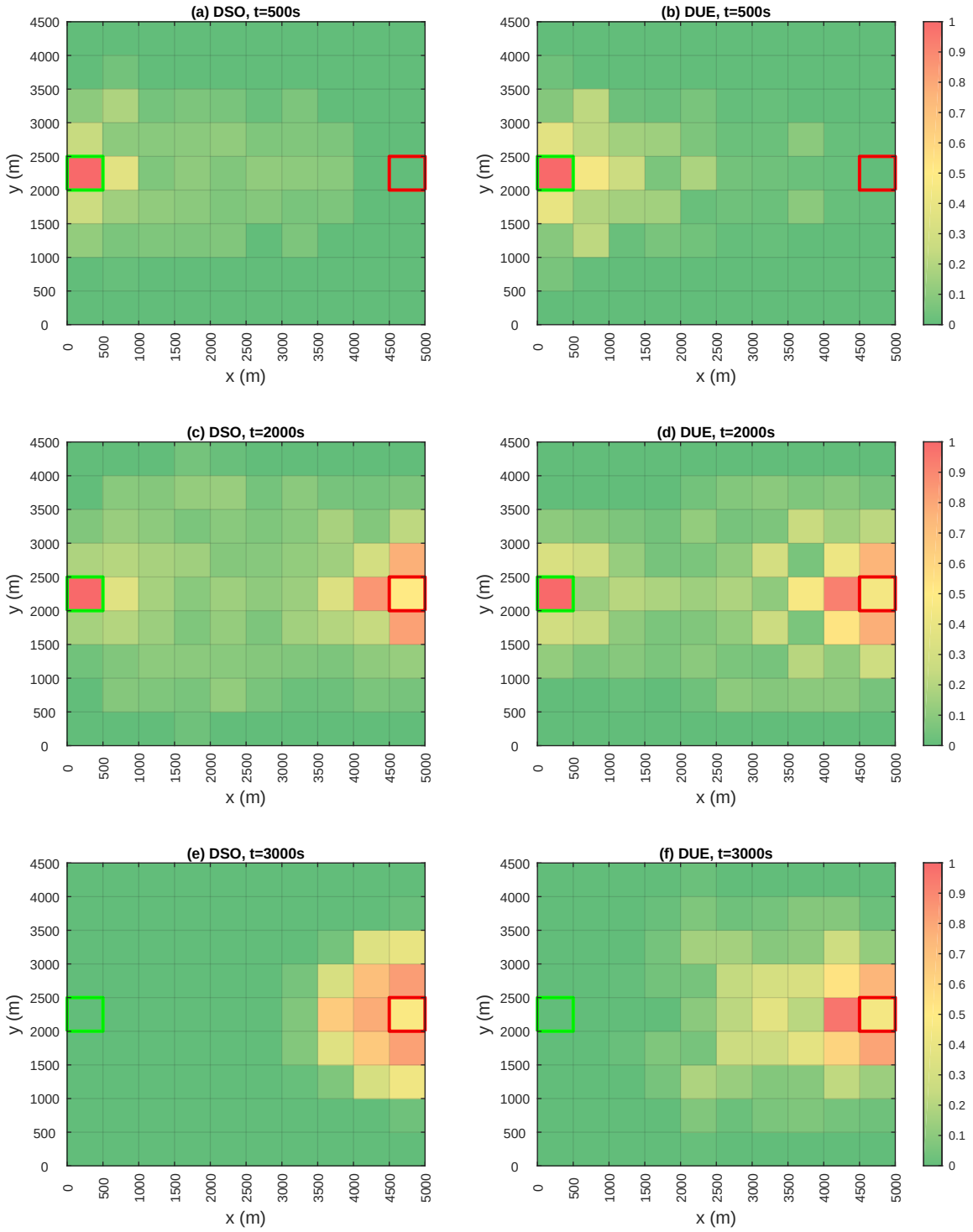

Figure 4.5: Occupancy heatmaps for experiment 1 at different time steps, Left: DSO, Right: RDUE

distributes the density more evenly throughout the analysis period, which eventually results in less average delay in the network.

4.4.2 Experiment 2: Presence of Physical Bottleneck

The second experiment is carried out on a similar network to the network in experiment 1 but in order to further highlight the advantages of the proposed DSO solution method over the RDUE method in Chapter 3, a physical bottleneck has been created in the network by placing two obstructions, which vehicles cannot exit or enter, as seen in Fig. 4.3(ii). There is an obvious shortest path in terms of travel distance from the origin to the destination, while all the other paths have significantly higher lengths. The other loading and partitioning properties are the same as experiment 1. The experiment is simulated once using the proposed DSO model in this chapter and once using the RDUE solution method in the previous chapter and the performances of these two solution methods are compared.

Fig. 4.6(i) demonstrates the cumulative demand, inflow and outflow curves for both DSO and RDUE models. It is noticeable that deviation of inflow and outflow curves in the RDUE method from the ones in the DSO model is more significant compared to experiment 1. Travel time histograms in Fig. 4.6(ii) also assert that on average the DSO model outperforms the RDUE model. The average delay for DSO and RDUE methods are 1256 and 1523 seconds, respectively, which indicates that the DSO method results in a reduction of 17.5% in average delay and total system cost compared to the RDUE method.

We can infer that the shortest path over-congestion problem in the RDUE model becomes more severe with presence of a distinct shortest path in the region. This can be confirmed by comparing the snapshots of distribution of vehicles over the network for the DSO and RDUE solutions in Fig. 4.7. At time $t = 500$, we can observe excessive congestion on the straight path from origin to destination in the

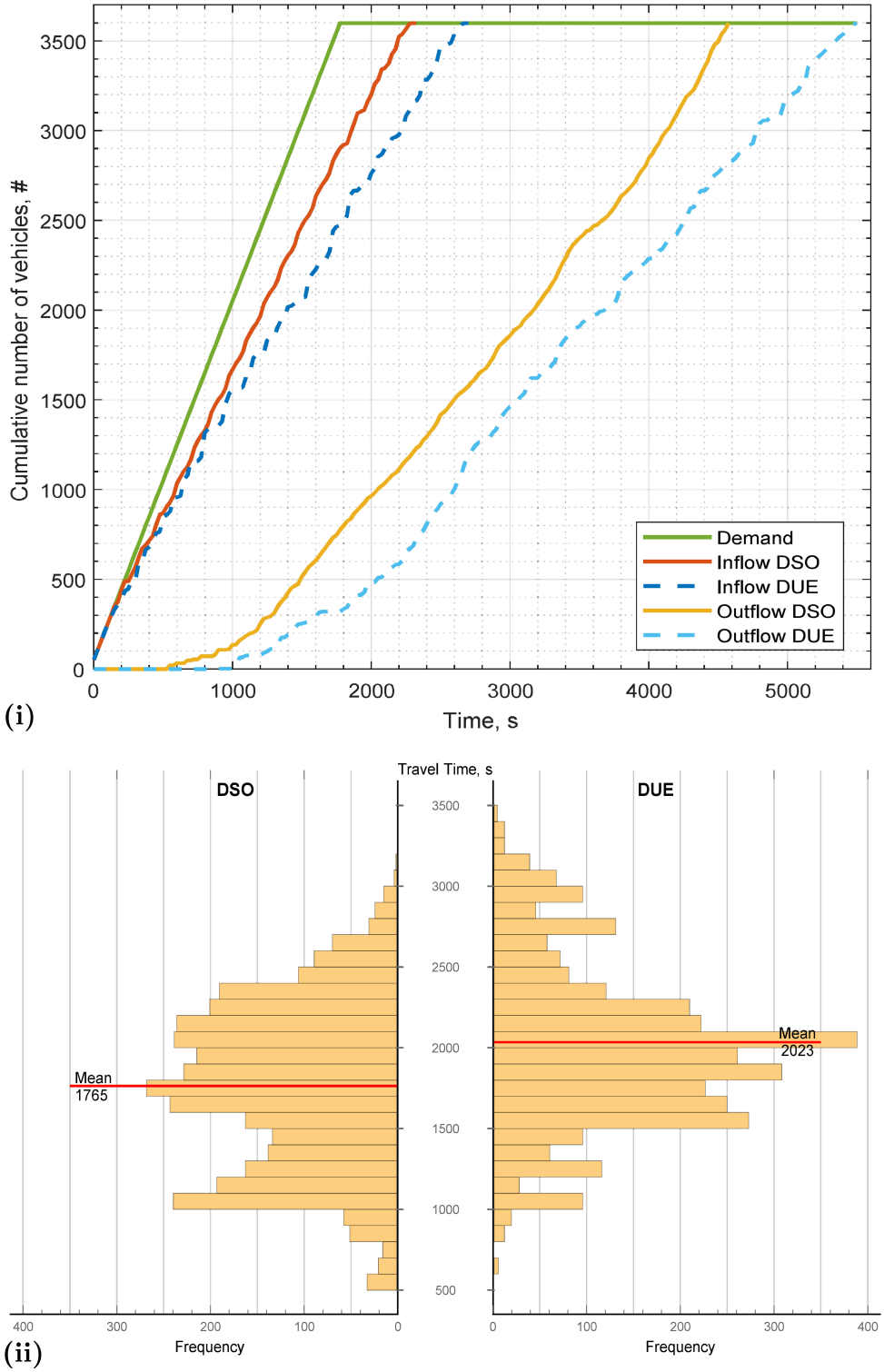

Figure 4.6: Experiment 2 results: (i) Cumulative curves, (ii) Travel time distributions

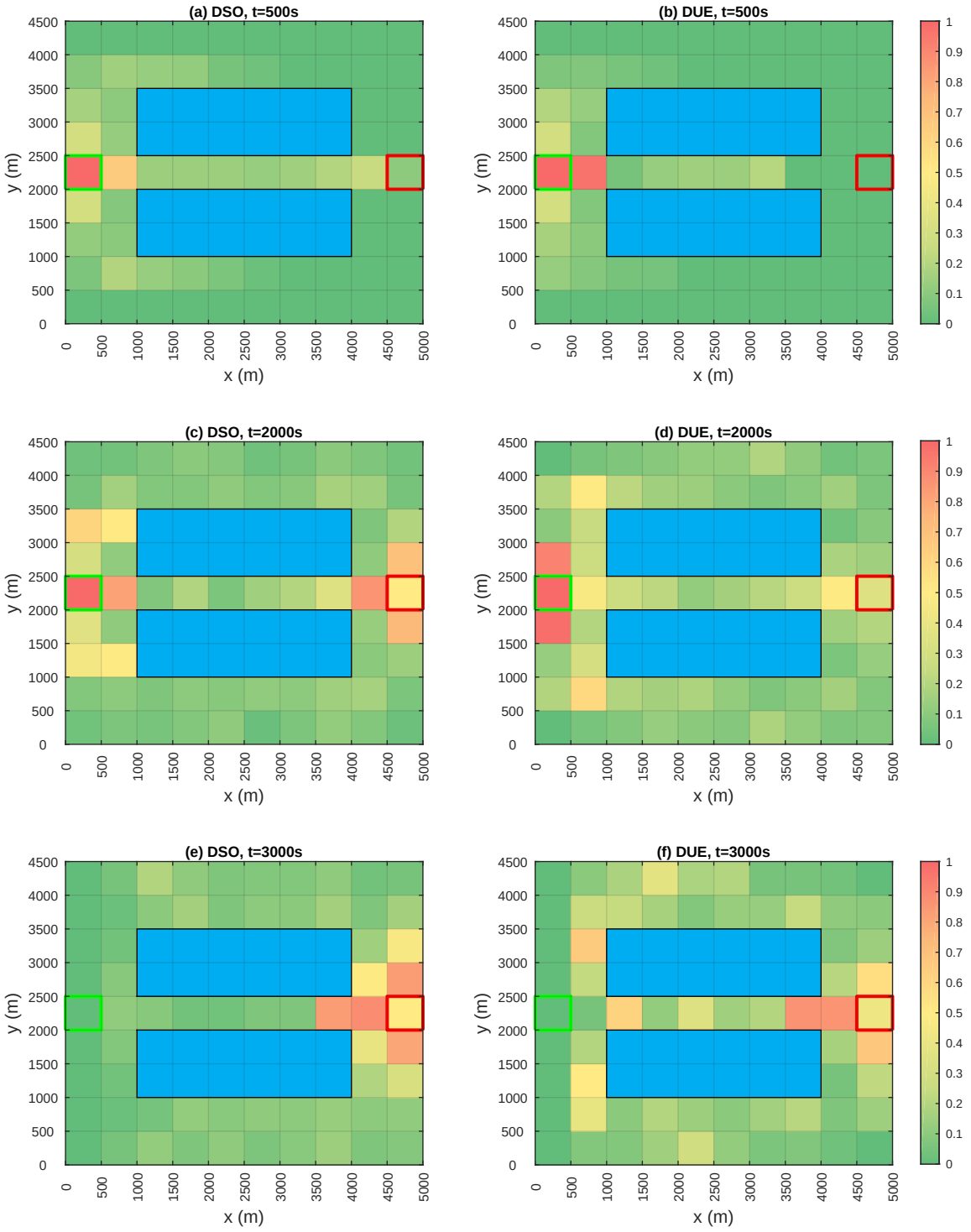

Figure 4.7: Occupancy heatmaps for experiment 2 at different time steps, Left: DSO, Right: RDUE

RDUE solution, while the density is more evenly distributed in the DSO solution by minimizing the total system cost rather than the individual travel time.

4.5 Chapter Summary

This chapter has presented a semi-DSO traffic assignment model in two-dimensional continuum space using instantaneous travel times. To the best of our knowledge, this is the first study investigating the DSO conditions for two-dimensional networks in a detailed manner. The results of two simulation experiments indicate that achieving the DSO conditions using the proposed solution method results in reduction of the total travel cost and delay in the network compared to the RDUE method in the literature.

The numerical results in Chapter 3 indicate that the accuracy of the solution method increases when the region is partitioned into smaller cells, which is also true for the DSO assignment method proposed in this chapter. The proposed solution method is based on instantaneous travel times, which might not result in the minimum experienced total travel cost. However, this is just a starting point for investigating DSO conditions in two-dimensional networks. An important future research direction is reformulating the solution method based on experienced travel time, which can solve the model to find the minimum total actual travel cost.

CHAPTER 5
INCORPORATING MACROSCOPIC FUNDAMENTAL DIAGRAM INTO EMISSIONS ESTIMATION USING THE MOVES MODEL

5.1 Introduction

The continuously increasing mobility demand in developing communities is resulting in several consequences where (i) delays caused by congestion directly result in economic loss, (ii) uncurbed fuel consumption exacerbates the depletion of non-renewable energy resources, and (iii) emissions from the vehicles provoke local (air pollution, acid rain, health issues) and global (climate change, global warming effect) damages. Therefore, from a sustainable development perspective, such effects must be studied and controlled in order to prevent any hindrance to the development process.

The motor vehicles emissions can be categorized into two general categories: (i) the air pollutants such as Particular Matter (PM), Volatile Organic Compounds (VOCs), Nitrogen Oxides (NO_x), Carbon Monoxide (CO) and Sulfur Dioxide (SO_2), which impact the air quality and can result in human health issues, and (ii) Greenhouse Gases (GHG) including Carbon Dioxide (CO_2), Methane (CH_4) and Nitrous Oxide (N_2O), which have broader impacts such as climate change and global warming. Although the predominant GHG emitted from motor vehicles is CO_2, the latter two GHGs have higher Global Warming Potential (GWP) despite their lower emission rates (Azar and Johansson, 2012).

According to a recent fact sheet published by the United States Environmental Protection Agency (EPA), transportation sector accounted for the largest portion (28%) of total U.S. GHG emissions in 2016 (EPA, 2018). Between 1990 and 2016, GHG emissions in the transportation sector has increased more in absolute terms

than any other sector (i.e. electricity generation, industry, agriculture, residential, or commercial) due in large part to increased demand for travel.

Several national and international policies including 1990 Clean Air Act Amendment, Kyoto Protocol and Paris Agreement have been proposed to control the GHG and pollutants emissions. The kernel technical basis of all these policies is the requirement of reliable emissions data, which is estimated using emission models. Emission models can be categorized as microscopic and macroscopic models (Rakha et al., 2003). Macroscopic models such as EPA's MOBILE and California Air Resources Board's EMFAC models use average aggregate network parameters to estimate network-wide energy consumption and emission factors. The major drawback of these macroscopic models is in the use of a single traffic-related variable to estimate emissions, thereby ignoring other important explanatory variables that can significantly affect emission estimates.

The project level of EPA's state-of-the-science emissions model, MOVES (MOtor Vehicle Emission Simulator; EPA, 2020a), is a microscopic model that incorporates the effects of different instantaneous speed and acceleration profiles on vehicle emissions, thereby representing real driving conditions. The latest approved version of MOVES is required by the regulations to be used in all transportation and air quality planning and assessment work. However, the complicated interface of MOVES, the need to comprehensive data and analysis for each link of the network and exhaustive run-time make it impractical for real-time analysis and control purposes on large-scale networks.

Up until recently, the main objective of most traffic control strategies has been to alleviate congestion and to reduce the delay incurred by travelers and emission and fuel consumption control have only gained interest in the past few years. However, the optimal emission and fuel consumption strategy may not lead to minimal delay and congestion (Csikós and Varga, 2011). Therefore, the optimal control strategy

considering all aspects must be a trade-off between optimum emissions and optimum delay conditions.

Estimation of emissions and energy consumption using MOVES or other microscopic models is resource-intensive for large-scale networks and intractable for real-time analysis and control purposes. Several efforts have lately been made to improve the efficiency of MOVES model, e.g. MOVES Lite (Liu and Frey, 2013) and MOVES-Matrix (Liu et al., 2016), but still the need to link-level computations is a hurdle to the real-time emission estimation for traffic control purposes.

The idea of an aggregated relationship between the network-wide traffic variables dates back to 1960s (Smeed, 1967; Godfrey, 1969) and a few other studies in 1980s (Herman and Prigogine, 1979; Mahmassani et al., 1984; Ardekani and Herman, 1987). However, the recent empirical verification of existence of a network-level Macroscopic Fundamental Diagram (MFD) on urban areas has opened up a new paradigm (Geroliminis and Daganzo, 2008).

The MFD gives average traffic variables as a function of the number of vehicles inside the region, n. For instance, the average flow MFD:

$$Q = Q(n), \qquad \text{(average flow MFD)} \tag{5.1}$$

gives the average flow, Q, on network arguably independently of trip origins and destinations, and route choice. This makes the MFD an invaluable tool to overcome the difficulties of the traditional link-level planning and control models including the microscopic emissions estimation models. If the MFD is accurately derived using empirical data or estimated using the analytical models on a region with homogeneous traffic distribution, it can be used to replace the links inside the region with a single entity, whose traffic dynamics is given by the MFD variables, in the emissions estimation method without significant loss of accuracy. The MFD has been widely used in

other aspects of traffic planning and management such as dynamic traffic assignment (Yildirimoglu et al., 2015; Batista et al., 2019; Aghamohammadi and Laval, 2020a; Aghamohammadi and Laval, 2020b), optimal control (Ramezani et al., 2015; Haddad, 2017; Kouvelas et al., 2017), parking (Cao and Menendez, 2015; Leclercq et al., 2017b) and pricing (Zheng et al., 2016; Yang et al., 2019).

Nevertheless, the implementation of MFD in emissions estimation is yet an underdeveloped field of research and practice, with only a handful of studies conducted on this topic up until now. Furthermore, since emissions analysis is required to be done according to certain regulations or using specific models by the national or local agencies, many authorities are limited to use the specific models and software packages for the emissions estimation purposes. Therefore, researchers tend to explore innovative ways to incorporate the state-of-the-art developments in traffic modeling realm into the existing emissions estimation methods to improve their efficiency.

In this chapter, we will first conduct an extensive review of the existing literature on implementation of the MFD for emissions estimation purposes in the recent years. Of foremost importance is the studies which try to incorporate the MFD dynamics into the existing well-established emissions estimation models mandated by the authorities to be used for emissions analysis. Later, a numerical experiment of estimating the emissions of a grid network in different aggregation levels using the MOVES model will be presented and the efficiency and accuracy of the results will be discussed.

5.2 Background

Considering the monotonous increase of the share of transportation industry in the pollutants and GHG emissions, there will be a need to develop traffic control and routing strategies in the near future not only based on the travel time minimization criterion, but also seeking to optimize the pollutants and GHG emissions and

energy consumption, which would require accurate real-time emissions estimations. In general, the emissions estimation models can be categorized into three different approaches based on their need to different levels of data (Yin et al., 2013):

I) The *emission factor models*, which assign an emission factor derived from repetitive experiments to the whole analysis period. The emission factor can be multiplied by correction factors to incorporate different vehicle speeds, fuel type, vehicle age, etc. The total amount of emission for any specific pollutant or GHG is found by multiplying the corresponding emission factor by the total vehicle miles travelled (VMT) for the analysis period.

