2021-04-23

Large scale flow modelling and control: a macroscopic fundamental diagram approach

Moshahedi, Nadia

UNIVERSITY OF CALGARY

Large scale flow modelling and control: a macroscopic fundamental diagram approach

by

Nadia Moshahedi

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN CIVIL ENGINEERING

CALGARY, ALBERTA
APRIL, 2021

Abstract

In recent decades, traffic congestion has become a major issue in traffic networks, especially in urban networks comprised of a set of short links and signalized intersections. To circumvent the issue, various traffic control and management strategies have been devised; however, the proposed strategies are rarely developed at network-wide level. Further, the modelling approach is based on microscopic models that cannot be adopted for centralized control or macroscopic models that have limited capacity to properly describe important phenomenon of traffic networks.

This thesis aims at modelling and control of a large-scale urban network comprised of multiple pockets of congestion. The modelling approach is based on macroscopic fundamental diagram (MFD), which assumes a well-defined relationship between average flow and average density for any traffic network with spatially homogeneous distribution of vehicles. This simpler representation of large-scale traffic networks using aggregated traffic variables facilitates a centralized and real-time control of urban networks. To use the system-wide benefits of MFD models, firstly, an anticipatory control scheme, integrating road users routing responses to the control model is advanced. The proposed anticipatory control approach is found to produce globally optimal solutions and move the network towards system optimum traffic condition. Thereafter, a proportionally fair control scheme that simultaneously enhances efficiency and fairness among road users is developed. The unique feature of the developed perimeter controller is consideration of road users' trip utility in the control model without much sacrificing efficiency for fairness.

Despite the computational advantages of aggregated MFD models, these models do not describe important phenomenon of traffic networks. In the final part of this thesis, the MFD dynamics is enhanced to capture multiple kinematic waves, congestion, and queuing with high precision and within reasonable computational effort. Further, an approach for incorporating connected/autonomous vehicles (CAV)s into MFD dynam-

ics is introduced; the network-wide effect of CAVs on network's traffic state is then investigated.

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my thesis advisor, Dr Lina Kattan, for giving me the opportunity to join her research group and pursue my Doctorate under her supervision. It was through her constant support and encouragement that I could finish my studies. During years of working together, I learnt a lot from her. She is not only a caring brilliant scholar but also a good friend. I feel extremely lucky to know her and work with her.

I am also very grateful to my co-supervisor, professor Richard Tay, whom I had the opportunity to discuss my research on several occasions. I would like to thank my committee members: Dr. Alexandre Gomez De Barros, Dr. Markus Dann, Dr. Alireza Sabouri, and Dr. Babak Mehran for taking time to review this dissertation. I would also like to thank Miss. Kate McGillis for her great job on editing this dissertation.

My deepest gratitude goes to my parents, Fereshteh and Ali, my sister, Nasim, and my brother, Nima, for their un-conditional love and continuous support not only during my PhD studies but through my whole life. Thank you for being there any time I needed you. Most importantly, I would like to express my gratitude to my husband, Mahdad, for his understanding, patience and support. Thank you for all that we had and we will have together.

This work is financially supported by Mitacs Accelerate PhD fellowship with Watt Consulting Group, Alberta Graduate Excellence Scholarship, Natural Sciences and Engineering Research Council of Canada (NSERC) (Grant No. RT735236: RGPIN/03942-2020), Discovery Accelerator Supplement (DAS) Grant on Urban Transportation Operation and Management in a Connected and Autonomous vehicle, Alberta Innovate Strategic Grant on Integrated Urban Mobility through Emerging Transportation Tech-

nologies (Grant No.10024748: G2018000894), and Urban Alliance Research Chair in Transportation Systems Optimization (Grant No. RT756247).