II) The *physical power-demand models*, which take into account the vehicle's operational conditions and driving environment to provide second-by-second emissions. This models usually consider different operational characteristics of vehicles such as engine power, engine speed, air/fuel ratio, fuel use, etc. for different vehicle operation states. These type of models estimate the emissions more accurately, however, their extensive need to obtain a large variety of detailed data is an obstacle in their application.

III) The *acceleration and speed-based models*, which compute emission rates as nonlinear functions of instantaneous speed and acceleration values. This approach has gained more popularity recently, since they provide more accurate estimates compared to the emission factor models and require less information compared to physical power-demand models.

Nevertheless, the current microscopic methods of emission estimation such as the MOVES model are unsuitable for real-time applications in large-scale networks with thousands of links despite their higher accuracy compared to the previous models. In this section, we will perform an extensive review of the state of the art in incorporating MFD in vehicular emissions estimation. It is worth mentioning that in order to keep

the consistency of the notation throughout this chapter, we have changed the notation of some of the studies reviewed here. Moreover, a **bold-typed reference** in the next section indicates the outset of its review.

5.2.1 Reservoir-based Models

Most of the MFD-based emissions estimation models in the literature model the entire network as a reservoir, whose traffic dynamics are given by the MFD of the reservoir. The network-wide aggregated MFD variables are then used to estimate the vehicular emissions inside the network.

Shabihkhani and Gonzales (2014) developed a methodology for GHG emissions estimation using the MFD of a single homogeneous network for estimating driving cycle components, i.e. (i) time spent cruising at free-flow speed, T_c, (ii) time spent idling while stopped, T_i, and (iii) the number of vehicle stops per distance traveled, n, without the need for extensive trajectory analysis using conventional microscopic methods. Then, the total emissions per vehicle distance traveled, E, is computed as:

$$E = e_c T_c + e_i T_i + e_s n, \qquad \text{(total emissions per distance travelled)} \quad (5.2)$$

where e_c is the emission factor per unit cruising time, e_i is the emission factor per unit idling time and e_s is the emission factor associated with complete cycle of deceleration from the free-flow speed, v_f, to 0 and then acceleration from 0 to v_f. In order to estimate the emission factors, a traffic simulation has been conducted on a simple ring network with a single intersection representing a long arterial or a network with homogeneous traffic conditions. The ring network has been loaded with the full range of possible densities from an empty network to complete jam. The vehicle trajectories are extracted to be used for microscopic emissions estimation via the project level of

the MOVES model as a comparison benchmark to compute the emission factors and the duration of a full deceleration and acceleration cycle, τ.

The next step is to approximate the aforementioned driving cycle components using the MFD. It has been assumed that the vehicles stop only once during each traffic signal cycle, C. The driving cycle components for each traffic state, with average MFD speed v are estimated as:

$$n = \frac{1}{vC}, \qquad \text{(number of stops per unit distance travelled)} \tag{5.3a}$$

$$T_c = \frac{1}{v_f} - \frac{\tau}{2}n, \qquad \text{(cruising time)} \tag{5.3b}$$

$$T_i = \frac{1}{v} - \frac{1}{v_f} - \frac{\tau}{2}n. \qquad \text{(idling time)} \tag{5.3c}$$

Later, sensitivity analysis has been done to see the impact of 3 main factors, green time to cycle length ratio, signal cycle length and block length, on GHG emissions estimates and their relative error compared to the benchmark emissions rate computed using the MOVES model. It has been found that the green time to cycle length ratio plays a key role in both the shape of the MFD and the rate of emissions and the estimation errors are significant and consistently positive at near-to-jam density values. A significant shortcoming of the proposed model is not being capable of taking different vehicle types into account.

Csikós et al. (2015) developed an optimal perimeter control model aiming to minimize the emissions from the vehicles inside the perimeter built on the emissions estimation model proposed in Csikós and Varga (2012) using the average speed (v), total travel distance (TTD) and total time spent (TTS) variables obtained using the MFD. Emission factors for different traffic states for the network are calculated similar to the link-wise emission factors formulation provided by Ntziachristos et al. (2009) using the MFD average speed, which has later been multiplied by the TTD of the network to find the total emissions of the network. The CO emission estimates for two

different case studies indicate that the proposed method has a 20% error compared to the link-level microscopic Versit+Micro model (Smit et al., 2007) but yields in very similar results compared to the link-level Copert IV model (Ntziachristos et al., 2009).

Amirgholy et al. (2017) proposes a model for optimal design of public transportation systems in congested urban networks. As a component of the total cost of the transportation system, the external cost of emissions is approximated by using the average speed of the network MFD in a VMT-based emissions estimation model proposed by Affum et al. (2003). In this model, the external cost of emissions (E_N) is computed as a function of the fuel consumption of automobiles (F_A) and transit vehicles (F_T), where the MFD speed for any given average density state in the network is only used for computation of the fuel consumption of the average automobile per unit distance traveled in the network.

Ingole et al. (2020) proposed an optimal perimeter control model for a network comprised of an inner city modeled as a reservoir with entrances and exits at three locations on the perimeter and bypass freeways connecting and exit points. The model aims to minimize the NO_x emissions inside the reservoir using a gating strategy based on Nonlinear Model Predictive Control (NMPC). The reservoir emissions are computed by integrating the fourth-degree polynomial formulation proposed by Lejri et al. (2018) into the COPERT IV model (Ntziachristos et al., 2009). The total emissions of pollutant $m \in \{CO_2, NO_x\}$ between t and $t + dt$, $E^m_{dt}(t)$, in [g] is found as:

$$E^m_{dt}(t) = EF^m(v(t)) \times n(t) \times v(t) \times dt, \tag{5.4}$$

where $EF^m(v(t))$ is the emission factor of pollutant m in [g/km], $n(t)$ is the accumulation of vehicles inside the reservoir at time t, $v(t)$ is the MFD mean speed at time t

in [km/h] and dt is the time step in [h]. The emission factors are calibrated through curve fitting to the actual microscopic emissions for different MFD mean speeds. The emissions of each internal route i inside the reservoir is found by replacing n in Eq. (5.4) with the partial accumulation of the route n_i. Furthermore, the emissions of each bypass link can be calculated by replacing the accumulation and average speed of the link in Eq. (5.4).

Saedi et al. (2020) developed a model to estimate the network-wide emissions by incorporating the MFD into the microscopic emission model considering different light and heavy-duty vehicle compositions. The benchmark microscopic emission rates are obtained using the polynomial model proposed in Panis et al. (2006) as:

$$E_n^m(t) = \max\left[0.0, (f_1)_n^m + (f_2)_n^m\, v_n + (f_3)_n^m\, v_n^2 + (f_4)_n^m\, a_n + (f_5)_n^m\, a_n^2 + (f_6)_n^m\, v_n a_n\right], \tag{5.5}$$

where $n \in \{\text{gasoline car, diesel car, LPG car, heavy-duty vehicles}\}$ is the vehicle type, $m \in \{CO_2, NO_x\}$ represents different emission types, v_n is the vehicle's speed in [m/s], a_n is the vehicle's acceleration in [m/s^2] and $(f_1)_n^m$ to $(f_6)_n^m$ are model constants determined by non-linear multiple regression methods. For the macroscopic emissions estimation the authors propose:

$$E_m(t) = \left(\sum_{i=1}^{N} \alpha_i^m p_i\right) k(t)\,(\beta_m + v(t)) = \left(\sum_{i=1}^{N} \alpha_i^m p_i\right)(\beta_m k(t) + q(t)), \tag{5.6}$$

where $E_m(t)$ is the rate of emission in [g/s] of pollutant $m \in \{CO_2, NO_x\}$ at time step t in the observation period, p_i is the penetration rate of vehicle type i in the traffic stream, $k(t)$, $v(t)$, and $q(t)$ are network-wide average density, speed, and flow, respectively, and α_i^m and β_m are the model parameters for pollutant m. β_m can be

perceived as a penalty factor for high density scenarios, when the traffic is congested and repeated stop and go movements have a significant toll on the emissions.

A micro-simulation is used to produce traffic data for different demand and vehicle composition scenarios in order to calibrate the proposed macroscopic emission model parameters using the microscopic emission estimates found by Eq. (5.5). The significant difference between the calibrated parameters for a congested central business district (CBD) inside the network and the entire network, which is less congested on average, demonstrate that the model parameters are very sensitive to the network topology and demand intensity and that the parameters have to be calibrated for each network. The results indicate that different vehicle compositions only have a scaling factor on the resulting total emissions. Using the results, 3-dimensional Network Emission Diagrams (NED) are developed showing the emission rate for any average flow and density pair for the studied network.

The next subsection will review another class of MFD-based emissions estimation models, which do not explicitly state the incorporation of MFD in their model but in fact are using MFD for their local traffic flow relationships similar to the continuum-space DTA models presented earlier in Chapter 2.

5.2.2 Continuum-space Models

Another major approach observed in the literature is to model the network as a continuum space where the vehicles can circulate at any point $\mathbf{x} = (x, y)$ of a Euclidian two-dimensional domain $\Omega \subset \mathbb{R}^2$, see Fig 5.1(a). The basic notion behind the continuum-space traffic models is that when the network is dense enough such that the distances between road segments are small compared to the size of network, the network can be approximated by a continuum space. There is a vast body of literature on continuum-space Dynamic Traffic Assignment (DTA), emanating from the seminal pedestrian flow model by Hughes (2002). Later, researchers have adapted this frame-

work to the vehicular traffic and have proposed various DTA models in continuum space (see e.g., Jiang et al., 2011; Du et al., 2013; Lin et al., 2017), which consist of a conservation law partial differential equation (PDE) in two dimensions, supplemented with an Eikonal or Hamilton-Jacobi PDE for the route choice component.

The connection between the continuum-space models and the MFD theory lies in the numerical solution of these models, where the continuum space is partitioned into a grid of small regions, see Fig. 5.1(b), and the traffic dynamics inside each region is described by a local speed-density relationship. A number of recent studies in the continuum-space DTA literature have mentioned the local speed-density relationship as MFD (Du et al., 2013; Du et al., 2015; Long et al., 2017). Aghamohammadi and Laval (2020b) shows that the speed-density relationship can in fact be interpreted as the MFD, since it is defined on a portion of the network and has to also take the network effects such as signal timing into account. Note that the local MFDs in the continuum-space models are defined only on a small portion of the network, while in the reservoir-based models there is a single MFD defined over the entire network.

As the first study investigating the emissions in the continuum space, **Yin et al. (2013)** proposes a bi-level optimization problem to optimize housing allocation to minimize vehicular emissions in an integrated land use and transportation modeling framework. The lower-level subprogram formulates the traffic assignment problem to

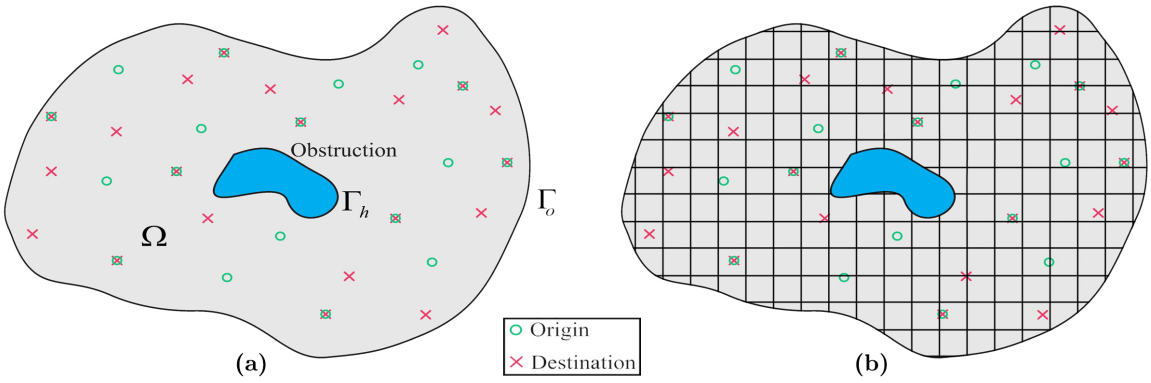

Figure 5.1: Illustration of (a) continuum space and (b) typical solution grid

achieve the user optimum (UE) solution, where the total cost for each user is consisted of the travel and housing costs, and solves the resulting system of PDEs using the finite element method (FEM). For a complete review of formulation and different solution methods of continuum-space traffic assignment models refer to Chapter 2.

In the upper-level subprogram, the housing allocation is optimized to minimize the CO_2 emissions. The emissions and fuel consumption estimation model proposed by Ahn et al. (2002) is adopted to estimate the emission rate of each type of emission as:

$$E^m(v,a) = exp\left(\sum_{i=0}^{3}\sum_{j=0}^{3}\omega_{i,j}^m v^i a^j\right), \tag{5.7}$$

where E^m is the instantaneous emission rate and fuel consumption with superscript m denoting different kind of emissions, HC and CO in [mg/s] and fuel consumption in [gal/h], v is the speed in [km/h], a is the acceleration in [km/h^2] and $\omega_{i,j}^m$ denotes the model regression coefficient for emission type m, speed power i and acceleration power j. However, this model cannot be directly used to estimate the CO_2 emissions and the CO_2 emission rate can be found using the carbon balance between fuel consumption and emissions of other gases including carbon as:

$$E^{CO_2}(v,a) = 2458.29F(v,a) - 3.17E^{HC}(v,a) - 1.57E^{CO}(v,a), \tag{5.8}$$

where F, E^{HC} and E^{CO} are Fuel consumption in [gal/h], HC emission rate in [mg/s] and CO emission rate [mg/s], respectively, which all can be computed using Eq. (5.7) using the available model coefficients. The total CO_2 emissions in the network are found by integrating the CO_2 emission rate multiplied by the norm of the optimum flow vector at any point $\mathbf{f}^*(x,y)$ over the entire network considering that the speed and acceleration values at the vicinity of any point is given by the local MFDs described earlier.

Jiang et al. (2018) proposes a second-order continuum-space DTA model, where the travelers seek to minimize their travel cost based on the dynamic user equilibrium (DUE) principle considering that the travelers only perceive their instantaneous travel times as the travel cost. After solving the DTA problem using the standard PDE solution methods, the instantaneous emission rate $E_n^m(\mathbf{x}, t)$ in [g/s/veh] for each vehicle type n and emission type $m \in \{CO_2, NO_x, PM, VOC\}$ at any location $\mathbf{x} \in \Omega$ and time t is found using the Panis et al. (2006) model given by Eq. (5.5). The concentration of the emission for the nth vehicle type and the mth emission type at location $\mathbf{x} \in \Omega$ and time t is expressed as $DE_n^m(\mathbf{x}, t)$ in [kg/km^2/h] and is calculated as:

$$DE_n^m(\mathbf{x}, t) = 3.6 \times \rho_n(\mathbf{x}, t) E_n^m(\mathbf{x}, t), \qquad \text{(emission concentration)} \quad (5.9)$$

where $\rho_n(\mathbf{x}, t)$ is the density of nth vehicle type at location $\mathbf{x}$ and time t. The total emission rate for the nth vehicle type and the mth emission type in the entire continuum space at time t is denoted as $TE_n^m(t)$ in [kg/h] and is obtained as:

$$TE_n^m(t) = \iint_{\Omega} DE_n^m(x, y, t) dx dy. \qquad \text{(total emission rate)} \quad (5.10)$$

The corresponding cumulative emission for the nth vehicle type and the mth emission type in the entire network since the beginning of analysis till time t, $CE_n^m(t)$ in [kg], can be found as:

$$CE_n^m(t) = \int_0^t TE_n^m(s) ds. \qquad \text{(cumulative emission)} \quad (5.11)$$

5.3 Discussion and Outlook

In this section, we discuss the main findings in terms of the observed trends in the literature, and identify the issues and challenges that deserve further research. Table 5.1 presents an overview of the studied papers, their specifications and main contributions.

5.3.1 From Single-reservoir to Multi-reservoir Models

The studies reviewed here demonstrate that the MFD can be a powerful tool in making the microscopic emissions estimation models less resource intensive and more efficient for real-time network control and management purposes. However, due to the heterogeneity of realistic large-scale networks, the single-reservoir models presented here cannot accurately estimate the emissions in large-scale networks. The current trend of dealing with heterogeneity in large-scale networks is to partition the

Table 5.1: Overview of studied papers

Source	Estimated Emissions	MFD Variables	Base Model	Main Contribution
Reservoir-based Models				
Shabihkhani and Gonzales (2014)	GHG	v	MOVES (EPA, 2020a)	Computed driving cycle components to be used in the MOVES model
Csikós et al. (2015)	CO	v, TTD, TTS	COPERT IV (Ntziachristos et al., 2009)	Replaced the MFD-based variables for network into the link-level COPERT IV model
Amirgholy et al. (2017)	External cost of emission	v	Affum et al. (2003)	Proposed optimal transit system design by incorporating external emissions cost in the total cost
Ingole et al. (2020)	CO_2, NO_x	v, n	COPERT IV (Ntziachristos et al., 2009)	Proposed optimal perimeter control minimizing the emissions inside the reservoir
Saedi et al. (2020)	CO_2, NO_x	v, q, k	Panis et al. (2006)	Developed the most detailed MFD-based emissions model in the literature using mean MFD v, q and k
Continuum-space Models				
Yin et al. (2013)	CO_2	v, a	Ahn et al. (2002)	Developed a housing allocation optimization model by minimizing the CO_2 emissions in the network
Jiang et al. (2018)	CO_2, NO_x, PM, VOC	v, a	Panis et al. (2006)	Developed a comprehensive method to estimate emissions in continuum space

network into smaller regions with homogeneous traffic conditions, which will exhibit accurate low-scatter MFDs. This approach has been recently very popular among the researchers for network control and traffic assignment purposes (Haddad, 2017; Yildirimoglu et al., 2015; Aghamohammadi and Laval, 2020a).

In a similar way, for emissions estimation purposes the large-scale networks can be divided into smaller regions in order to estimate the emissions inside the network more accurately. The size of regions can be determined based on the physical reality of the networks or in a way to make a balance between the accuracy and efficiency of the model. The smaller the regions inside the network, the more accurate the emissions estimations will be but on the other hand the higher number of regions would increase the computation times and reduce the model efficiency.

After appropriately partitioning the network into smaller regions, the same single-reservoir emissions estimation methods presented here can be applied to estimate the emissions from each region, which would sum up to the emissions from the entire network. Furthermore, the knowledge about emissions at each part of the network will pave the path to develop traffic control and routing strategies aiming to minimize the emissions inside the network.

5.3.2 Challenges in Estimation of the MFD

Although the proposed multi-reservoir approach sounds promising, deriving the MFDs for each and every region inside the network seems to be a hurdle in its implementation. The MFD can be empirically derived using empirical data such as loop detector data (LDD). However, the required empirical data is not available for many networks and even when available it is subject to significant errors and bias. Another approach is to determine the MFD analytically, which does not require empirical data and estimates the shape of the MFD using network topology and control

characteristics such as block length, existence of turn-only lanes, and traffic signal settings.

Up until recently, estimating the MFD was not an easy task because the Method of Cuts (MoC) in Daganzo and Geroliminis (2008) becomes intractable for real-life networks and one needs to resort to simulation methods, which defeats the purpose of macroscopic modeling. Laval and Castrillón (2015) develops the Stochastic Method of Cuts (SMoC) by extending the MoC to stochastic corridors with different block lengths and signal timing settings and shows that (the probability distribution of) the MFD can be well approximated by a function of mainly two parameters: *the density of traffic signals* (λ) and *the mean red to green ratio of traffic signals* (μ). Chapter 6 of this dissertation will present an extension of the SMoC to general networks and and develop a statistical inference method to estimate the shape of the MFD and network parameters using available limited empirical data.

Adopting the analytical methods discussed above or other analytical MFD estimation methods in the literature can help to overcome the obstacle of determining the MFDs of regions in the multi-reservoir approach by only needing some information about the network topology and control characteristics. Moreover, if the data for analytical methods are not easy to obtain, one can resort to micro-simulation methods to derive the MFD, which would need precise calibration and validation to represent the ground truth accurately.

5.4 Incorporating the MFD in the MOVES model

Out of the studied papers, only one study builds an MFD-based emissions estimation model based on the MOVES model (EPA, 2020a) required by the US EPA for emissions analysis. The European researchers tend to utilize the COPERT IV model (Ntziachristos et al., 2009), which is developed and required by the European Environment Agency. Therefore, it would be better if the researchers in the US put more

emphasis on improving the efficiency of the MOVES model by incorporating recent advances in traffic modeling domain such as the MFD.

The project level of the MOVES model allows the traffic data to be included in the model via three different methods: (i) the average link speed for the analysis period, (ii) the second-by-second link drive schedule demonstrating instantaneous link speeds, or (iii) the operating mode distribution for all links in the network. The derivation of the second set of data might be unattainable from the MFD only, unless we have information about the evolution of the MFD over time. The sequential MFD data can help us to derive the driving cycle components more accurately compared to the method presented in Shabihkhani and Gonzales (2014), which uses only the mean MFD speed in order to find the driving cycle components. Furthermore, the variance of MFD at any given density value might be helpful in determining the driving cycle components, which needs further investigation.

On the other hand, the average link speed input method provides a straightforward method to incorporate MFD variables in the MOVES model. Following the aforementioned recommendations, after dividing the network into several regions with homogeneous traffic conditions, all the links inside each region can be replaced by a single entity with its average speed given by the mean MFD speed. However, case-specific correction factors may be needed to estimate the emissions accurately by this approach, which can be found by calibrating the results using the results of link-level emissions estimates.

The next section will present the results from a numerical experiment incorporating the aggregated traffic variables in the MOVES model and will compare the efficiency and accuracy attained by 4 different aggregation levels of network representation.

5.5 Numerical Experiment

The project scale is the most in-detail module of the MOVES model requiring detailed inputs describing the vehicle population and activity at the site. This module is capable of estimating the emissions of different pollutants emitted from a transportation network for a one-hour period in a specific month of the year at a specific county given information on network geometry, vehicle type, fuel type and age distribution of the fleet, meteorology data, and any ongoing maintenance programs at the project location.

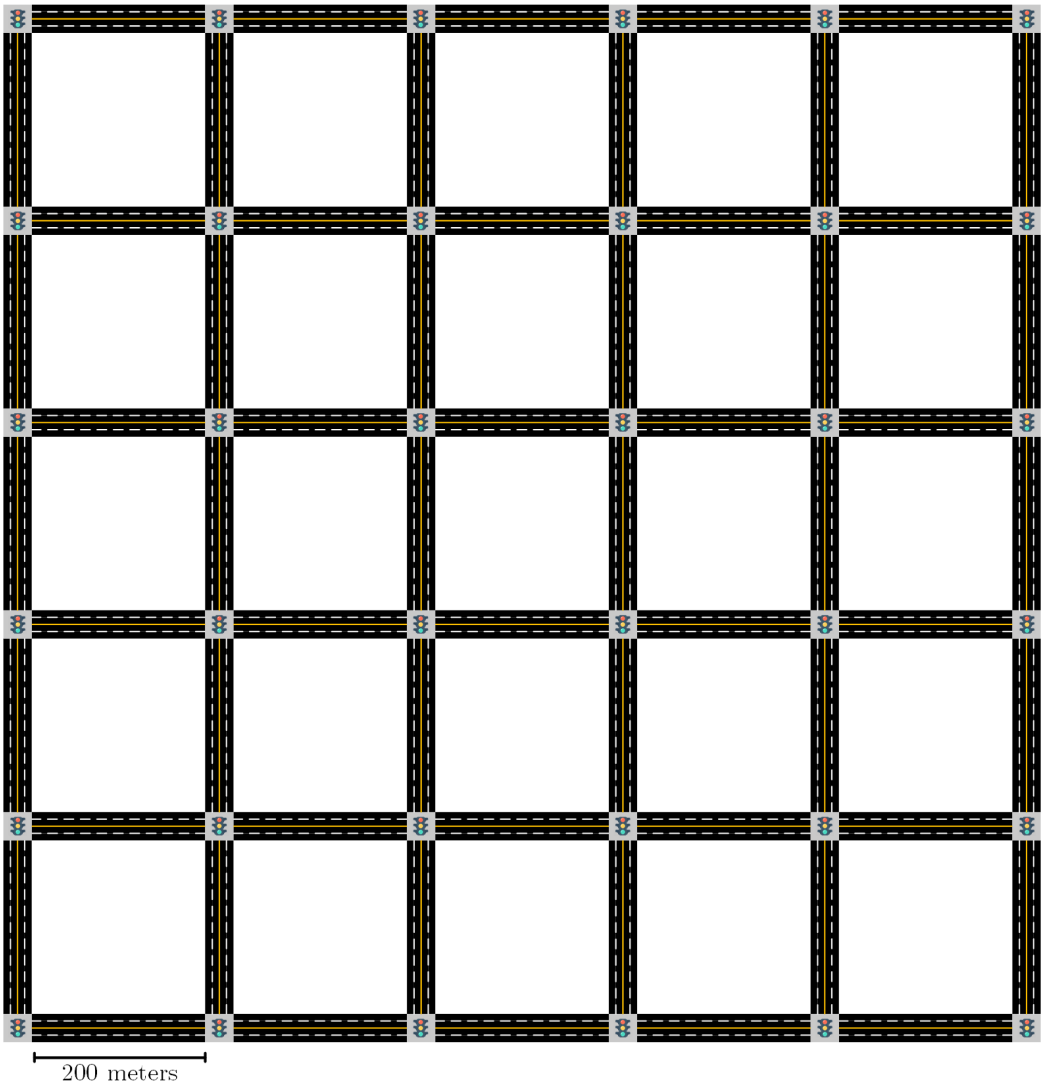

Figure 5.2: Illustration of the 5-by-5 grid network

The numerical experiment here aims to demonstrate the potential of implementing aggregated-level representation of a network in the MOVES model by comparing the accuracy and efficiency of emissions estimations conducted in 4 different aggregation levels: (i) lane, (ii) link, (iii) corridor, and (iv) network. Toward this purpose, a 5-by-5 homogeneous grid network with identical block lengths of 200 meters, 2 lanes in each direction and traffic signals with a cycle length of 90 seconds at all intersections, as shown in Fig. 5.2, has been loaded with a constant demand rate of 5 vehicles per second for 2100 seconds with randomly-distributed origins and destinations across the network using the SUMO traffic simulation package (Lopez et al., 2018). The evolution of departures, arrivals, accumulation and network average speed during the analysis period is demonstrated in Fig. 5.3. The accumulation curve, indicating the number of running vehicles on the network, can be derived as the difference between the cumulative departures and arrivals curves at any time.

Fig. 5.4 exhibits the average speed versus accumulation and average flow versus average density MFD plots for the network. The two different branches in each of

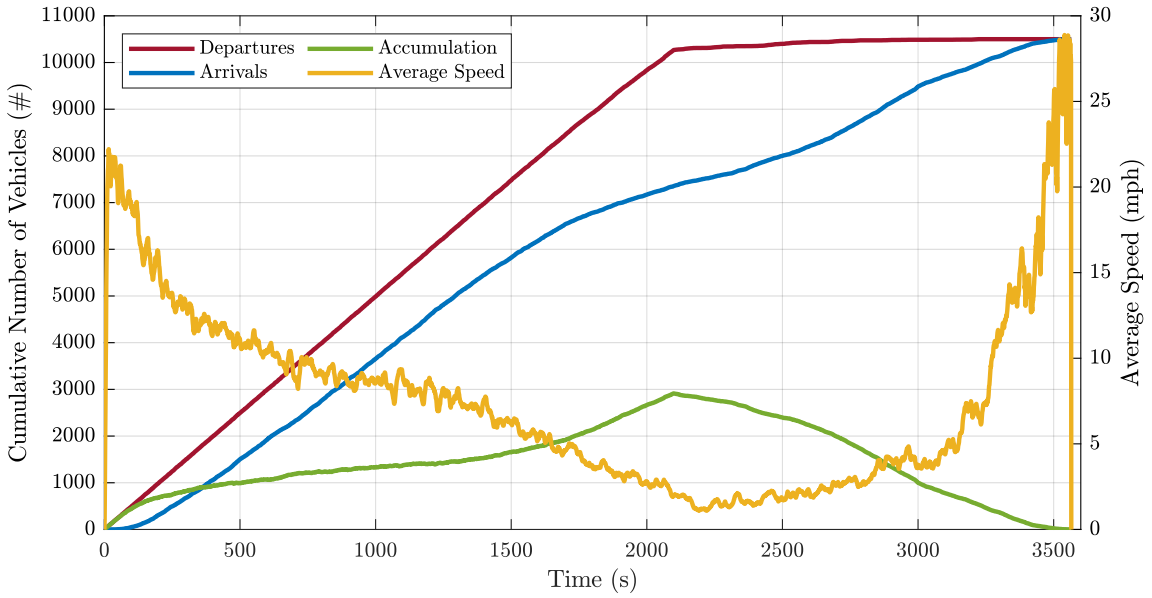

Figure 5.3: Evolution of the cumulative curves and average speed over time

these diagrams, known as the "hysteresis loops" in the literature, are associated with the loading and unloading phases of the network and different behavior of the urban networks in loading and unloading periods due to abrupt and immense changes in the demand as in this experiment. However, in real life, the changes in demand pattern are usually gradual and the networks do not exhibit distinct loading and unloading behavior. This will reduce the likelihood of observing hysteresis loops and will result in well-defined MFD functions. The analysis in this experiment will not be impacted by this phenomenon since we can directly obtain the speed values at any point of time, needless of referring to the MFD.

The simulation provides second-by-second speed, density and flow measurements for each lane in the network, which can be inputted to the MOVES model to compute the lane-level emissions, which results in the most accurate estimates and will be used as a benchmark for comparison purposes. Furthermore, 3 other emissions estimations have been performed the link, corridor and network levels by replacing all the lanes inside each of the aforementioned aggregated representations by a single entity, whose traffic conditions are computed by averaging the traffic variables of all lanes inside

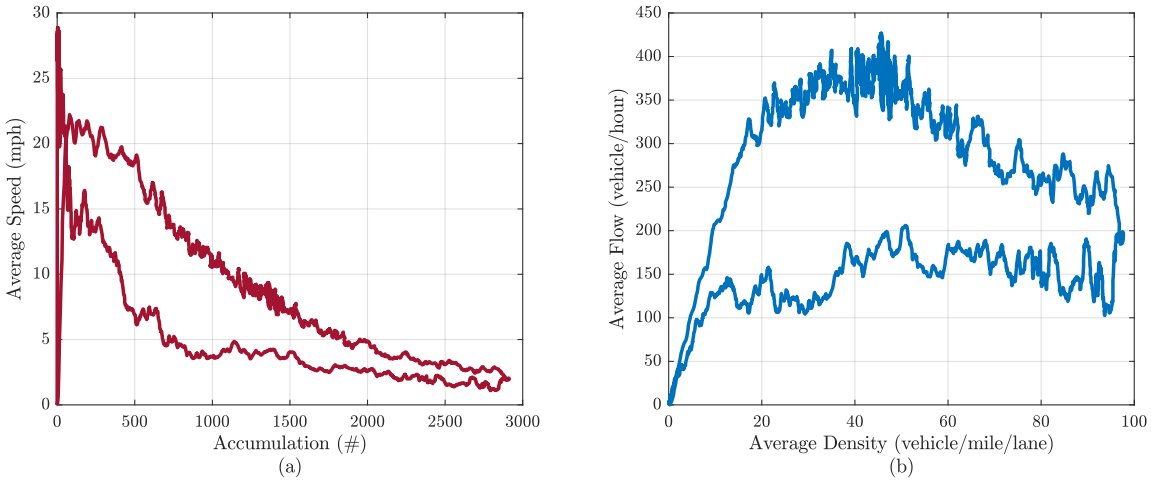

Figure 5.4: (a) Average speed vs. Accumulation, and (b) Average Flow vs. Average Density MFDs

Table 5.2: MOVES input parameters for the numerical experiment

Parameter	
Year, month, day	Weekday in June 2021
Time of day	17:00 to 18:00
Location	Fulton county, Georgia
Vehicle type	Passenger cars
Fuel type	Gasoline
Road type	Urban unrestricted access

them. The other required inputs of the MOVES model, which will be the same through all runs, are given in Table 5.2.

The estimated emissions, total energy consumption and run-times for different aggregation levels are presented in Table 5.3. The lane-level network representation estimates are the most accurate results and have been selected as a benchmark to compare the accuracy and efficiency of the estimates using other aggregation levels.

Table 5.3: Emissions estimates and run-times for different aggregation levels

Component	**Aggregation level**			
	Lane	*Link*	*Corridor*	*Network*
Carbon Monoxide (CO) [kg]	64.51	59.36 (-7.97%)	66.14 (2.53%)	76.4 (18.43%)
Carbon Dioxide Equivalent (CO_2) [kg]	3646.90	3753.68 (2.92%)	3875.27 (6.26%)	4162.17 (14.12%)
Oxides of Nitrogen (NO_x) [g]	996.26	923.57 (-7.29%)	896.41 (-10.02%)	744.6 (-25.25%)
Primary Exhaust $PM_{2.5}$ [g]	32.95	31.13 (-5.51%)	27.95 (-15.14%)	24.71 (-24.99%)
Methane (CH_4) [g]	127.43	126.03 (-1.09%)	130.19 (2.17%)	116.9 (-8.25%)
Total Energy Consumption [10^9J]	50.59	52.07 (2.92%)	53.75 (6.26%)	57.75 (14.15%)
Run-time [s]	1098.7	612.5 (-44.25%)	206.2 (-81.24%)	84.0 (-92.35%)

Relative percent change with respect to the lane-level results in parentheses

The numbers in parentheses in this table show the relative percent change in the estimates with respect to the lane-level estimates, which have been further plotted against the run-time of each estimation process in Fig. 5.5. The results indicate that using an aggregated network representation in link, corridor and network levels, will result in 1.8, 5.3, and 13.1 times faster estimation, respectively, compared to the lane-level estimation. However, this higher efficiency comes at the expense of losing some accuracy. While the link-level estimates have a maximum relative error rate of 8%, the relative error of network-level estimates is as high as 25%. On the other hand, the relative error of the corridor-level results for all components except $PM_{2.5}$ is below 10%. Considering the 81% reduction in the execution time, the corridor-level aggregation looks as an efficient and plausible approximation method for the microscopic emissions estimation at the lane level.

Although the results presented here are for an arbitrary and simplified network, the two main impacting factors for the aggregated-level emissions estimation in the MOVES model can be listed as:

1) The lengths provided in the network configuration to the MOVES model serve as the trip length on the given entity. This is valid for the lane- and link-

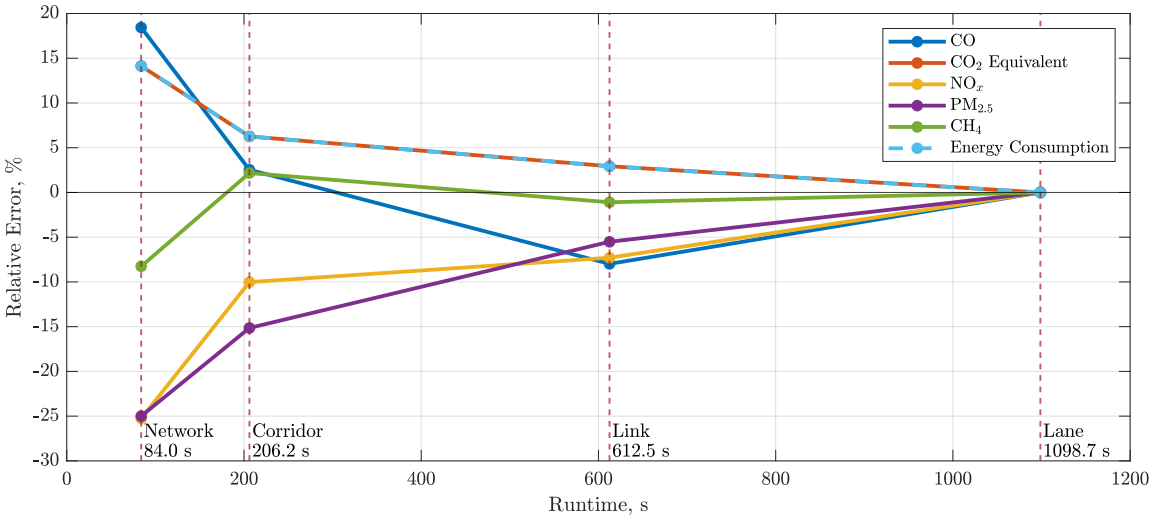

Figure 5.5: Emission estimates vs. Run-times for different aggregation levels

level network representations; however, in the corridor and network levels, the travelers do not necessarily travel the entire length and their trip length is usually lower than the length of the corridor or network. Therefore, using the actual corridor and network lengths will most probably result in overestimation of the emissions. Using average trip length instead of entity length in corridor- and network-level representations can help to mitigate this issue and increase the accuracy of the results.

2) Implementing an aggregated representation will result in a less-detailed rendition of the traffic flow evolution in the emissions estimation process, which will prevent the estimation process from capturing the entire driving cycles in the network and presumably result in underestimation of the emissions. Although this was expected by implementing aggregated representations of the network, it can be diminished by calibrating the aggregated estimates to the microscopic estimates.

As it can be seen in Fig. 5.5, the estimated components exhibit different behaviors in different aggregation levels, except the CO_2 emissions and total energy consumption which have totally similar behaviors, stemming from individual formulations with different weights of the impacting factors used to compute them. Therefore, any calibration process for the emissions estimates in different aggregation levels should be network and component specific.

Once the results from a training set has been calibrated for a network, a general recipe can be provided to estimate the emissions with high accuracy using the aggregated traffic variables. The main advantage of such method in addition to its significantly higher efficiency would be that the emissions inside any region of the network can be easily estimated by tracking a single traffic variable: the number of vehicles inside the region or accumulation. This will enable the practitioners to develop perimeter control strategies aiming to minimize the emission of any compo-

nent for a multi-region network with well-known MFDs and calibrated MFD-based emissions estimation methods.

5.6 Chapter Summary

This chapter presented a comprehensive review of studies incorporating MFD dynamics for emissions estimation using various microscopic estimation frameworks. These studies show the potential of applicability of MFD-based tools for emissions estimation. However, the accuracy of existing models in estimating the emissions of large-scale urban networks is questionable due to their inability in capturing the variations in traffic conditions across such networks.

As a solution to this problem, we have proposed to develop multi-reservoir emissions estimation framework by partitioning large-scale networks into smaller regions with homogeneous traffic conditions and low-scatter MFDs similar to the multi-reservoir DTA solution method proposed in Chapter 3, which can result in more accurate network-wide emissions estimation. The key component of this framework is finding a method to accurately estimate the emissions using aggregated network representation and its corresponding variables. A numerical experiment on an arbitrary network shows that the estimation efficiency can increase significantly by implementing aggregated network representation, albeit the results will be less accurate the more aggregated the representation becomes. The possible reasons and considerations for future application have been discussed, which would lead to developing calibrated aggregated-level methods, which can estimate the emissions efficiently and accurately.

After calibrating the MFD-based emissions estimation method to acceptable levels of accuracy and efficiency, traffic control strategies can be proposed to optimize the energy consumption and emissions of CO, CO_2, NO_x, $PM_{2.5}$, CH_4, VOC, etc. The proposed control strategies can include perimeter control strategies in the boundaries

of the regions, ramp-metering strategies at the connections to the freeways and signal timing strategies, which is known to influence the shape of the MFD.

Acknowledgments

The research in this chapter has received funding from NCST research project #DOT 69A3551747114.

CHAPTER 6
PARAMETER ESTIMATION OF THE MACROSCOPIC FUNDAMENTAL DIAGRAM

This chapter is modified from the following published preprint at the time of submission of this dissertation:

Rafegh Aghamohammadi and Jorge A. Laval (2021). "Parameter Estimation of the Macroscopic Fundamental Diagram: A Maximum Likelihood Approach". In: *Preprints.org*, 2021100310.

6.1 Introduction

The previous chapters have discussed the application of the macroscopic fundamental diagram (MFD) in dynamic traffic assignment (DTA) and emissions estimation, given that the MFD of each zone relating the aggregated traffic variables inside the zone is already known. Existence of such well-defined relationship, i.e., MFD, was empirically verified for the first time by Geroliminis and Daganzo (2008) using the loop detector and taxi GPS data from Yokohama. Since then, many researchers have collected network-wide data to investigate the relationship between average flow, density and speed across urban networks and have derived the MFD for several cities around the world (see e.g., Buisson and Ladier, 2009; Tsubota et al., 2014; Wang et al., 2015; Ambühl et al., 2018; Knoop et al., 2018).

Nevertheless, deriving the MFD using the empirical data is challenging since (i) the required loop detector data is not available in most of the cities, (ii) in the networks with available loop detector data, the loop detectors cover only a fraction of streets in the network, and (iii) the data coming from various loop detectors is prone to bias and inaccuracy, which makes the data cleaning and processing cumbersome. These

challenges make it difficult to obtain systematic insight into MFD characteristics and the factors impacting its shape.

To overcome these challenges, some studies have resorted to micro-simulation approaches to see how the changes in network topology and control can affect the shape of the MFD (see e.g., Ji et al., 2010; Knoop et al., 2014; Gayah et al., 2014). However, these studies are mostly carried out on small arbitrary networks since the calibration of the micro-simulation models on large-scale networks is a cumbersome task due to their complex and heterogeneous nature. Therefore, the micro-simulation approach might not be a feasible option to study the impact of network characteristics on the shape of the MFD of real-life networks.

Analytical approximation can be a useful tool to estimate the MFD without having to face the data collection and preparation hurdles. Daganzo and Geroliminis (2008) introduced the Method of Cuts (MoC) as an analytical approximation method for estimation of an upper bound for the MFD of homogeneous corridors using the variational theory (VT; Daganzo, 2005b; Daganzo, 2005c). Later, Laval and Castrillón (2015) developed the stochastic method of cuts (SMoC) by extending the MoC to heterogeneous corridors by introducing stochasticity and using renewal theory. This model is the most detailed analytical MFD estimation model to date, incorporating the most number of the influential network topology and control characteristics and the variations in their values over time and space.