Table of Contents

List of Figures and Illustrations

List of Tables

List of Symbols, Abbreviations and Nomenclature

Abbreviations	Definition
AC	anticipatory control
BC	basic control
BLP	bi-level programming
CAV	connected autonomous vehicle
CBD	central business district
CTM	cell transmission model
DNL	dynamic network loading
DTA	dynamic traffic assignment
FC	fair control
IOA	iterative optimization and assignment
KWT	kinematic wave theory
LTM	link transmission model
LWR	Lighthill-Whitham-Richards
MFD	macroscopic fundamental diagram
MPC	model predictive control
MRDNL	multi reservoir dynamic network loading
MSA	method of successive averages
NC	no control
PC	perimeter control
PFPC	proportionally fair perimeter control
PQ	point queue
RM	ramp metering
SO	system optimum
SSO	stochastic system optimum
SUE	stochastic user equilibrium
SVG	sufficient variational graph
TTS	total time spent
UE	user equilibrium
V2V	vehicle 2 vehicle
V2I	vehicle 2 infrastructure
VT	variational theory

Symbol	Definition
I	set of time intervals
R	set of regions
W	set of OD region pairs
P_w	set of routes between OD region pairs w
H_r	set of neighbor regions of region $r \in R$
$G(.)$	MFD function (veh/s)
$P(.)$	travelling production function (veh.m/sec)
$E(.)$	travel time function (s)
$F(.)$	user equilibrium function (veh/s)
$R(.)$	route cost function
$r(.)$	link cost function
a, b, c	MFD defining parameters
α, β	MFD scaling parameters
n	accumulation (veh)
$\varepsilon(n)$	accumulation error term (veh)
n_{crt}	critical accumulation (veh)
n_{jam}	jam accumulation (veh)
n_T	transferring vehicle accumulation (veh)
n_Q	queuing vehicle accumulation (veh)
n_I	internal vehicle accumulation (veh)
$n_{Q_{max}}$	maximum queuing vehicle accumulation (veh)
N	cumulative accumulation (veh)
N_{in}	cumulative upstream accumulation (veh)
N_{out}	cumulative downstream accumulation (veh)
d	demand (veh/s)
$\varepsilon(d)$	demand noise term (veh/s)
f	flow (veh/s)
$\hat{f}$	restricted flow (veh/s)
f_{in}	inflow (veh/s)
$\hat{f}_{in}$	restricted inflow (veh/s)
f_{out}	outflow (veh/s)
$\hat{f}_{out}$	restricted outflow (veh/s)
ψ	internal flow (veh/s)
Ψ	internal accumulation (veh)
u	perimeter control rate
θ	regional split ratios
λ	route split ratios
δ	region-route incidence matrix
C	regional boundary capacity (veh/s)
$\tilde{C}$	regions' remaining capacity (veh/s)

χ	parameter defining critical point of receiving capacity functio
u_{min}	lower control bound
u_{max}	upper control bound
T	total simulation time (s)
T_c	control sampling time (s)
N_p	prediction horizon
k	density (veh/m)
k_{jam}	jam density (veh/m)
q	flow (veh/sec)
V	speed (veh/m)
$\bar{V}$	average regional speed (veh/s)
x	distance (m)
t	time (s)
$\bar{L}$	average regional trip length (m)
L	arterial length (m)
l	link length (m)
τ	regional travelling time (s)
τ_Q	regional queuing delay (s)
$\tilde{\tau}$	marginal regional travelling time (s)
$\mathcal{T}$	route travelling time (s)
$\tilde{\mathcal{T}}$	marginal route travelling time (s)
ϕ	random term indicating drivers' knowledge of travel time
ϵ	MSA step size
κ	threshold value for the convergence error of MSA
ξ	heterogeneity coefficient
η	lower range of heterogeneity coefficient parameter
ρ	number of lanes
ν	number of links
m	number of vehicle classes
γ	free-flow speed (veh/m)
ω	backward speed (veh/m)
q_c	maximum arterial capacity (veh/s)
c	traffic signal cycle (s)
g	traffic signal green time interval (s)
r	traffic signal red time interval (s)
o	traffic signal offset (s)
sf	sending flow (veh/s)
rc	receiving capacity (veh/s)
ζ	reduction factor on demanding inflow