Both of the mentioned analytical methods are proposed for estimation of the MFD of signalized corridors, while limited discussion have also been presented on their extension for network MFD estimation. These extensions come along with a number of assumptions and simplifications that make them impractical for estimating the MFD of real-life large-scale networks due the inherent heterogeneity of their topological and control characteristics. The reasons behind complexity of traffic dynamics in large-scale networks compared to signalized corridors includes but is not limited to

significantly higher number of links, inclusion of streets from different functional classifications and different speed limits, distribution of origins and destinations (trip start and end points) all across the network, directionality of the links, different number of lanes in each link, turning movements at intersections, different control strategies at intersections (e.g. pre-timed signals, actuated signals, roundabouts, stop and yield signs), etc. Therefore, some considerations have to be made before implementing the current methods for estimation of network MFDs.

The purpose of this chapter is twofold: (i) first, to extend the SMoC for analytical estimation of the MFD of large-scale real-life networks and (ii) second, to propose a method to estimate the parameters of the stochastic analytical MFD using empirical data for both corridors and networks and investigate the impact of different network topology and control characteristics on the shape of the MFD. The network-wide dataset studied here is obtained from the recently-published loop detector data (LDD) in UTD19 (Loder et al., 2019) for 40 cities around the world.

6.2 Background

The MFD estimation methods in the literature can in general be classified into 3 main categories: (i) empirical, (ii) simulation-based and (iii) analytical studies. Although the empirical method is the most reliable and accurate way to estimate the MFD, the required data is not available for many cities and even when available it is subject to significant errors and bias. The exorbitant dependency of empirical approach on the availability and accuracy of network-wide data makes it an impractical option for estimation of MFD in many cities. While simulation seems to be a promising tool to overcome the dependency of empirical approach on data, calibration of detailed micro-simulation models for large-scale urban networks is a cumbersome and prohibitive task.

The aforementioned drawbacks of the empirical and simulation-based MFD estimation methods highlight the significance of the analytical approach to determine the network MFD using network topology and control characteristics and find out the impact of each factor on the shape of the MFD. However, no universal recipe yet exists in the literature for analytical approximation of MFD for large-scale networks. The present efforts in the literature investigate the analytical estimation of MFD for signalized corridors and there are some limited extensions to estimate network-wide MFDs. In this section, we will first present the literature on analytical estimation of corridor MFDs and later review the analytical methods for estimation of network MFDs.

6.2.1 Analytical Corridor MFD Estimation Methods

Daganzo and Geroliminis (2008) has proposed the Method of Cuts (MoC) as the first analytical method for approximation of the MFD of signalized homogeneous corridors implementing the VT. In this method a set of linear cuts are plotted on the average flow-density space based on moving observers travelling in the corridor, forming an upper bound for the corridor MFD. The average flow-density MFD is defined as the steady-state flow at any location, x, as:

$$q(k) = \min_{v}\{\phi(k) + vk\}, \qquad \text{(MoC)} \qquad (6.1)$$

where q and k are average flow and density, respectively, v is the average speed of the moving observer and $\phi(v)$ is its maximum passing rate. However, evaluating $\phi(k)$ for all cuts with different speeds is time-consuming and redundant, therefore, Daganzo and Geroliminis (2008) proposes three families of "practical cuts" using stationary, forward-moving and backward-moving observers in the corridor. These cuts form an upper bound for the MFD, which might not tightly define the MFD.

Leclercq and Geroliminis (2013) generalizes the concept of practical cuts by including a sufficient number of additional cuts representing optimal paths with different mean speeds to provide a tighter upper bound for the MFD. Xie et al. (2013) presents two methods for estimation of the corridor MFDs with presence of public transport by (i) including buses in the method proposed in Leclercq and Geroliminis (2013) and (ii) accounting for buses in the time-space diagram according to the kinematic wave theory.

Nevertheless, the MoC is proposed for homogeneous corridors with equal block lengths and signal settings, which makes it complicated to be scaled up for the network MFD estimation. Daganzo and Lehe (2016) has extended the concept of the MoC to account for inhomogeneities in block length and offset of adjacent traffic signals and formulated the method as a linear program.

To more generally account for the variations in the block length and signal timing settings, Laval and Castrillón (2015) has extended the MoC to inhomogeneous corridors developing the Stochastic Method of Cuts (SMoC). It first defines the stochastic corridor random variable, where each inhomogeneous corridor is a realization of this random variable with three main independent random variables: block length (ℓ), signal green (g) and red (r) times which can vary in time and across different segments of the corridor. In order to simplify the representation of the problem, it has reformulated the individual link fundamental diagram (FD) and network parameters in dimensionless form by measuring time in units of critical headway and space in units of jam spacing, assumed the coefficients of variations (δ) of the three main parameters are equal and performed a density transformation to have symmetric fundamental diagram (FD) and MFD.

This method results in normally-distributed stochastic cuts, $q_s(k)$; therefore, the yielding upper-bound MFD found using this model will also be stochastic, whose

CDF, $F_{q(k')}(q)$, is given by:

$$F_{q(k')}(q) = 1 - (1 - F_{s_0}(q))^B \prod_{s \in \Omega} (1 - F_{s,k'}(q)), \qquad \text{(SMoC MFD CDF)} \qquad (6.2)$$

where $\Omega = \{s_1^\sharp, s_1^\flat, s_2^\sharp, s_2^\flat\}$ denotes the set of different forward-moving (denoted as $\cdot^\sharp$) and backward-moving (denoted as $\cdot^\flat$) strategies, $F_{s,k'}(q)$ is the CDF of cut s, s_0 is the stationary cut and B is the minimum number of stationary cuts. The authors found that the shape of the MFD mainly depends on two dimensionless parameters: (i) *mean block length to critical length ratio*, λ, and (ii) *mean red to green signal time ratio*, ρ. The SMoC has shown promising results in capturing the shape of MFD for corridors with different values of λ, ρ and δ.

Castrillón and Laval (2018) revisits the MoC to include the impact of buses on the MFD of homogeneous corridors. They found that presence of buses has two major impacts: (i) "moving bottleneck effect" due to lower speed of buses compared to other vehicles and (ii) reduction in capacity due to the "short block" effect. Each of these effects result in an additional cut in the MoC providing a tighter upper bound for the corridor MFD.

Recently, Tilg et al. (2020) has compared the performance of MoC and SMoC for an empirical corridor MFD obtained by data from the city of Munich, Germany. The empirical data is for the northbound direction of an arterial segment of 1.1 km with 5 intersections, where all intersections are controlled by actuated traffic signals. Although there are some discrepancies between the estimated analytical MFDs and the empirical MFD, which might partially be due to the turning movements at intersections and other real-life network complexities which are not considered in the analytical methods, the results show that the MoC provides a good estimate for the corridor capacity and the SMoC delivers a good fit for the free-flow branch of the MFD.

6.2.2 Analytical Network MFD Estimation Methods

Although the analytical MFD estimation methods discussed in the previous section are proposed for estimation of corridor MFDs, most studies have discussed the potential extension of their method to network MFD estimation and have tested the model on toy networks or simplified real-life networks under some assumptions. Daganzo and Geroliminis (2008) has tested the MoC on the empirical data from Yokohama and the simulation data for San Francisco by decomposing the network into 1-lane links, thereby eliminating lane changing. For the Yokohama data, assumptions are made for the link FD and signal setting parameters. While the MoC results in a good fit for the simulation data for San Francisco assuming pre-timed signals with a common cycle, it overestimates the speed and flow for the Yokohama network, which authors have mentioned can be partly due to the measurement errors in the empirical data, heterogeneity of the network and actuated signal settings varying with time.

Geroliminis and Boyacı (2012) introduces a number of extensions and refinements to the VT to analytically investigate the effects of different traffic signal offsets and block lengths on the homogeneous network MFD. Leclercq and Geroliminis (2013) has extended its method to estimate the MFD of a simple network consisting a number of parallel hyperlinks to investigate the impacts of different route choice models on the MFD. The results show that variation in densities over the network can affect the shape of the MFD.

Laval and Castrillón (2015) claims that their proposed SMoC can provide a good approximation for networks as well as corridors if the following conditions are met: (i) the density is distributed homogeneously over the network, and (ii) the block lengths and signal green and red times are i.i.d. random variables. If these conditions are met, all network paths can be considered as realizations of the stochastic corridor random variable and the SMoC will provide a good approximation for the network MFD. The approximated analytical MFD for the Yokohama network using the assumed

network parameters in Daganzo and Geroliminis (2008) shows good agreement with the empirical MFD.

Ambühl et al. (2018) developed a method to derive a functional form with a physical meaning for the upper bound MFD. The proposed model results in a trapezoidal function using a smoothing parameter and four physical properties of the network: free-flow speed, backward wave speed, jam density and intersection capacity. The authors have fitted the model to empirical loop detector data from 4 cities in Europe, which has resulted in viable estimates for the model parameters. However, this model does not take the block lengths into account, which has been shown to have a significant impact on the shape of the MFD in most of the other studies.

6.3 Experimental Setup

As discussed earlier, several analytical methods are presented in the literature for the MFD estimation. Although the models are mainly proposed for analytical estimation of corridor MFDs, there have been limited discussions about extending them to estimate the network MFDs. Among the models presented in the literature, the SMoC seems to be the most detailed model incorporating the most number of the influential network topology and control characteristics taking the spatial and temporal variations in the parameters into account. However, this method has not been empirically verified for estimation of network MFDs. Here, we will first present the original SMoC model and explore how it can be extended to approximate the MFDs of large-scale networks and later a method will be presented to estimate the SMoC parameters to fit them to the empirical network MFDs and draw insights into the impact of different parameters on the shape of the MFD.

6.3.1 Structure of the SMoC

The stochastic method of cuts (SMoC) proposed by Laval and Castrillón (2015) introduces an stochastic extension to the MoC by allowing for variations in the network topology and control parameters. The variables used in the model are listed in Table 6.1. This model builds on the VT and MoC using cuts based on the same three categories of forward-moving, stationary and backward-moving observers. The main addition of the model is its ability to capture the spatial variations in block lengths and the spatial and temporal variations in signals cycle and phase durations, enabling its application for estimating the MFD where the signal timing settings are different for different intersections or actuated signal controls are in use.

The model assumes that all links follow a triangular FD with capacity Q and jam density K. The FD has been normalized by expressing the flow q and density k in terms of Q and K, respectively. Laval and Chilukuri (2016) shows that by applying a shear transformation to the FD one can reach a symmetric FD that can help to simplify the analytical computations. Laval and Castrillón (2015) articulates that if such transformation is applied to the link FDs in a network, the resulting MFD

Table 6.1: SMoC model parameters

Link Fundamental Diagram	
Q	Capacity (veh/h/ln)
K	Jam Density (veh/km)
u	Free-flow speed (km/h)
w	Backward wave speed (km/h)
θ	Free-flow speed to backward wave speed ratio (dimensionless)
Network Parameters	
μ_ℓ	Mean block length (m)
μ_g	Mean signal green time (s)
μ_r	Mean signal red time (s)
λ	Mean block length to critical block length ratio (dimensionless)
ρ	Mean signal red-to-green ratio (dimensionless)
δ	Coefficient of variation of network parameters (dimensionless)

of the network will also have a symmetric shape. This density transformation is a mapping $(q, k) \rightarrow (q, k')$ that keeps the same flow but transforms the density to the transformed density k' via the following equation:

$$k' = k - \frac{1}{2}\left(1 - \left(\frac{1}{w} - \frac{1}{u}\right) q\right), \qquad \text{(density transformation)} \qquad (6.3)$$

where the transformed density domain is $k' \in [-1/2, 1/2]$. This transformation is illustrated graphically in Fig. 6.1a. One of the main benefits of this transformation would be that as a result of the transformation the empirical data fed to the MLE method will be parametric itself, which will make finding the PDF of the given dataset more tractable providing the method more degrees of freedom to seek for a better fit.

After calculating the cumulative distribution function (CDF) of cuts for each strategy, s, the CDF of the corridor MFD, $F_{q(k')}(q)$, can be found using Eq. (6.2). It has been shown that the cuts can analytically be determined by the mean block length, μ_ℓ, the mean green, μ_g, and the mean red, μ_r, times and their respective standard deviations δ_ℓ, δ_g and δ_r, which for the sake of further simplification of the problem are all assumed to be equal $\delta_\ell = \delta_g = \delta_r = \delta$.

The authors further claim that the shape of the MFD is mainly influenced by two dimensionless parameters: (i) *the mean block length to critical length ratio* $\lambda = \frac{\mu_\ell}{\mu_{\ell^*}}$, and (ii) *the mean red to green signal time ratio* $\rho = \frac{\mu_r}{\mu_g}$. The critical block length is

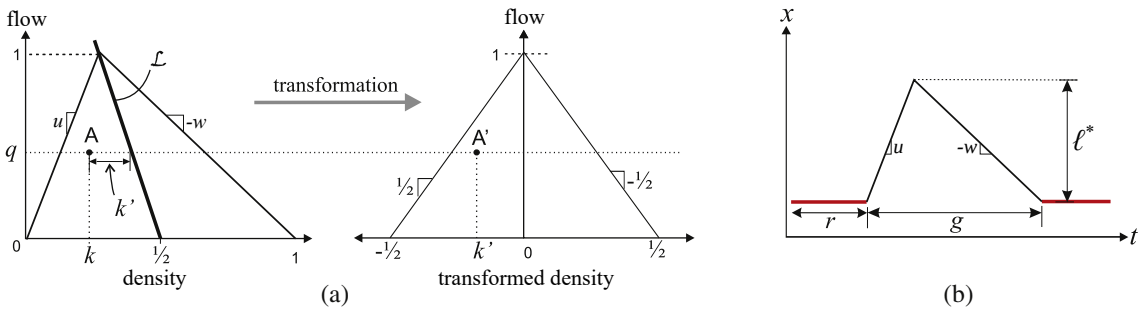

Figure 6.1: (a) Density transformation and (b) Critical block length on time space diagram

the minimum block length to avoid the short block effect; see Fig. 6.1b. As seen in this figure, the critical block length can be computed as:

$$\ell^* = ug\frac{k_c}{K} = \frac{Q}{k_c}g\frac{k_c}{K} = \frac{gQ}{K}, \qquad \text{(critical block length)} \tag{6.4}$$

where k_c is the critical density. If the dimensionless forms of flow and density are used in the equation above, one will have:

$$\ell^* = g. \qquad \text{(dimensionless critical block length)} \tag{6.5}$$

Knowing the values of the required parameters for a corridor or network, one can compute the cut CDFs and eventually determine the corresponding CDF of the MFD. However, gathering the data required to find the model parameters is not a straightforward task. In many cities the data is not readily available due to the lack of measurement equipment and even when it is available, it is prone to different biases and measurement errors. Computing the parameters becomes more challenging if the method is applied for network MFD estimation due to the higher complexity and heterogeneity of the network topology and control settings. We will delve into these challenges in more detail later in this chapter. Here, we will present a method to statistically estimate the parameters by applying the maximum likelihood estimation (MLE) method to the empirical corridor and network MFDs based on loop detector data (LDD). But before that, we need to check how the original model can be extended to approximate the MFDs of large-scale networks.

6.3.2 Extension of the SMoC for Network MFD Estimation

The SMoC has been originally developed for corridor MFD estimation. The authors have claimed that the model can be applied for estimation of network MFDs as well. However, the model cannot be readily implemented to estimate the network MFDs.

Although the model can be extended in many different ways to take into account additional network properties such as the presence of buses by including additional cuts, here, we want to stick to the original model framework and cuts, and only explore whether there would be any need for redefinition of any of the model parameters while scaling up from corridors to networks.

A brief look at Table 6.1 reveals that link FD and the mean block length parameters can be defined and calculated for networks in the same fashion as the corridors. However, this is not true for the signal timing parameters and especially for the mean red to green signal time ratio, ρ, which is one of the most influential parameters impacting the shape of the MFD. In the corridor case, the calculation of ρ is straightforward since at each intersection the green and red times for the corridor of interest are known explicitly. However, in the network case, there arises confusion in selecting the green and red times for each intersection, since there are several approaches with different green and red times. If the signal timing data is available for all approaches of all intersection across the network, assuming that most signals have only two phases, the sum of green times will almost be equal to the sum of red times (neglecting the lost times) at the approach, intersection and network level, therefore resulting in $\rho \approx 1$.

Nevertheless, in many cities there are intersections with protected left turn phases or more than two phases, which will make the mean red time higher than the mean green time and ultimately $\rho > 1$. To gain a better understanding of how the presence of traffic signals with any general number of phases, n_i, impacts the value of ρ, let us first focus on a single phase of the intersection. By definition, one can find the mean red to green signal time ratio for a phase, ρ_j, as:

$$\rho_j = \frac{r_j}{g_j} = \frac{\sum_{k=1, k\neq j}^{n_i} g_k}{g_j}, \qquad (\rho_j \text{ for a single phase, } j) \tag{6.6}$$

where g_j is the average green time for phase j and g_k represents the average green time for the rest of the phases, k.

The value of ρ in the model is in fact an indicator of the available intersection capacity. Higher ρ values mean that more time is assigned to the red phase compared to the green time and less intersection capacity is being utilized, while lower ρ values would mean higher utilization of the cycle time and intersection capacity. Therefore, the rational way to compute the average ρ^i for an intersection, i, is to take a weighted average of the values for all phases in the signal timing with respect to their respective demand or flow. Without loss of generality, one can assume that the signal timing is efficiently programmed to be proportionate to the demand values at each phase. Hence, the weight of each phase, w_j, can be found as:

$$w_j = \frac{f_j}{\sum_{j=1}^{n_i} f_j} = \frac{g_j}{\sum_{j=1}^{n_i} g_j} = \frac{g_j}{C}, \qquad \text{(weighting factor of each phase)} \tag{6.7}$$

where f_j is the demand or flow of each phase j and C is the total cycle time for the intersection. Now, the mean red to green time ratio for intersection i, ρ^i, can be computed as:

$$\rho^i = \sum_{j=1}^{n_i} w_j \rho_j = \sum_{j=1}^{n_i} \frac{g_j}{C} \times \frac{\sum_{k=1,k\neq j}^{n_i} g_k}{g_j} = \frac{1}{C} \sum_{j=1}^{n_i} \sum_{\substack{k=1 \\ k\neq j}}^{n_i} g_k = \frac{(n_i - 1)C}{C} = n_i - 1. \tag{6.8}$$

Now that we have obtained a simple expression for ρ^i, the mean red to green signal time ratio for the entire network with N intersections can be calculated as:

$$\rho = \frac{\sum_{i=1}^{N} \rho^i}{N} = \frac{\sum_{i=1}^{N} n_i - 1}{N} = \frac{\sum_{i=1}^{N} n_i}{N} - \frac{N}{N} = \bar{n} - 1, \qquad (\rho \text{ of a network}) \tag{6.9}$$

where $\bar{n}$ is the average number of signal phases of the intersections in the network. Now that we have found how all the model parameters can be determined to approximate the network MFDs, we can proceed to estimation of model parameters using empirical data.

6.3.3 Parameter Estimation Using Maximum Likelihood Estimation Method

Once the probability distribution of a dataset is known, we can estimate the population parameters using the distribution fitting methods such as the method of moments or the maximum likelihood estimation (MLE) method. In this study, we will use the MLE due to its large-sample or asymptotic properties to estimate the analytical MFD given by the SMoC and its parameters for several different cases using their empirical MFDs. The MLE estimators become minimum-variance unbiased estimators as the sample size increases and have approximate normal distributions and approximate sample variances that can be used to generate confidence bounds and hypothesis tests for parameters.

Let the probability density function (PDF) of the MFD, which can be derived as the derivative with respect to q of the CDF given in Eq. (6.2), be denoted as $f(x|\Theta)$. The likelihood function for a given dataset $\mathbf{X}$ is the joint density of n independent and identically distributed (i.i.d.) observations, which is the product of their individual densities:

$$L(\Theta|\mathbf{X}) = f(x_1, \ldots, x_n|\Theta) = \prod_{x_i \in \mathbf{X}} f(x_i|\Theta), \qquad \text{(likelihood function)} \quad (6.10)$$

where Θ is the unknown parameter vector. Considering that some of the model parameters presented in Table 6.1 are dependent and can be derived using the other parameters, we choose $\Theta = \{Q, K, \theta, \mu_\ell, \mu_g, \rho, \delta\}$ as the independent parameter vector. It is worth noting that the likelihood function is a function of the parameters

conditioned on the data whereas the joint density is a function of the data conditioned on the parameters. In the MLE method, we aim to find the parameter values that maximize the likelihood function for the observed data. However, for the sake of more convenience in the computations, it is preferred to maximize the natural logarithm of the likelihood function, l:

$$l(\Theta|\mathbf{X}) = \ln L(\Theta|\mathbf{X}) = \ln \prod_{x_i \in \mathbf{X}} f(x_i|\Theta) = \sum_{x_i \in \mathbf{X}} \ln f(x_i|\Theta). \quad \text{(log-likelihood)} \quad (6.11)$$

By maximizing the log-likelihood function, we find the set of parameter estimators $\widehat{\Theta}$ for each dataset, under which observing the dataset is most probable. Under certain regularity conditions, the estimators are known to asymptotically tend to have a multivariate normal distribution for large samples:

$$\Theta \sim \mathcal{N}_p(\widehat{\Theta},\ J^{-1}(\widehat{\Theta})), \quad \text{(multivariate normal distribution of estimators)} \quad (6.12)$$

where for our model the number of variables is $p = 7$ and the covariance matrix of parameters, $J^{-1}(\widehat{\Theta})$, is the inverse of the observed Fisher's information matrix, $J(\Theta)$, calculated as:

$$J(\Theta) = -\frac{\partial^2 l(\Theta)}{\partial \Theta^2}, \quad \text{(observed information matrix)} \quad (6.13)$$

where the J^{-1} is the covariance matrix of the model parameters and its diagonal elements are the variance of the parameters. The regularity conditions include: (i) the true parameter values must be interior to the parameter space, (ii) the log-likelihood function must be thrice differentiable and (iii) the third derivatives must be bounded (Geyer et al., 1994). Knowing the distribution of the estimators will enable us to perform hypothesis testings for the estimated values.

6.3.4 Data Sources and Preparation

The proposed estimation method will first be applied on a empirical MFD dataset for a 4-block corridor from Munich provided by Tilg et al. (2020). This dataset is already consisted of the average flow versus average density data for the corridor and, thus, does not need any further preparation to be used by the proposed method in this study.

In order to evaluate the method for analytical network MFD and parameter estimation, the network-wide loop detector data published in UTD19 (Loder et al., 2019) for 40 cities around the world is used. The dataset includes flow, loop detector occupancy and sporadic speed measurements for 3 up to 60 minute intervals. We first need to derive the empirical MFDs for the networks in the dataset and then apply the proposed estimation method to approximate the analytical MFD and estimate the network topology and control parameters. However, the provided raw data needs to be cleaned and processed to produce valid empirical MFDs, which will be elucidated in the ensuing subsection for the sake of reproducibility of the research.

Network empirical data cleaning and processing

The UTD19 dataset includes the loop detector measurements, loop detector descriptions and link descriptions for 40 urban networks worldwide. We first separate the respective data for each network before processing the data further. Our initial evaluation of the dataset reveals that the time intervals of the measurements for the city of Paris is 1 hour, which is very long for the purpose of finding the MFD. Therefore, we exclude this city from further analysis in this study. Furthermore, the data for the 9 cities of Bolton, Birmingham, Groningen, Innsbruck, Manchester, Melbourne, Rotterdam, Torino and Utrecht provide partial or no occupancy measurements. Hence, these 9 cities are also excluded from further investigation in this study. For the remaining 30 cities, the following data cleaning steps are applied:

- Similar to all other empirical data, the UTD19 dataset includes outliers and faulty values. Some of the measurements have already been marked as erroneous in the dataset, which are eliminated from the dataset. In addition to the marked measurements, other faulty and missing values have been excluded from the dataset based on the following criteria:

  - Negative or nonnumerical flow or occupancy measurements,
  - Occupancy measurements larger than 1,
  - Flow measurements larger than 2500 veh/hr/ln,
  - Flow measurements less than 10 veh/hr/ln for occupancy values between 0.2 and 0.75,
  - Flow measurements higher than 100 veh/hr/ln for occupancy values larger than 0.95.

- After eliminating all the outliers and faulty values, we select only the loop detectors with valid data for more than 80% of the time intervals.
- Moreover, we only keep the intervals which have valid data from more than 80% of the loop detectors selected in the previous step.

After cleaning and processing the raw data as explained in the steps above, we obtain the average flow versus average occupancy MFDs for 30 cities as shown in Fig. 6.2. One of the most conspicuous issues observed in these empirical MFDs is that the resulting diagrams only cover a very limited range of occupancy (density), resulting in a lack of the capacity state and the descending congested branch of the MFD for most cases. This deficiency in the empirical data might impact the performance of the proposed parameter estimation method and its results, which will be discussed further in the next section.

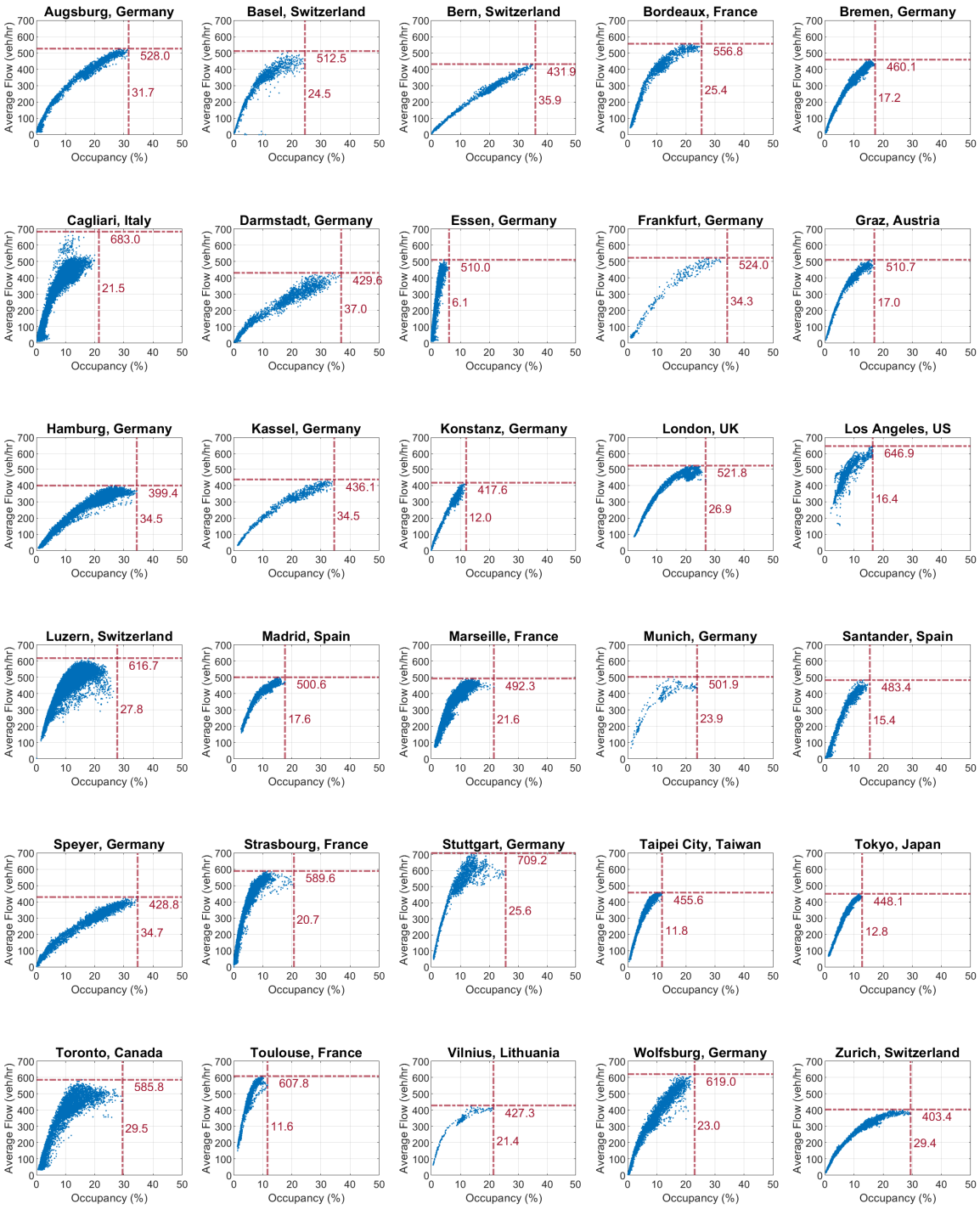

Figure 6.2: Empirical Average Flow vs. Occupancy MFDs for the 30 networks investigated in this study. The vertical and horizontal lines and numbers indicate the maximum occupancy and flow values, respectively.

In order to derive the average flow versus average density MFDs, one should convert the occupancy measurements for each individual loop to density values before aggregation, which would need further information on the length of each loop detector and the average vehicle length for each city to approximate the average field length for each loop detector. However, the proposed estimation method in this study is based on the normalized flow and density values by the capacity flow and jam density, respectively. Therefore, if a linear relationship between the occupancy and density values is assumed, one can regard the occupancy values as the normalized density values and there will not be any need to further loop detector and average vehicle length data for occupancy to density transformation.

6.3.5 Application Considerations

Our preliminary results indicate that the proposed MLE method might result in unrealistic estimated values, e.g. negative values for strictly positive model parameters or extremely large or small values that do not conform to the expected range of the ground truth values, while producing satisfactory graphical fits. In order to prevent the estimation process from resulting in such unrealistic results, we can instead solve a constrained MLE (CMLE) problem, which allows the method to seek for optimum parameter estimates inside a set of rational parameter boundaries on the parameter space. The upper and lower bounds of these constraints for all model parameters are given in Table 6.2, which have been selected in a way to provide a large feasible

Table 6.2: Lower and upper boundary values for MLE method parameters

Parameter	μ_ℓ [m]	μ_g [s]	ρ	Q [veh/h]	K [veh/km]	θ	δ
Lower bound	100	20	1	1000	80	1	0.1
Upper bound	500	70	2	2500	200	4	0.4

space for the parameters so that the optimizer can find optimum parameters inside the parameter space. In addition to the parameter constraints given in Table 6.2, another constraint have also been imposed to keep the free-flow speed, u, less than 60 km/h.

Although the CMLE problem will ensure that the estimated values are in plausible ranges, the optimization method might not be able to find a global optimum inside the parameter space and one or more of the estimated values might be on the boundaries at the maximum log-likelihood value. This will result in a violation of the regularity conditions discussed earlier, since the optimum parameter set will not be interior to the parameter space. In such case, the first derivative of the log-likelihood function with respect to the parameter on the edge of the parameter space is not necessarily equal to zero, therefore the approximate multivariate normal distribution for the parameters will be invalid. Therefore, we should be careful about the selection of the boundaries of the parameter space to prevent this problem from happening.

However, as it will be discussed in the next section, hitting the boundaries in the present problem might be inevitable due to high number of parameters needed to be estimated, the probable errors and biases in the empirical data and the other impacting factors which are not considered in the SMoC model such as the presence of buses creating both stationary and moving bottlenecks along the network.

The next section will present the results of application of the proposed method for estimation of analytical MFD parameters for both corridors and networks under different assumptions.

6.4 Results

In this section, we will estimate the analytical MFD parameters using the proposed MLE method for two different sets of empirical MFD data. The first set is the data from a corridor and the second set includes empirical MFD datasets for several

networks obtained from Loder et al. (2019). For the network datasets, the proposed method will be tested under two different sets of assumptions.

Toward this purpose, the MFD CDF for the SMoC model has been coded in Wolfram Mathematica®. Then, the derivative of the CDF with respect to q is computed to find the MFD PDF, which has been later used to find the log-likelihood function for any given dataset and set of model parameters. A constrained optimization method incorporating the constraints given in Table 6.2 has been utilized to find the optimum parameter value maximizing the log likelihood function. Afterwards, the second derivatives of the likelihood function are calculated to find the observed information matrix and the variances of the model parameters, which will be used in hypothesis testing for the estimated parameter values.

6.4.1 Corridor MFD Parameter Estimation

As the original SMoC model was initially proposed for corridor MFD estimation, we first intend to test the proposed parameter estimation method on a corridor MFD dataset. Tilg et al. (2020) has collected the LDD for a 4-block segment of one direction of an arterial with total length of 1.1 km in Munich, Germany for 24 hours. The signal timings for 5 intersections along the segment, which are all actuated, are also collected for the duration of study. The authors have assumed the link capacity flow $Q = 1690$ veh/hr/ln, the jam density $K = 150$ veh/hr and the free-flow speed $u = 45$ km/hr. The mean green and red times are not reported in the study; however, we have approximated these values by the median green and red times, respectively, which are presented in a figure in the study. The other required parameter values for the SMoC have been calculated based on the available data and the analytical MFD have been derived using these values.

Fig. 6.3 shows the fitted MFD CDF based on our proposed method. The figure provides the 50% MFD CDF (solid black line) and 16% and 84% CDF limits (the gray

Table 6.3: Empirical and estimated parameter values for the Munich dataset

	λ	μ_ℓ [m]	μ_g [s]	ρ	Q [veh/h]	K [veh/km]	θ	δ
Empirical Values	2.08	275	42.4	1.2	1690	150	3	NA
Estimation Results	2.07	267.3 (2.26)	45.2 (2.39)	1.57 (18.02)	1452 (18.23)	141 (11.5)	3.2 (4.54)	0.59 (7.07)

t-statistics in parentheses

shaded area) along with the mean values for the stationary and moving cuts. The method results in a good fit to the empirical data and the estimated values presented in Table 6.3 are all statistically significant and most of them show close agreement with the measured empirical values.

In the original study, although the resulting analytical MFD provided a good fit to the free-flow branch of the empirical MFD, it overestimated the corridor capacity. In particular, we see that the original study uses a higher capacity and lower ρ compared

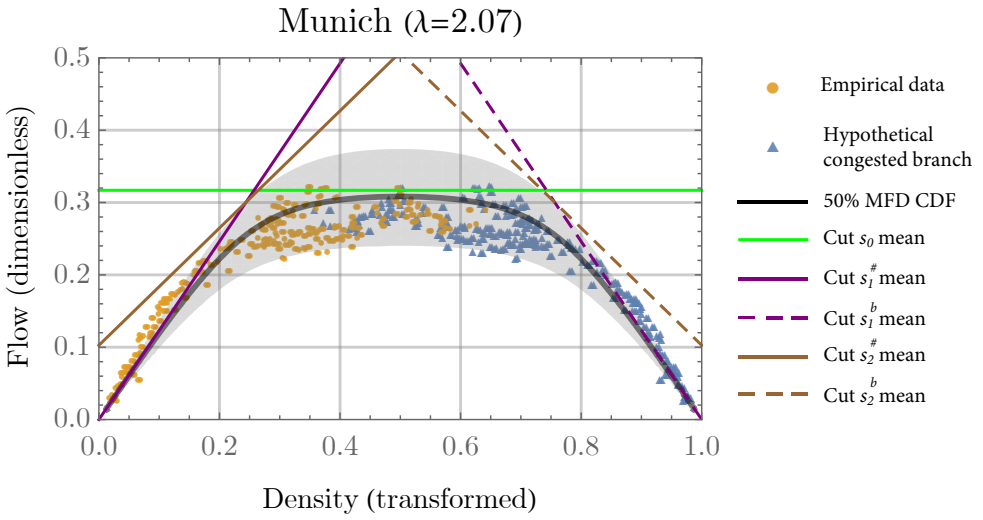

Figure 6.3: Estimated analytical corridor MFD and cuts for the Munich dataset displayed in the normalized flow units with respect to the estimated link capacity, Q, and the transformed and normalized density, k', as discussed in subsection 6.3.1. Note that the yellow markers show the empirical MFD points and the blue markers represent the hypothetical congested branch mirrored with respect to the line $k' = 0.5$. The solid black line is the 50% CDF and the shaded area shows the 16% and 84% CDF limits for the corresponding density values.

to the estimated values in this study. The observed discrepancies can be attributed to a number of issues. First, buses travel through the corridor and there is a major public transportation hub in the middle of the corridor. Presence of buses is known to lower the capacity due to their lower speed and frequent stops, imposing a moving bottleneck effect. The signals along the corridor prioritize the movement of buses causing further delay at the intersections, which is represented by higher ρ values in the SMoC.

Furthermore, there are a few minor unsignalized intersections along the corridor. Although vehicles on the corridor do not fully stop at these intersections, the turning movements from and to the intersecting streets may cause the vehicles on the corridor to lower their speeds which will eventually result in lower capacity. The large variations in the signal green and red times and the fact that we have approximated the empirical value of ρ based on the median green and red times instead of their mean values might also have contributed to the discrepancy. Taking all into account, we conclude that the SMoC can capture the shape of the corridor MFD satisfactorily provided that the parameters are carefully estimated. Our proposed method is a step in that direction.

6.4.2 Network MFD Parameter Estimation

After testing the proposed method on empirical corridor and simulation network cases, we aim to check the performance of the method to estimate the MFD parameters for empirical network data. The data source and cleaning process has been explained earlier in subsection 6.3.4. Here, we will present and discuss two sets of estimation results for this network data under different assumptions.

Estimation of all model parameters

We first assume that we do not have any a posteriori knowledge about the real value of any of the model parameters $\Theta = \{Q, K, \theta, \mu_\ell, \mu_g, \rho, \delta\}$ for the given networks and all of these parameters are to be estimated. The cleaned and processed MFD data for 30 cities from the UTD19 dataset, shown in Fig. 6.2, is given to the program coded based on the proposed method and the SMoC analytical MFD is fitted to the data from each city and the model parameters for each network are estimated.

Despite the large feasible space provided by the parameter boundaries given in Table 6.2, the estimation method results in boundary values for many datasets. For 16 cities out of the total 30 cities at least one estimated parameter value is on the determined boundaries while the graphical fits to the empirical data look satisfactory. This issue will result in a violation of the MLE regularity conditions and we will not be able to make any statistical inference from the estimation results. The issue can be resolved by changing the boundary values for the problematic boundaries; however, we refrain from doing so since the imposed boundaries are selected to keep the parameter estimates in meaningful ranges and further broadening of them would result in unrealistic estimations. Therefore, we deem the results from the 16 cities with at least one parameter estimate on the boundaries unsatisfactory and direct our further analysis on the results from the 14 cities, where this problem is not encountered. The satisfactory results mostly come from the datasets covering a wider range of occupancy values or exhibiting downward-bending trends reaching a near capacity state.

The estimated analytical MFDs and cuts for five cities, in which the method exhibits a satisfactory performance, as illustrated in Fig. 6.4, show that the method results in satisfactory fits of the analytical MFD CDF to the empirical data. The estimated parameter values for these networks are given in Table 6.4. The t-statistics for the estimated values (shown in parentheses) indicate that all of the estimated

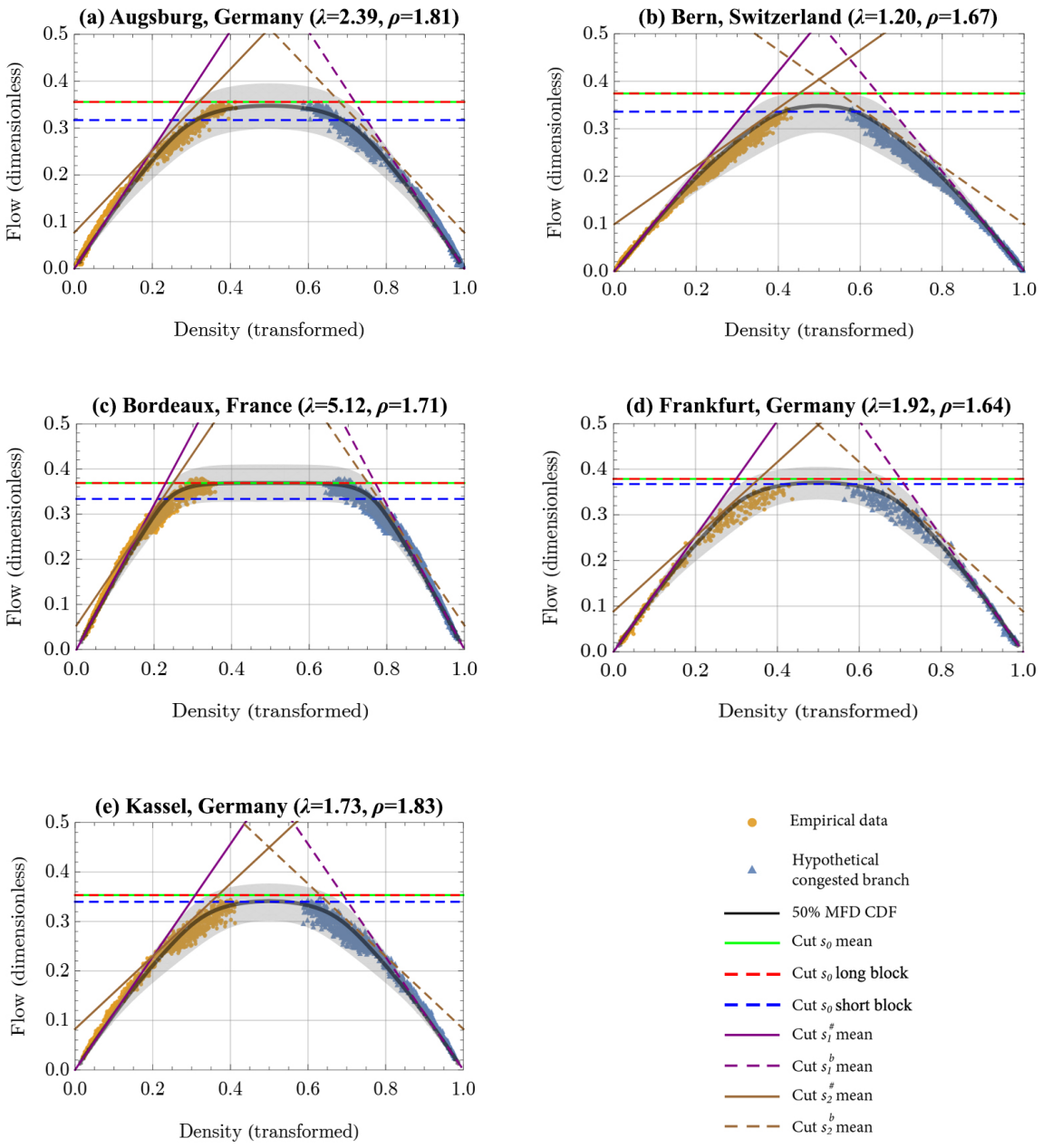

Figure 6.4: Estimated analytical MFDs and cuts using all 7 model parameters for a number of studied networks displayed in the normalized flow units with respect to the estimated link capacity, Q, and the transformed and normalized density, k', as discussed in subsection 6.3.1. Note that the yellow markers show the empirical MFD points and the blue markers represent the hypothetical congested branch mirrored with respect to the line $k' = 0.5$. The solid black line is the 50% CDF and the shaded area shows the 16% and 84% CDF limits for the corresponding density values.