Chapter 1

Introduction

1.1 Background

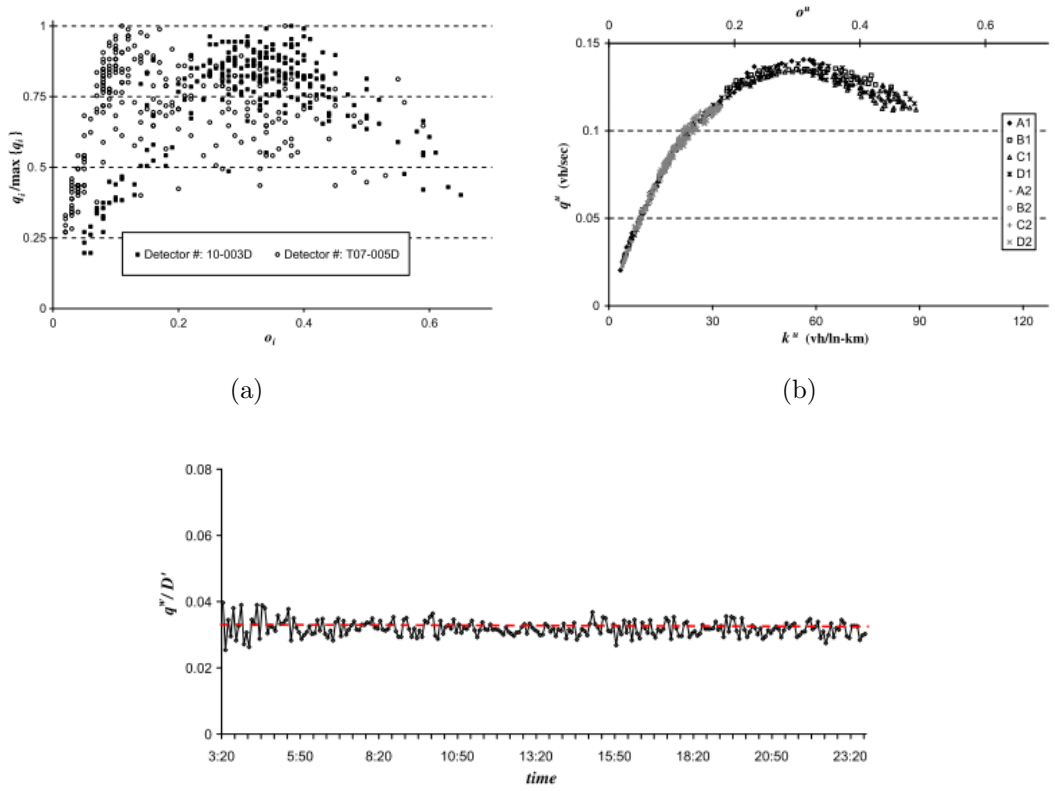
In recent decades, traffic congestion has become a major issue in traffic networks due to the rapid increase of urbanization, the upward trend of travel demand, and the limited supply of road network facilities. Building new urban infrastructure to mitigate or relieve traffic congestion is not a sustainable solution. Therefore, various traffic management and control schemes have been developed to relieve traffic congestion, improve mobility, enhance safety, and reduce emissions.

While various flow modelling and control schemes have been discussed in the literature for large-scale freeway traffic networks, there is a lack of research on developing such schemes for urban networks due to their specific structures that have short links and traffic signals. Local controllers locally control traffic networks, but a network-wide centralized controller based on a detailed microscopic or mesoscopic traffic model is not operational for real-time applications. These models are based on highly detailed link-level information, which is difficult to acquire and analyze in real time, thereby imposing a huge computational burden on the system. Furthermore, the unpredictability of road users' behaviour adds more complexity to the network-wide control problem. However, parsimonious or approximate aggregated traffic models are more computationally efficient compared to their detailed microscopic counterparts and can be used to solve the problem [3].

1.1.1 Macroscopic fundamental dagram

Macroscopic fundamental diagram (MFD) has emerged as a promising traffic modelling tool. [4] described the dynamics of a homogeneous urban network by considering it as a single reservoir where outflow is a function of the number of vehicles in the network. MFD provides a unimodal, demand-insensitive, and low-scatter relationship between network density or accumulation (number of vehicles in a network) and network space-mean flow or outflow (trip completion rate) in roughly homogeneous urban networks. MFD was initially proposed by [5] and was widely recognized due to the seminal work of [4]. Analytical estimation of MFD was done by [6] based on variational theory method introduced in [7, 8] for a single arterial consisting of a set of consecutive links separated by traffic signals and was later extended [9, 10].