Table 6.4: Estimated parameter values for estimating all 7 model parameters

City	λ	μ_ℓ [m]	μ_g [s]	ρ	Q [veh/h]	K [veh/km]	θ	δ
Augsburg [1273.9]	2.39	218.3 (55.68)	40.6 (66.92)	1.81 (127.68)	1443.2 (86.35)	178.4 (55.67)	3.58 (73.94)	0.43 (27.86)
Bern [2743.3]	1.20	152.1 (58.64)	41.2 (75.41)	1.67 (126.79)	1251.2 (61.38)	112.7 (58.64)	2.61 (74.33)	0.19 (16.26)
Bordeaux [4089.8]	5.12	386.3 (11.14)	31.2 (9.09)	1.71 (9.25)	1457.6 (27.61)	167.3 (11.43)	3.80 (56.29)	0.41 (3.73)
Frankfurt [874.2]	1.92	183.6 (5.75)	38.3 (9.46)	1.64 (12.50)	1432.6 (30.24)	159.4 (5.93)	3.50 (5.41)	0.31 (9.59)
Kassel [2258.7]	1.73	234.3 (17.19)	37.1 (18.63)	1.83 (16.72)	1235.6 (85.59)	94.2 (17.19)	2.63 (17.09)	0.32 (46.65)

Maximum log likelihood value in brackets
t-statistics in parentheses

values are statistically significant and most of them are reasonable and in the expected range of values in reality. However, at first sight, the estimated link capacity values draw attention, which are lower than the expected uninterrupted link capacity (about 2000 veh/hr). Moreover, in some cases such as Kassel, the estimated jam density is significantly lower than the anticipated jam density value (about 160 veh/km).

The estimated ρ values are all between 1 and 2, indicating that the average number of signal phases for the studied networks is between 2 and 3 based on the formulation presented in Section 6.3.2. Unfortunately, we do not have any information on the signal timing configurations for the studied network, but the estimation results sound reasonable based on what we can expect of the real-life configurations in urban network with many signals having more than 2 phases.

Out of the estimated parameters, we can only obtain information on the link lengths inside the analysis area of each network to compute the average block lengths, μ_ℓ, using the provided data. Fig. 6.5 illustrates the 95% confidence intervals for the average block length estimations along with the measured average block length using

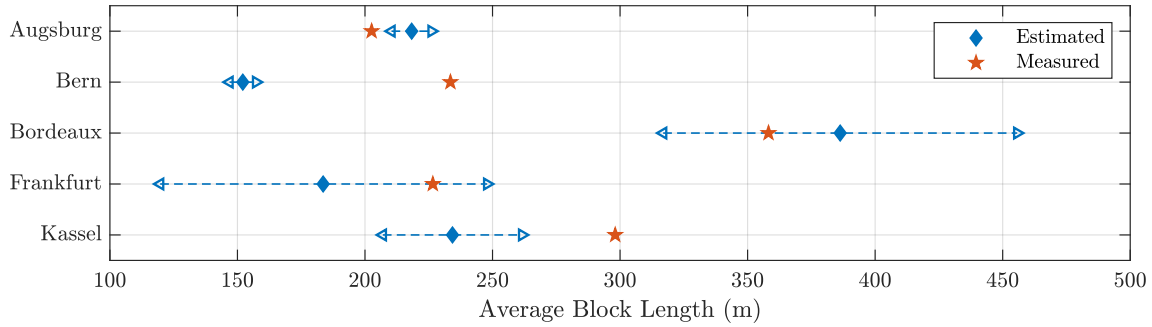

Figure 6.5: 95% estimation confidence intervals and the measured values for average block length.

the available data for the 5 networks presented in Fig. 6.4. Although the graphical fits look up to scratch, the measured real values of μ_ℓ do not necessarily fall in the 95% confidence interval of the estimations for all datasets. This might have happened due to a host of reasons, out of which the most important ones seem to be: (i) the position of loop detectors along the links, which impacts their measurements, (ii) measuring the average block length based on only the links with loop detectors or all the links inside the analysis area, and (iii) the presence of unsignalized intersections along the links, which can impact their effective block length to be included in the analytical model. All of these issues will be discussed in further detail in the next section.

The next subsection will present another set of estimation results assuming that the values for the average block length and jam density parameters are already known for all networks and the method only will estimate the 5 remaining model parameters to fit the analytical MFD to the empirical data.

Estimation by substituting the measured average block length and jam density

The UTD19 dataset further provides the location of loop detectors, the distance of loop detectors to the downstream traffic signal and the length of their corresponding block. Among the SMoC model parameters, the only parameter that we can readily

compute using the available data is the actual average block length, ℓ. For each network, ℓ can been computed as the average length of unique links with valid loop detectors.

The jam density can be defined as the traffic density when all the vehicles are jammed on the road bumper-to-bumper. Assuming that the average vehicle length across different urban networks does not vary significantly, the jam density can be considered as a universal constant that does not vary from a link to another link or a network to another network. Considering that average vehicle length is about 5 meters and allowing extra spacing about 1 meter between the vehicles, the universal jam density would approximately be 160 vehicles per kilometer.

Now that we have approximate knowledge on the values of average block length and jam density parameters, we intend to revisit the parameter estimation results by replacing the the values of these 2 parameters in the model and estimating only the remaining 5 model parameters. This will result in a reduction of the degree of freedom of the estimation method and thus a less optimum fit compared to the previous results. However, if the replaced parameter values are computed precisely based on the ground truth information, it will result in more accurate estimations for the other parameters, which would also be closer to the real values. On the contrary, if the replaced values are not accurate, they can put the estimation on the wrong track and result in inaccurate estimated values, which will not match up to the real values.

The MFD CDFs and cuts fitted by estimating the 5 other model parameters are shown in Fig. 6.6. The new graphical fits look different than the ones in the previous subsection, which can also be told by the change in the λ and ρ values that are the most impactful parameters on the shape of the analytical MFD CDF. The estimated parameter values are given in Table 6.5, where all the newly-estimated parameter

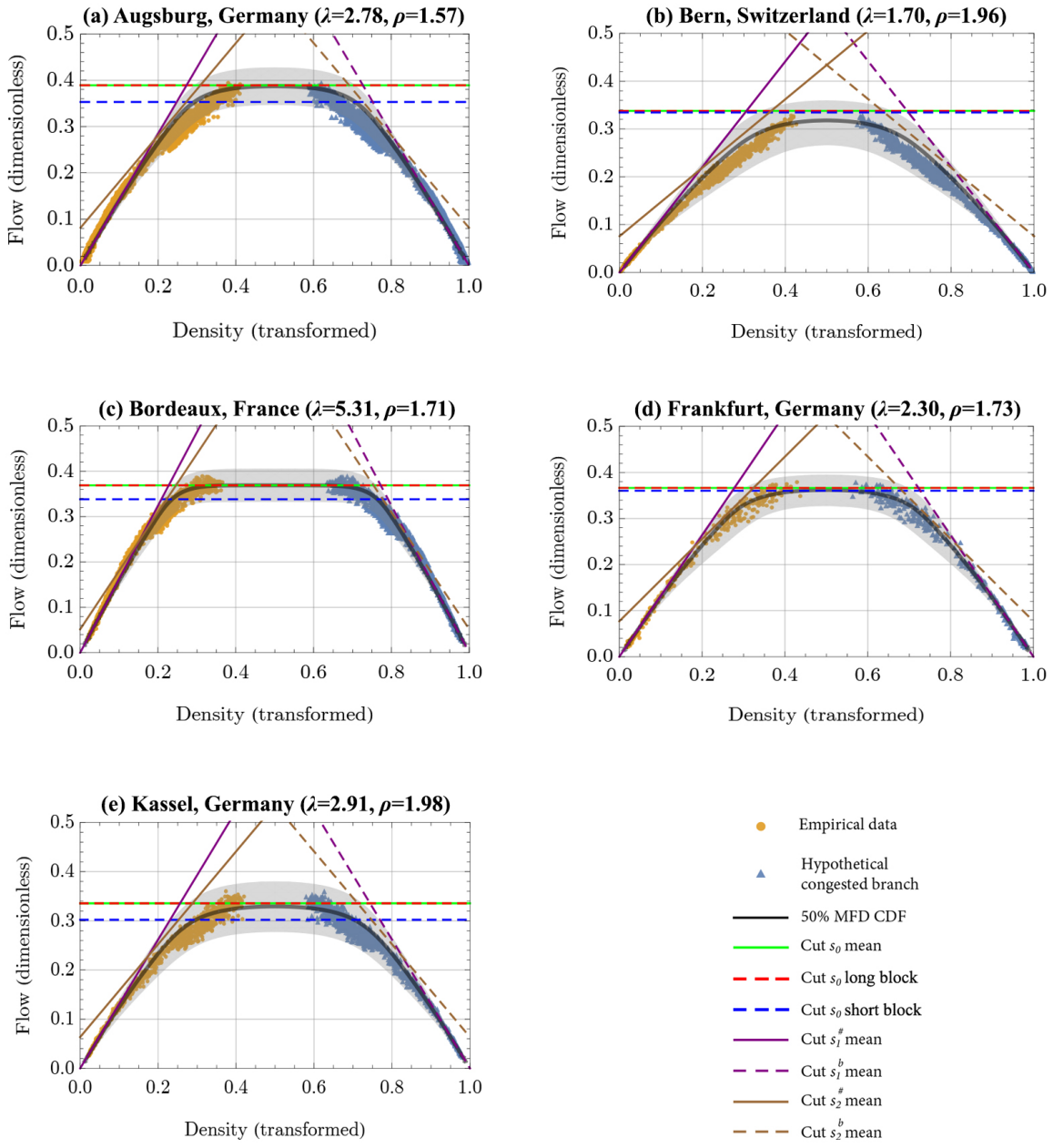

Figure 6.6: Estimated analytical MFDs and cuts using only 5 model parameters for the same networks as Fig. 6.4 displayed in the normalized flow units with respect to the estimated link capacity, Q, and the transformed and normalized density, k', as discussed in subsection 6.3.1. Note that the yellow markers show the empirical MFD points and the blue markers represent the hypothetical congested branch mirrored with respect to the line $k' = 0.5$. The solid black line is the 50% CDF and the shaded area shows the 16% and 84% CDF limits for the corresponding density values.

Table 6.5: Estimated parameter values for estimating only 5 model parameters

	λ	μ_ℓ [m]	μ_g [s]	ρ	Q [veh/h]	K [veh/km]	θ	δ
Augsburg [234.2]	2.78	202.6	31.5 (54.63)	1.57 (21.35)	1331.9 (140.6)	160	2.97 (92.43)	0.41 (37.71)
Bern [911.4]	1.70	233.5	59.8 (35.85)	1.96 (86.65)	1304.7 (21.01)	160	2.53 (18.23)	0.28 (6.54)
Bordeaux [4020.5]	5.31	358.2	27.3 (53.4)	1.71 (28.25)	1423.0 (97.99)	160	3.95 (67.96)	0.39 (15.07)
Frankfurt [792.6]	2.3	225.6	41.2 (7.51)	1.73 (6.41)	1375.2 (12.30)	160	3.31 (10.17)	0.29 (3.79)
Kassel [2160.9]	2.91	298.1	49.0 (7.36)	1.98 (13.93)	1203.4 (23.89)	160	2.73 (10.87)	0.41 (13.77)

Maximum log likelihood value in brackets
t-statistics in parentheses
μ_ℓ and K values are known a posteriori and are not estimated

values are statistically significant. However, most of these values are different than the previously-estimated values, as it was expected.

Unfortunately, the available data do not include further information about most of the model parameters, e.g. signal timing settings, and we are not able to compare the estimated values to their corresponding real values and verify which of the presented results conform more to the ground truth. By comparing the two sets of results, we can conclude that the proposed MLE method can be highly sensitive to the imposed assumptions or a posteriori information. For instance, many loop detectors are located close to the downstream end of the links and their measurements might not be representative of the average traffic condition in the whole link. Therefore, the average of link lengths used as the average block length parameter, μ_ℓ, in the second set of estimations might need an adjustment to take the location of loop detectors within the links into account.

6.5 Discussion

As demonstrated in the previous section, for the corridor case, the proposed method results in a well-fitted analytical MFD with mostly plausible parameter estimates, which are consistent with the ground truth. This shows that the underlying analytical MFD estimation method, SMoC, is well-suited for estimation of corridor MFDs. On the other hand, for the network case, we have observed that for many cities the estimated values for one or more parameters lie on the predefined parameter boundaries and therefore, we have deemed them as unsatisfactory and only presented the results from a few cities, where this issue is not encountered. Nevertheless, this shortcoming of the proposed method is a major issue, which needs to be addressed.

Furthermore, we do not have enough information about the real values of many of the model parameters for the studied networks. This prevents us from comparing the estimated and real-life parameter values to reach a solid conclusion about the performance of the proposed estimation method and the accuracy of the estimated values. The two different sets of results in the previous section, shown in Tables 6.4 and 6.5, highlight the sensitivity of the estimation results on any prior assumption about the value of some of the model parameters. Hence, in this section, we will discuss the possible reasons resulting in unsatisfactory estimations and other issues and challenges with both measured and estimated parameter values. The discussions here will provide numerous future research directions.

6.5.1 Missing Capacity and Congested States in the Empirical Data

An important issue with the available network-wide data is that the empirical MFDs only cover a limited range of occupancy (density), resulting in a lack of the capacity state and the descending congested branch of the MFD for most cases, which is obvious in Fig. 6.2. We highly believe that this is one of the main reasons of unsatisfactory

estimation results, since the satisfactory results are mostly from the networks whose empirical MFDs cover wider ranges of occupancy values or demonstrate bending trend toward the descending congested branch.

The lack of a bending trend in the empirical MFD will prevent the proposed method from accurately locating the MFD cuts and estimating their associated parameter values. Therefore, the significance of many parameters in the model will drop and in order to maximize the value of the log-likelihood function, the CMLE method may choose the boundary values for the insignificant parameters. This would also violate the regularity conditions of the MLE method, which will prevent us from calculating the covariance matrix and making inferences from the estimation results. Considering that the parameter boundaries are selected in a way that any value beyond the boundaries would be inconsistent with the ground truth, any estimation result with one or more estimated values on the boundaries has been deemed as unsatisfactory.

The missing congested branch of the empirical MFDs is a very well-known issue demonstrated by many studies in the literature, however, further research is still required to identify the reasons behind this issue thoroughly. Of foremost interest is the behavior of the MFD at the capacity state and whether the network exhibits a continuous MFD or there is a breakdown in average flow values at the capacity state.

6.5.2 Distribution of Empirical Data Points

To gain further understanding of the possible reasons behind the unsatisfactory results, we want to look into the distribution of the average flow values at different density values. Fig. 6.7 presents the histograms of the average flow values at the vicinity of certain density values for several different cases. The histograms in the first row of this figure illustrate the distribution of average flow values at the vicinity of 3 different density values for the city of Augsburg, where the method has resulted

in plausible graphical fit and estimates as shown in Fig. 6.4(a). As it can be seen, the empirical values in the Augsburg dataset demonstrate bell-shaped distributions similar to the estimated analytical distributions shown by the red curves.

On the other hand, there are many counterexamples in the datasets from the networks with unsatisfactory estimations, where the empirical histograms do not conform to the expected bell-shaped analytical distribution of the average flow values. The second row of Fig. 6.7 shows 3 histograms from 3 cities with unsatisfactory results, where the empirical average flow histograms at the vicinity of certain average densities do not demonstrate bell-shaped curves, It can be observed that the histograms of empirical data in the second row. This might be due to a host of reasons including measurement errors in individual loop detectors or the fact that the empirical MFD points might not be representative of the evolution of traffic in the entire network due to the heterogeneous or sporadic distribution of the loop detectors across the network.

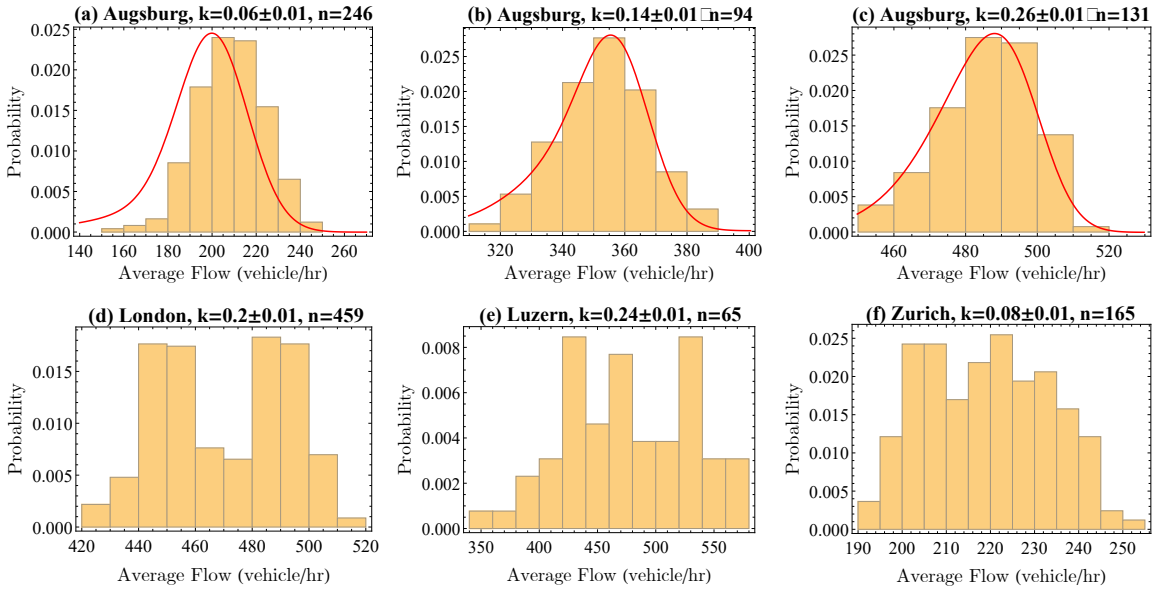

Figure 6.7: Empirical histograms and estimated analytical distributions (red curves) of the average flow values at the vicinity of certain density values for different cities.

6.5.3 Dependency of Empirical MFDs on Loop Detector Locations

While aggregation over a time interval decreases the dependency of flow measurements on the location of the loop detector along the links, the occupancy measurements will represent the occupancy at the loop detector's location rather than the entire link. The location of loop detectors is of essence to the shape of MFD and uneven distribution of loop detectors within the links may introduce significant bias in the MFD (Buisson and Ladier, 2009; Courbon and Leclercq, 2011). Fig. 6.8 illustrates the distribution of block lengths and distance of loop detectors to their respective downstream intersection for a number of studied networks.

The network map and location of loop detectors on the links for London and Marseille are shown in Fig. 6.9. It is evident that the loop detectors' distances to their downstream intersection are significantly lower than their respective block lengths, which is due to the fact that most loop detectors are placed just upstream of the intersections for signal control purposes, which will eventually result in an overestimation of link densities obtained by the loop detector occupancy measurements. On the other hand, if the loop detectors are located too close to the upstream end of the links, the queues formed on the links will rarely reach them making the loop detec-

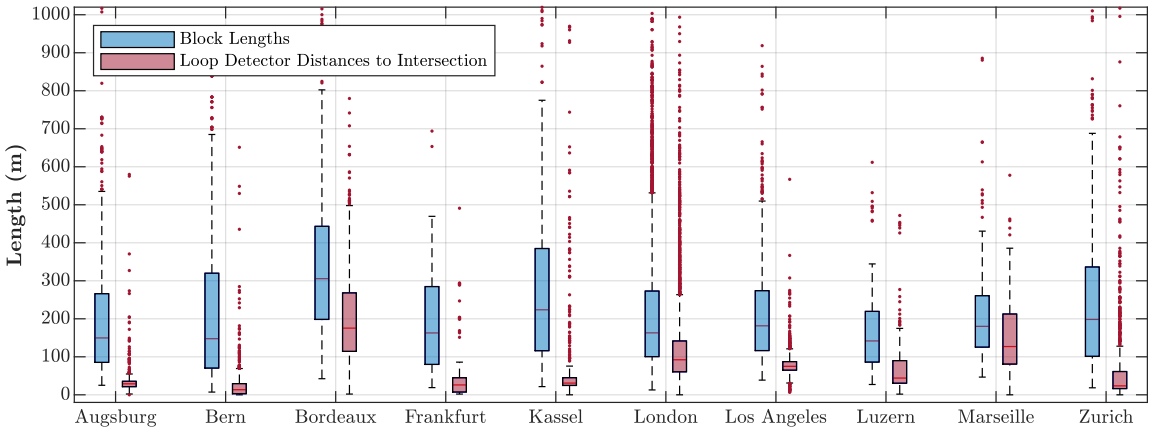

Figure 6.8: Box-and-whisker plots for block lengths and distances of loop detectors to downstream intersections for a number of the studied networks

Figure 6.9: Street network and loop detectors' location and coverage for (a) London and (b) Marseille [Network maps are obtained from OpenStreetMap (2020)]

tors unable to capture the congestion over the network. This issue needs foremost attention in future MFD research since it impacts the accuracy of both empirical and estimated analytical MFDs.

6.5.4 Empirical MFDs Not Representing the Entire Network

Fig. 6.9 illustrates the street network inside the analysis boundary for London and Marseille, where the bold links indicate the links with loop detectors. As expected, not all links in the network are equipped with loop detectors. While the loop detectors in London cover an interconnected system of major arterials, we observe that in Marseille, the loop detectors are placed in sparsely-distributed disjointed parcels across the analysis zone, which might have heterogeneous density distributions violating one of the main assumptions of the MFD existence. Even if the data shows homogeneous density distribution, the resulting MFD at best represents the subnetwork that we have loop detectors on. If the sample is carefully selected, e.g. London in this study, the resulting MFD can be a good proxy for the network MFD. Otherwise, it is not reasonable to conclude that the empirical MFD represents the average traffic flow of the entire network.

6.5.5 Block Length Measurement Considerations

Theoretically, the block length should be equal to the length of the uninterrupted segment of the approach upstream of an intersection, where the vehicles might need to make a full stop due to signal control. This length is clear and unique in the case of a grid network, where every intersection is signalized. However, in real-life networks, not all intersections are signalized. For example, we have minor unsignalized intersections, which are controlled by stop signs on the minor approach and no control on the major approach. An example from Zurich network is illustrated in Fig. 6.10. This figure shows a section of the Zurich network where there are six unsignalized inter-

sections with stop signs on the minor approaches along a link between two signalized intersections.

We can safely assume that the incoming vehicles from the minor approaches, which have to stop behind the stop signs, can enter the link only if there is sufficient headway in the major link stream. Hence, they will not affect the flow of the vehicles that were already on the main link and we can neglect the effect of incoming vehicles from those minor intersection. However, the vehicles going from the main link to the minor links will slow down and might affect the speed of the vehicles behind them if the main link has only one lane in the movement direction. Since the MFDs are mainly determined by the data from arterials having more than one lane in each direction, it seems safe to neglect the effect of turning movements from the major link at uncontrolled intersections on the flow of the major link. Therefore, it seems appropriate to neglect the minor intersections without any controls on the major link in block length calculations.

One of the main purposes of inclusion of block length in the model is to check the "short block" effect. Above a certain length, the block length does not seem to have significant impact on the speed of vehicles. In other words, if for a network we have $\ell^* = 200$ meters, a 300 meters link will not have a different average speed than a 700 meters link, hence the extra link length will not have any impact on the

Figure 6.10: Example from Zurich network with unsignalized intersections between two signalized intersections

average performance of the network and the shape of MFD. An upper limit for the block length may be needed to prevent the long blocks from skewing the mean block length. The validity of this idea and determining case-specific upper limits for the block length needs to be investigated using further link performance data.

6.5.6 Heterogeneities in Network Topology

The impacts of heterogeneity in the block lengths on the MFD has already been discussed in this section. The urban street networks include links from different functional classes, i.e. major and minor arterials, collectors, local streets. In general, the streets from lower functional classes have less lanes and lower speed limits and as a result lower operating speeds. It should not be expected to observe the same speed as arterials on the local streets. Therefore, MFDs cannot include both arterials and local streets, otherwise there will be huge variations in the speed and flow measurements, introducing high scatter to the MFD. Based on the placement of the loop detectors, we can conclude that the empirical MFDs represent the MFD of the grid of arterials inside the network, which are of foremost importance for network management and control purposes.

6.6 Chapter Summary

This chapter extends the SMoC to analytical estimation of network MFDs and proposes a method to estimate the parameters and shape of analytical MFD using empirical data in order to investigate the impact of different network topology and control characteristics on the shape of the MFD. The SMoC model has been selected as the base model after an exhaustive review of the present analytical MFD estimation models in the literature and the MLE method has been implemented to fit the analytical MFD and estimate its parameters for several datasets.

We have found that to extend the SMoC to estimate the analytical network MFDs, we need to redefine the mean red to green signal time, ρ, parameter in the original formulation as $\rho = \bar{n} - 1$, where $\bar{n}$ is the average number of signal phases for all the intersections in the study area. The parameter estimation method imposes constraints on the parameter space in order to keep the parameter estimates realistic. The lower and upper boundaries for each parameter are selected in a way that any values beyond them would not be attainable in real world.

Testing the proposed MLE method on the empirical data for a corridor yields that the estimates are very close to the measured values. The parameter estimation method is later used to estimate the analytical network MFD parameters for 30 networks using their empirical MFDs. While the estimation method results in statistically significant estimated parameter values for many of the studied datasets, there are also many networks where at least one parameter estimate is on the imposed boundaries. Such results have been considered as unsatisfactory and the possible reasons leading to this issue have been discussed in the next section. The satisfactory estimation results are revisited by assuming that the average block length and jam density values are already known a posteriori. The different parameter estimates and analytical MFD shapes indicate that the proposed method can be highly sensitive to the input values and any assumptions made throughout the process.

The observed discrepancies and limitations in the results are discussed in full length in section 5. Many of these issues most probably stem from the shortcomings of the underlying empirical data. The distribution of the loop detectors within the links is an impactful factor considering the fact that the loop detectors measure the occupancy at the point that they are located, which might not represent the density of the link if they are placed too close to the beginning or end of the link. Furthermore, it is discussed that the empirical MFDs might not be a good representative for the network MFD due to the fact that the loop detectors do not cover all the links in a

network. The computation of block lengths and signal control parameters for networks are also discussed in this chapter. However, reaching clear methods to tackle these issues needs further consideration using additional empirical data, which should be addressed in future research efforts.

CHAPTER 7
CONCLUSIONS AND FUTURE RESEARCH

This dissertation contributes to the state of the art on macroscopic traffic flow modeling and control by (i) identifying the gaps in the MFD-based DTA literature, (ii) proposing a novel MFD-based DTA framework to fill the gap between the multi-reservoir and continuum-space DTA approaches, (iii) demonstrating the potential of the MFD to improve the efficiency of the standard emissions estimation methods, and (iv) extending an analytical MFD estimation method to large-scale urban networks and verifying it using network-wide empirical data from 41 cities around the world.

Similar to many other engineering problems, the fundamental matter in the application of macroscopic traffic flow models narrows down to a trade-off between the accuracy and efficiency. While more detailed link-level traffic flow models are able to accurately replicate the traffic behavior in each element of the network, they might suffer from prohibitive calibration and run times hindering their application, especially in real time. The MFD relating the average network traffic variables can come in handy for solving this issue by replacing all the links inside a region with a single entity in the modeling and analysis process and bringing down the computational costs. However, this will reduce the accuracy of the results to some extent by assuming that all the links inside a region will have the same traffic conditions at any time. Therefore, the entirety of this dissertation is based on the idea to develop efficient traffic flow models which yield accurate results conforming to real-life traffic behavior.

Toward this end, Chapters 2 through 4 of this dissertation explore the application of the MFD in DTA. Chapter 2 presents an extensive review on macroscopic DTA models and categorizes them to (i) discrete-space multi-reservoir, and (ii) continuum-space approaches. This review reveals the shortcomings and gaps in the literature,

the most important of which are (i) both modeling approaches have limitations and are incapable of capturing the complex traffic dynamics in large-scale networks, and (ii) the connection between the two approaches is not well understood. It has been asserted that the speed-density relationship in the continuum-space models can be interpreted as the local MFD of the area under consideration.

Building on this finding, Chapter 3 develops a new continuum-space DTA framework and proposes a semi-Lagrangian solution method partitioning the network into small regions with homogeneous traffic conditions, which is consistent with the premise that a well-defined discrete-space model can be interpreted as the numerical solution of a continuum-space model; i.e., when the size of regions tends to zero while keeping the total area constant, the discrete-space model tends to the numerical solution of the continuum-space model. This framework allows the distribution of origin and destination points across the network, incorporates a more realistic local speed-density function in the continuum space by utilizing the MFD of each region, and resolves the stalling problem of the discrete-space models by incorporating the supply and demand MFDs rather than using the speed MFD for each region. The numerical experiments indicate that the accuracy of the model can be increased by partitioning the network into a finer grid of regions. The proposed framework can be extended in a number of directions including but not limited to reservoirs of arbitrary shapes, different cost functions, and inhomogeneous city networks.

The main goal of the DTA models is to reduce the travel costs and mitigate the congestion as much as possible. To demonstrate such application of the proposed DTA framework, Chapter 4 develops a semi-DSO solution trying to minimize the total instantaneous travel time of the users in the network by reactively minimizing the total marginal cost of route choice decisions at each time step. Although this solution method will not result in the fully DSO solution with the minimum total experienced travel times of all users, the numerical experiments exhibit that compared

to the original RDUE solution in Chapter 3, the semi-DSO solution results in (i) more homogeneous distribution of density over the network, (ii) less congestion over time, and eventually, (iii) less total travel time and average delay. An important future research direction is to predictively incorporate the experienced travel times in the solution method to find the DSO solution with the minimum total actual user travel times and to compare the efficiency and accuracy of the proposed semi-DSO solution against.

The aggregated network representation can be implemented in other aspects of traffic flow modeling as well. Chapter 5 explores the application of the MFD in emissions estimation. This is yet an underdeveloped field of research and the limited studies have been reviewed in this chapter highlighting their advantages and disadvantages. In light of these findings, the possibility of implementing the MFD in EPA's MOVES model, which is the mandatory emissions estimation model for transportation projects in the US. The emissions estimation for a grid network has been conducted in 4 different levels: (1) lane, (2) link, (3) corridor, and (4) network in the MOVES model. The comparison between results reveals that the efficiency of the emissions estimation process can significantly improve if a more aggregated representation of the network is implemented without sacrificing much accuracy.

For the specific grid network, the corridor-level results, which have a more homogeneous traffic distribution compared to the entire network, give acceptable relative error rates of less than 10% for most of the components compared to the lane-level estimates, which can be computed about 5 times faster than the lane-level estimation. This finding yields that if the network can be appropriately partitioned into small regions with homogeneous traffic conditions, a similar framework to the proposed DTA framework in Chapter 3 can be developed to estimate the emissions from large-scale networks efficiently and accurately. This can open the door for real-time emissions

estimation and developing control strategies aiming to reduce emissions in real time by only knowing and managing the number of vehicles inside each region.

The previous chapters demonstrate applications of the MFD in traffic modeling and analysis assuming that MFDs are already known for the network and its regions. Chapter 6 argues that in reality, deriving the MFD using empirical or simulation approaches can become a cumbersome task, especially for large-scale networks, and the proposed analytical methods in the literature are originally conceived for estimating the corridor MFDs. In this chapter, an analytical corridor MFD estimation method, the SMoC, has been first extended to estimation of network MFDs and later has been tested by estimating the model parameters using empirical data from 41 urban networks around the globe. Although the results cannot be fully validated due to unavailability of additional information from the studied networks, they reveal the drawbacks and further considerations required in both empirical and analytical MFD estimation approaches, which can guide the focus of future research in this field. The analytical model can be further improved by including additional cuts to incorporate the impacts of other factors such as (i) the "short block" effect, and (ii) the "moving bottleneck" effect caused by operation of buses in the network, which usually move slower than the general traffic and make frequent stops.

REFERENCES

Affum, J. K., A. L. Brown, and Y. C. Chan (2003). "Integrating air pollution modelling with scenario testing in road transport planning: the TRAEMS approach". In: *Science of the total environment* 312.1-3, pp. 1–14.

Aghamohammadi, Rafegh and Jorge A. Laval (2020a). "A continuum model for cities based on the macroscopic fundamental diagram: A semi-Lagrangian solution method". In: *Transportation Research Part B: Methodological* 132, pp. 101–116.

Aghamohammadi, Rafegh and Jorge A. Laval (2020b). "Dynamic traffic assignment using the macroscopic fundamental diagram: A Review of vehicular and pedestrian flow models". In: *Transportation Research Part B: Methodological* 137, pp. 99–118.

Aghamohammadi, Rafegh and Jorge A. Laval (2021). "Parameter Estimation of the Macroscopic Fundamental Diagram: A Maximum Likelihood Approach". In: *Preprints.org*, 2021100310.

Aghamohammadi, Rafegh, Ludovic Leclercq, and Jorge A. Laval (2020). "A Semi-Lagrangian Dynamic System Optimum Traffic Assignment model in Continuum Space". In: 99^{th} *Annual meeting of Transportation Research Board*. Washington, D.C.

Ahn, Kyoungho et al. (2002). "Estimating vehicle fuel consumption and emissions based on instantaneous speed and acceleration levels". In: *Journal of transportation engineering* 128.2, pp. 182–190.

Ambühl, Lukas et al. (2018). "A functional form with a physical meaning for the macroscopic fundamental diagram". In: *Transportation Research Part B: Methodological*.

Amirgholy, Mahyar, Mehrdad Shahabi, and H Oliver Gao (2017). "Optimal design of sustainable transit systems in congested urban networks: A macroscopic approach". In: *Transportation Research Part E: Logistics and Transportation Review* 103, pp. 261–285.

Ardekani, Siamak and Robert Herman (1987). "Urban Network-Wide Traffic Variables and Their Relations". In: *Transportation Science* 21.1, pp. 1–16.

Azar, Christian and Daniel J. A. Johansson (2012). "On the relationship between metrics to compare greenhouse gases–the case of IGTP, GWP and SGTP". In: *Earth System Dynamics* 3.2, pp. 139–147.

Batista, SFA, Ludovic Leclercq, and Nikolas Geroliminis (2019). "Estimation of regional trip length distributions for the calibration of the aggregated network traffic models". In: *Transportation Research Part B: Methodological* 122, pp. 192–217.

Boyce, David E., Bin Ran, and Larry J. Leblanc (1995). "Solving an Instantaneous Dynamic User-Optimal Route Choice Model". In: *Transportation Science* 29.2, pp. 128–142.

Buisson, Christine and Cyril Ladier (2009). "Exploring the Impact of Homogeneity of Traffic Measurements on the Existence of Macroscopic Fundamental Diagrams". In: *Transportation Research Record* 2124.1, pp. 127–136.

Cao, Jin and Monica Menendez (2015). "System dynamics of urban traffic based on its parking-related-states". In: *Transportation Research Part B: Methodological* 81, pp. 718–736.

Carey, Malachy and Mark McCartney (2004). "An exit-flow model used in dynamic traffic assignment". In: *Computers & Operations Research* 31.10, pp. 1583–1602.

Castrillón, Felipe and Jorge A. Laval (2018). "Impact of buses on the macroscopic fundamental diagram of homogeneous arterial corridors". In: *Transportmetrica B: Transport Dynamics* 6.4, pp. 286–301.

Chabini, Ismail (1998). "Discrete Dynamic Shortest Path Problems in Transportation Applications: Complexity and Algorithms with Optimal Run Time". In: *Transportation Research Record: Journal of the Transportation Research Board* 1645, pp. 170–175.

Courbon, Thomas and Ludovic Leclercq (2011). "Cross-comparison of macroscopic fundamental diagram estimation methods". In: *Procedia-Social and Behavioral Sciences* 20, pp. 417–426.

Csikós, Alfréd, Tamás Tettamanti, and István Varga (2015). "Macroscopic modeling and control of emission in urban road traffic networks". In: *Transport* 30.2, pp. 152–161.

Csikós, Alfréd and István Varga (2011). "Real-time estimation of emissions emerging from motorways based on macroscopic traffic data". In: *Acta Polytechnica Hungarica* 8.6, pp. 95–110.

Csikós, Alfréd and István Varga (2012). "Real-time modeling and control objective analysis of motorway emissions". In: *Procedia-Social and Behavioral Sciences* 54, pp. 1027–1036.

Daamen, Winnie and Serge Hoogendoorn (2003). "Experimental Research of Pedestrian Walking Behavior". In: *Transportation Research Record: Journal of the Transportation Research Board* 1828, pp. 20–30.

Daganzo, Carlos F. (1994a). "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory". In: *Transportation Research Part B: Methodological* 28.4, pp. 269–287.

Daganzo, Carlos F. (1994b). "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory". In: *Transportation Research Part B* 28.4, pp. 269–287.

Daganzo, Carlos F. (1995). "The cell transmission model, part II: Network traffic". In: *Transportation Research Part B: Methodological* 29.2, pp. 79–93.

Daganzo, Carlos F. (2005a). "A Variational Formulation of Kinematic Wave Theory: basic theory and complex boundary conditions". In: *Transportation Research Part B* 39.2, pp. 187–196.

Daganzo, Carlos F. (2005b). "A variational formulation of kinematic waves: basic theory and complex boundary conditions". In: *Transportation Research Part B: Methodological* 39.2, pp. 187–196.

Daganzo, Carlos F. (2005c). "A variational formulation of kinematic waves: Solution methods". In: *Transportation Research Part B: Methodological* 39.10, pp. 934–950.

Daganzo, Carlos F. (2007). "Urban gridlock: Macroscopic modeling and mitigation approaches". In: *Transportation Research Part B: Methodological* 41.1, pp. 49–62.

Daganzo, Carlos F. and Nikolas Geroliminis (2008). "An analytical approximation for the macroscopic fundamental diagram of urban traffic". In: *Transportation Research Part B: Methodological* 42.9, pp. 771–781.

Daganzo, Carlos F. and Lewis J Lehe (2016). "Traffic flow on signalized streets". In: *Transportation Research Part B: Methodological* 90, pp. 56–69.

Daganzo, Carlos F., W. Lin, and J. Del Castillo (1997). "A simple physical principle for the simulation of freeways with special lanes and priority vehicles". In: *Trans. Res. B* 2.31, pp. 105–125.

Dijkstra, Edsger W (1959). "A note on two problems in connexion with graphs". In: *Numerische mathematik* 1.1, pp. 269–271.

Du, Jie et al. (2013). "Revisiting Jiang's dynamic continuum model for urban cities". In: *Transportation Research Part B: Methodological* 56, pp. 96–119.

Du, Jie et al. (2015). "Reformulating the Hoogendoorn-Bovy predictive dynamic user-optimal model in continuum space with anisotropic condition". In: *Transportation Research Part B: Methodological* 79, pp. 189–217.

EPA (2018). "Fast Facts: US Transportation Sector Greenhouse Gas Emissions". In:

EPA (2020a). *MOVES3 technical guidance: Using MOVES to prepare emission inventories for state implementation plans and transportation conformity*. Tech. rep. United States Encironmental Protection Agency.

EPA (2020b). "US transportation sector greenhouse gas emissions: 1990–2018". In: *Office of Transportation and Air Quality EPA-420-F-20-037*.

Fosgerau, Mogens (2015). "Congestion in the bathtub". In: *Economics of Transportation* 4.4, pp. 241–255.

Friesz, Terry L. et al. (1989). "Dynamic Network Traffic Assignment Considered as a Continuous Time Optimal Control Problem". In: *Operations Research* 37.6, pp. 893–901.

Gayah, Vikash V, Xueyu Shirley Gao, and Andrew S Nagle (2014). "On the impacts of locally adaptive signal control on urban network stability and the macroscopic fundamental diagram". In: *Transportation Research Part B: Methodological* 70, pp. 255–268.

Geroliminis, Nikolas and Burak Boyacı (2012). "The effect of variability of urban systems characteristics in the network capacity". In: *Transportation Research Part B: Methodological* 46.10, pp. 1607–1623.

Geroliminis, Nikolas and Carlos F. Daganzo (2008). "Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings". In: *Transportation Research Part B: Methodological* 42.9, pp. 759–770.

Geyer, Charles J. et al. (1994). "On the asymptotics of constrained M-estimation". In: *The Annals of Statistics* 22.4, pp. 1993–2010.

Godfrey, JW (1969). "The mechanism of a road network". In: *Traffic Engineering & Control* 8.8.

Gomes, Gabriel and Roberto Horowitz (2006). "Optimal freeway ramp metering using the asymmetric cell transmission model". In: *Transportation Research Part C: Emerging Technologies* 14.4, pp. 244–262.

Haddad, Jack (2017). "Optimal perimeter control synthesis for two urban regions with aggregate boundary queue dynamics". In: *Transportation Research Part B: Methodological* 96, pp. 1–25.

Haddad, Jack, Mohsen Ramezani, and Nikolas Geroliminis (2013). "Cooperative traffic control of a mixed network with two urban regions and a freeway". In: *Transportation Research Part B: Methodological* 54, pp. 17–36.

Helbing, Dirk and Péter Molnár (1995). "Social force model for pedestrian dynamics". In: *Phys. Rev. E* 51 (5), pp. 4282–4286.

Herman, Robert and Siamak Ardekani (1984). "Characterizing traffic conditions in urban areas". In: *Transportation Science* 18.2, pp. 101–140.

Herman, Robert and Ilya Prigogine (1979). "A two-fluid approach to town traffic". In: *Science* 204.4389, pp. 148–151.

Hoogendoorn, Serge P. and Piet H. L. Bovy (2004). "Dynamic user-optimal assignment in continuous time and space". In: *Transportation Research Part B: Methodological* 38.7, pp. 571–592.

Hoogendoorn, Serge P. et al. (2015). "Continuum theory for pedestrian traffic flow: Local route choice modelling and its implications". In: *Transportation Research Part C: Emerging Technologies* 59, pp. 183–197.

Huang, Ling et al. (2009a). "Dynamic continuum model for bi-directional pedestrian flows". In: *Proceedings of the Institution of Civil Engineers - Engineering and Computational Mechanics* 162.2, pp. 67–75.

Huang, Ling et al. (2009b). "Revisiting Hughes' dynamic continuum model for pedestrian flow and the development of an efficient solution algorithm". In: *Transportation Research Part B: Methodological* 43.1, pp. 127–141.

Hughes, Roger L. (2002). "A continuum theory for the flow of pedestrians". In: *Transportation Research Part B: Methodological* 36.6, pp. 507–535.

Hänseler, Flurin S. et al. (2014). "A macroscopic loading model for time-varying pedestrian flows in public walking areas". In: *Transportation Research Part B: Methodological* 69, pp. 60–80.

Hänseler, Flurin S. et al. (2017). "A dynamic network loading model for anisotropic and congested pedestrian flows". In: *Transportation Research Part B: Methodological* 95, pp. 149–168.

Ingole, Deepak, Guilhem Mariotte, and Ludovic Leclercq (2020). "Minimizing network-wide emissions by optimal routing through inner-city gating". In: *Transportation Research Part D: Transport and Environment* 86, p. 102411.

Inrix (2020). *2019 Traffic Scorecard*. Retrieved from http://www.inrix.com/scorecard/. Accessed 14 August 2020.

Ji, Yangbeibei et al. (2010). "Investigating the shape of the macroscopic fundamental diagram using simulation data". In: *Transportation Research Record* 2161.1, pp. 40–48.

Ji, Yuxuan and Nikolas Geroliminis (2012). "On the spatial partitioning of urban transportation networks". In: *Transportation Research Part B: Methodological* 46.10, pp. 1639–1656.

Jiang, Yan Q., Pei Jie Ma, and Shu Guang Zhou (2018). "Macroscopic modeling approach to estimate traffic-related emissions in urban areas". In: *Transportation Research Part D: Transport and Environment* 60, pp. 41–55.

Jiang, Yan Q. and Peng Zhang (2011). "Macroscopic simulation of pedestrian flow through a bottleneck". In: *Applied Mechanics and Materials* 97-98, pp. 1168–1175.

Jiang, Yan Q., Wei Zhang, and Shu G. Zhou (2016a). "Comparison study of the reactive and predictive dynamic models for pedestrian flow". In: *Physica A: Statistical Mechanics and its Applications* 441, pp. 51–61.

Jiang, Yan Q. and Shu G. Zhou (2014). "Macroscopic simulation of traffic flow on continuum urban networks". In: *Applied Mechanics and Materials* 641-642, pp. 887–891.

Jiang, Yan Q. et al. (2009). "A reactive dynamic continuum user equilibrium model for bi-directional pedestrian flows". In: *Acta Mathematica Scientia* 29.6, pp. 1541–1555.

Jiang, Yan Q. et al. (2010). "A higher-order macroscopic model for pedestrian flows". In: *Physica A: Statistical Mechanics and its Applications* 389.21, pp. 4623–4635.

Jiang, Yan Q. et al. (2011). "A dynamic traffic assignment model for a continuum transportation system". In: *Transportation Research Part B: Methodological* 45.2, pp. 343–363.

Jiang, Yan Q. et al. (2012). "Numerical simulation of a continuum model for bi-directional pedestrian flow". In: *Applied Mathematics and Computation* 218.10, pp. 6135–6143.

Jiang, Yan Q. et al. (2016b). "Macroscopic modeling of pedestrian flow based on a second-order predictive dynamic model". In: *Applied Mathematical Modelling* 40.23-24, pp. 9806–9820.

Jin, Wenlong (2012). "A kinematic wave theory of multi-commodity network traffic flow". In: *Transportation Research Part B: Methodological* 46.8, pp. 1000–1022.

Jin, Wenlong (2017). "A Riemann solver for a system of hyperbolic conservation laws at a general road junction". In: *Transportation Research Part B: Methodological* 98, pp. 21–41.

Jin, Wenlong and R Jayakrishnan (2005). "First in, first out properties of a commodity-based kinematic wave simulation model". In: *Transportation Research Record: Journal of the Transportation Research Board* 1934, pp. 197–207.

Jin, Wenlong and H. M. Zhang (2003). "On the distribution schemes for determining flows through a merge". In: *Transportation Research Part B: Methodological* 37.6, pp. 521–540.

Knoop, Victor L., David De Jong, and Serge P Hoogendoom (2014). "Influence of road layout on network fundamental diagram". In: *Transportation Research Record* 2421.1, pp. 22–30.

Knoop, Victor L. et al. (2018). "Empirical MFDs using google traffic data". In: *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. Ieee, pp. 3832–3839.

Kouvelas, Anastasios, Mohammadreza Saeedmanesh, and Nikolas Geroliminis (2017). "Enhancing model-based feedback perimeter control with data-driven online adaptive optimization". In: *Transportation Research Part B: Methodological* 96, pp. 26–45.

Kurganov, Alexander, Sebastian Noelle, and Guergana Petrova (2001). "Semidiscrete central-upwind schemes for hyperbolic conservation laws and Hamilton–Jacobi equations". In: *SIAM Journal on Scientific Computing* 23.3, pp. 707–740.

Lamotte, Raphaël and Nikolas Geroliminis (2017). "The morning commute in urban areas with heterogeneous trip lengths". In: *Transportation Research Part B: Methodological*.

Laval, Jorge A. (2009). "Graphical solution and continuum approximation for the single destination dynamic user equilibrium problem". In: *Transportation Research Part B: Methodological* 43.1, pp. 108–118.

Laval, Jorge A. and Felipe Castrillón (2015). "Stochastic approximations for the macroscopic fundamental diagram of urban networks". In: *Transportation Research Part B: Methodological* 81, pp. 904–916.

Laval, Jorge A. and Bhargava R. Chilukuri (2016). "Symmetries in the kinematic wave model and a parameter-free representation of traffic flow". In: *Transportation Research Part B: Methodological* 89, pp. 168–177.

Laval, Jorge A., Guillaume Costeseque, and Bhargava R. Chilukuri (2016). "The impact of source terms in the variational representation of traffic flow". In: *Transportation Research Part B: Methodological* 94, pp. 204–216.

Laval, Jorge A. and Ludovic Leclercq (2013). "The Hamilton-Jacobi partial differential equation and the three representations of traffic flow". In: *Transportation Research Part B* 52, pp. 17–30.

Laval, Jorge A., Ludovic Leclercq, and Nicolas Chiabaut (2017). "Minimal parameter formulations of the dynamic user equilibrium using macroscopic urban models: Freeway vs city streets revisited". In: *Transportation Research Part B: Methodological* 23, pp. 517–530.

Laval, Jorge A. et al. (2015). "Real-time congestion pricing strategies for toll facilities". In: *Transportation Research Part B: Methodological* 71, pp. 19–31.

Leclercq, Ludovic (2007). "Hybrid approaches to the solutions of the Lighthill-Whitham-Richards model". In: *Transportation Research Part B: Methodological* 41.7, pp. 701–709.

Leclercq, Ludovic and Nikolas Geroliminis (2013). "Estimating MFDs in simple networks with route choice". In: *Procedia-Social and Behavioral Sciences* 80, pp. 99–118.

Leclercq, Ludovic, Jorge A. Laval, and Estelle Chevallier (2007). "The Lagrangian coordinate system and what it means for first order traffic flow models". In: *17th International Symposium on Transportation and Traffic Theory*. Ed. by B Heydecker, M Bell, and R Allsop. Elsevier, New York.

Leclercq, Ludovic, Almeria Senecat, and Guilhem Mariotte (2017a). "Dynamic macroscopic simulation of on-street parking search: A trip-based approach". In: *Transportation Research Part B: Methodological* 101, pp. 268–282.

Leclercq, Ludovic, Almeria Senecat, and Guilhem Mariotte (2017b). "Dynamic macroscopic simulation of on-street parking search: A trip-based approach". In: *Transportation Research Part B: Methodological* 101, pp. 268–282.

Lejri, Delphine et al. (2018). "Accounting for traffic speed dynamics when calculating COPERT and PHEM pollutant emissions at the urban scale". In: *Transportation research part D: Transport and Environment* 63, pp. 588–603.

Lighthill, M. J. and G. B. Whitham (1955). "On Kinematic Waves. II. A Theory of Traffic Flow on Long Crowded Roads". In: *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* 229.1178, pp. 317–345.

Lin, Z.Y. et al. (2017). "A predictive continuum dynamic user-optimal model for a polycentric urban city". In: *Transportmetrica B: Transport Dynamics* 5.3, pp. 233–252.

Liu, Bin and H Christopher Frey (2013). "Development of a simplified version of MOVES and application to iterative case studies". In: *106th Annual Meeting of the Air & Waste Management Association Chicago, IL, USA*.

Liu, Haobing et al. (2016). "Improved energy and emissions modeling for project evaluation (MOVES-Matrix)". In: *UC Davis: National Center for Sustainable Transportation*.

Loder, Allister et al. (2019). "Understanding traffic capacity of urban networks". In: *Scientific reports* 9.1, pp. 1–10.

Long, Jiancheng et al. (2017). "A dynamic taxi traffic assignment model: A two-level continuum transportation system approach". In: *Transportation Research Part B: Methodological* 100, pp. 222–254.

Lopez, Clélia et al. (2017). "Spatiotemporal Partitioning of Transportation Network Using Travel Time Data". In: *Transportation Research Record: Journal of the Transportation Research Board* 2623, pp. 98–107.

Lopez, Pablo Alvarez et al. (2018). "Microscopic traffic simulation using SUMO". In: *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, pp. 2575–2582.

Mahmassani, Hani S, James C Williams, and Robert Herman (1984). "Investigation of network-level traffic flow relationships: some simulation results". In: *Transportation Research Record* 971, pp. 121–130.

Merchant, Deepak K. and George L. Nemhauser (1978). "A Model and an Algorithm for the Dynamic Traffic Assignment Problems". In: *Transportation Science* 12.3, pp. 183–199.

Muñoz, JC and Jorge A. Laval (2005). "System Optimum Dynamic Traffic Assignment Graphical Solution Method for a Congested Freeway and One Destination". In: *Transportation Research Part B* 40.1, pp. 1–15.

Newell, Gordon Frank (1980). *Traffic flow on transportation networks*. Monograph.

Ntziachristos, Leonidas et al. (2009). "COPERT: a European road transport emission inventory model". In: *Information technologies in environmental engineering*. Springer, pp. 491–504.

OpenStreetMap Contributors (2020). *Planet dump*. Retrieved from https://planet.osm.org. Accessed 17 May 2020.

Panis, Luc Int, Steven Broekx, and Ronghui Liu (2006). "Modelling instantaneous traffic emission and the influence of traffic speed limits". In: *Science of the total environment* 371.1-3, pp. 270–285.

Peeta, Srinivas and Athanasios K. Ziliaskopoulos (2001). "Foundations of Dynamic Traffic Assignment: The Past, the Present and the Future". In: *Networks and Spatial Economics* 1.3, pp. 233–265.

Rakha, Hesham, Kyoungho Ahn, and Antonio Trani (2003). "Comparison of MOBILE5a, MOBILE6, VT-MICRO, and CMEM models for estimating hot-stabilized light-duty gasoline vehicle emissions". In: *Canadian Journal of Civil Engineering* 30.6, pp. 1010–1021.

Ramezani, Mohsen, Jack Haddad, and Nikolas Geroliminis (2015). "Dynamics of heterogeneity in urban networks: aggregated traffic modeling and hierarchical control". In: *Transportation Research Part B: Methodological* 74, pp. 1–19.

Richards, Paul I (1956). "Shock waves on the highway". In: *Operations research* 4.1, pp. 42–51.

Romero, Luis M. and Francisco G. Benitez (2010). "Traffic flow continuum modeling by hypersingular boundary integral equations". In: *International Journal for Numerical Methods in Engineering* 82.1, pp. 47–63.

Saedi, Ramin et al. (2020). "Comparison of Support Vector and Non-Linear Regression Models for Estimating Large-Scale Vehicular Emissions, Incorporating Network-Wide Fundamental Diagram for Heterogeneous Vehicles". In: *Transportation Research Record* 2674.5, pp. 70–84.

Saeedmanesh, Mohammadreza and Nikolas Geroliminis (2017). "Dynamic clustering and propagation of congestion in heterogeneously congested urban traffic net-

works". In: *Transportation Research Part B: Methodological* 105.Supplement C, pp. 193–211.

Saumtally, Tibye, Jean Patrick Lebacque, and Habib Haj-Salem (2013). "A dynamical two-dimensional traffic model in an anisotropic network". In: *Networks and Heterogeneous Media* 8.3, pp. 663–684.

Shabihkhani, Rooholamin and Eric Gonzales (2014). "Macroscopic relationship between network-wide traffic emissions and fundamental properties of the network". In: *Transportation Research Circular E-C197. Proc., Symposium Celebrating*. Vol. 50, pp. 284–299.

Smeed, Reuben J. (1967). "The road capacity of city centers". In: *Highway Research Record* 169.

Smit, Robin, Richard Smokers, and Elke Rabé (2007). "A new modelling approach for road traffic emissions: VERSIT+". In: *Transportation Research Part D: Transport and Environment* 12.6, pp. 414–422.

Sossoe, Kwami Seyram and Jean Patrick Lebacque (2017). "Reactive Dynamic Assignment for a Bi-dimensional Traffic Flow Model". In: *Advances in Systems Science: Proceedings of the International Conference on Systems Science 2016 (ICSS 2016)*. Ed. by Jerzy Świkatek and Jakub M. Tomczak. Cham: Springer International Publishing, pp. 179–188.

Sossoe, Kwami Seyram et al. (2015). "Traffic flow within a two-dimensional continuum anisotropic network". In: *Transportation Research Procedia* 10, pp. 217–225.

Staniforth, Andrew and Jean Cote (1991). "Semi-Lagrangian Integration Schemes for Atmospheric Models - A Review". In: *Monthly Weather Review* 119.9, pp. 2206–2223.

Tao, Y. Z. et al. (2014). "Dynamic system-optimal traffic assignment for a city using the continuum modeling approach". In: *Journal of Advanced Transportation* 48.7, pp. 782–797.

Tilg, Gabriel, Sasan Amini, and Fritz Busch (2020). "Evaluation of analytical approximation methods for the macroscopic fundamental diagram". In: *Transportation Research Part C: Emerging Technologies* 114, pp. 1–19.

Tsubota, Takahiro, Ashish Bhaskar, and C Edward (2014). "Brisbane Macroscopic Fundamental Diagram: Emperical Findings on Network Partitioning and Incident Detection". In: *Trans Res Rec*.

Vickrey, William S (1969). "Congestion theory and transport investment". In: *The American Economic Review* 59.2, pp. 251–260.

Wang, Peng Fei et al. (2015). "An Empirical Analysis of Macroscopic Fundamental Diagrams for Sendai Road Networks". In: *Interdisciplinary Information Sciences* 21.1, pp. 49–61.

Weidmann, Ulrich (1993). "Transporttechnik der fußgänger: transporttechnische eigenschaften des fußgängerverkehrs, literaturauswertung". In: *IVT Schriftenreihe* 90.

Wong, S. C. et al. (2010). "Bidirectional Pedestrian Stream Model with Oblique Intersecting Angle". In: *Journal of Transportation Engineering* 136.3, pp. 234–242.

Xia, Yinhua, S. C. Wong, and Chi-Wang Shu (2009). "Dynamic continuum pedestrian flow model with memory effect". In: *Phys. Rev. E* 79 (6), p. 066113.

Xie, Xiaoyan, Nicolas Chiabaut, and Ludovic Leclercq (2013). "Macroscopic fundamental diagram for urban streets and mixed traffic: cross comparison of estimation methods". In: *Transportation research record* 2390.1, pp. 1–10.

Xiong, Tao et al. (2011). "High-Order Computational Scheme for a Dynamic Continuum Model for Bi-Directional Pedestrian Flows". In: *Computer-Aided Civil and Infrastructure Engineering* 26.4, pp. 298–310.

Yang, Kaidi, Monica Menendez, and Nan Zheng (2019). "Heterogeneity aware urban traffic control in a connected vehicle environment: A joint framework for congestion pricing and perimeter control". In: *Transportation Research Part C: Emerging Technologies* 105, pp. 439–455.

Yildirimoglu, Mehmet and Nikolas Geroliminis (2014). "Approximating dynamic equilibrium conditions with macroscopic fundamental diagrams". In: *Transportation Research Part B: Methodological* 70, pp. 186–200.

Yildirimoglu, Mehmet, Mohsen Ramezani, and Nikolas Geroliminis (2015). "Equilibrium analysis and route guidance in large-scale networks with MFD dynamics". In: *Transportation Research Part C: Emerging Technologies* 59, pp. 404–420.

Yin, Jun et al. (2013). "A continuum model for housing allocation and transportation emission problems in a polycentric city". In: *International Journal of Sustainable Transportation* 7.4, pp. 275–298.

Zheng, Nan, Guillaume Rérat, and Nikolas Geroliminis (2016). "Time-dependent area-based pricing for multimodal systems with heterogeneous users in an agent-

based environment". In: *Transportation Research Part C: Emerging Technologies* 62, pp. 133–148.

Ziliaskopoulos, A. K. et al. (2004). "Large-Scale Dynamic Traffic Assignment: Implementation Issues and Computational Analysis". In: *Journal of Transportation Engineering* 130.5, pp. 585–593.

VITA

Rafegh Aghamohammadi was born in Maku, Iran and raised in Urmia, Iran. At the age of 18, he moved to the Iranian capital city of Tehran to pursue his undergraduate studies at the Sharif University of Technology, after being ranked 22^{nd} in the nationwide undergraduate entrance exam among more than $350,000$ participants. He received his BS degree in Civil Engineering and MS degree in Highway and Transportation Engineering form Sharif in 2013 and 2015, respectively.

Rafegh joined Georgia Tech in April 2017 as a doctoral student in the Transportation Systems Engineering program of the School of Civil and Environmental Engineering under the supervision of Prof. Jorge Laval. He has also obtained a second MS degree in Computational Science Engineering from Georgia Tech during his PhD. His PhD research has been on macroscopic traffic flow modeling and its application in different aspects of traffic control such as dynamic traffic assignment and vehicular emissions estimation. This research has been funded by the National Science Foundation (NSF) and National Center for Sustainable Transportation (NCST) and the results of his work have been published and presented in several high-impact journals and prestigious conferences such as *Transportation Research Part B: Methodological*, *Transportation Research Part C: Emerging Technologies*, *International Symposium of Traffic and Transportation Theory (ISTTT)*, *Transportation Research Board (TRB) Annual Meetings*, and *International Symposium on Dynamic Traffic Assignment (DTA)*.

Rafegh has received several awards and fellowships during his studies at Georgia Tech, including: *President's Excellence in Graduate Studies Fellowship*, *Elizabeth B. and William E. Higginbotham Beyond the Classroom Experience Fellowship*, *CEE Future Faculty Fellowship*, and *ASHE Georgia Babs Abubakari Scholarship*. He has

also been serving as a reviewer for several transportation journals and conferences. During summer 2019, Rafegh spent 3 months as visiting researcher in the LICIT lab of the University of Lyon, France under supervision of Prof. Ludovic Leclercq.

Immediately after graduating from Georgia Tech, Rafegh will join the University of Michigan as a postdoctoral research fellow in the LIMOS lab under supervision of Prof. Yafeng Yin. His long-term career plan is to combine his traffic modeling background with the emerging technologies in the transportation realm to find innovative ways to deal with the future transportation systems and mitigate their undesirable consequences.