[1] verified the existence of MFD in the Yokohamas (Japan) traffic network by integrating highly scattered plots of flow versus density obtained using a combination of fixed detectors and GPS-equipped taxis. Figure 1.1 (a) shows the scatter plots of flow vs. occupancy for a weekday collected every 5 minutes from two single detectors in the Yokohamas traffic network. The relationship between flow and occupancy is scattered and does not provide much detail about traffic conditions. Figure 1.1 (b) shows the aggregated plot of average flow vs. average density obtained from two sensors over two distinct days (labels A1-D2 denote different times in two days). The unimodal low-scatter relationship between flow and density suggests the existence of MFD. Since demand and origin-destination (OD) values vary substantially over the two days, we can conclude that MFD is demand-insensitive and independent of O-D. From field data, it was also found that the ratio of production to trip completion is a relatively constant value and equal to average trip length of the region (see Figure 1.1 (c)). For the link fundamental diagram, the relationship between flow and density is described at any point of links while for network MFD, the relationship is drawn between average flow and average density of a large-scale network comprising of hundred thousand of links.

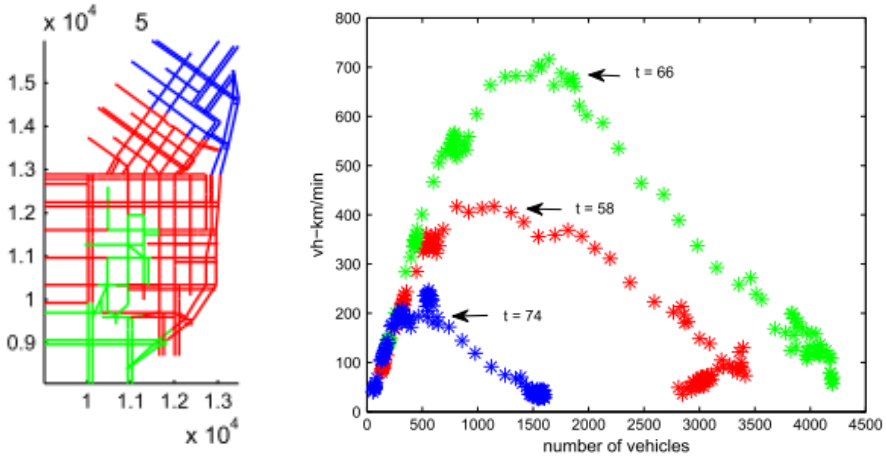


(a) (b)

(c)

Figure 1.1: Loop detector data from Yokohamas (Japan) traffic network: (a) flow vs. occupancy for two single detectors across a day (b) average flow vs. average density from all the detectors across two days (c) ratio of production to trip completion [1].

It was demonstrated that a well-defined, low-scatter MFD can only be identified for networks with a spatially homogeneous distribution of density. Heterogeneity undermines the existence of a well-defined MFD and reduces outflow [11, 12, 13]; therefore, to facilitate modelling and control of a heterogeneous network, it can be partitioned into several smaller homogeneous sub-regions with a well-defined MFD. From Figure 1.2 (a), it can be observed that traffic network is divided into three regions. This is done using clustering technique and based on spatial features of congestion in traffic networks by minimizing variance of link density within each cluster [2]. The estimated MFD for each region is shown in Figure 1.2 (b), in which t denotes the time that outflow is at its maximum level. Accumulation at this time corresponds to the critical accumulation.



(a) (b)

Figure 1.2: (a) Clustering result of a heterogeneous urban network (b) MFDs for three homogeneous regions [2].

1.1.2 Modeling and control based on MFD

Two basic approaches to model MFD exist: accumulation-based MFD [14, 15, 16, 17, 18] and trip-based MFD [19, 20, 21, 22]. Interested readers are referred to [23] for analytical and numerical comparison of the existing MFD models. In this thesis, we focus on accumulation-based modelling. In accumulation-based MFD, dynamics of a single reservoir is described by a simple conservation equation, where outflow is a function of total number of vehicles within the network. While this approach is attractive due to its simplicity, it relies on heavily simplifying assumptions, such as assumption of constant average trip length. In reality, the average trip length varies as traffic congestion propagates throughout an urban network and influences travelers' route choice. Also, accumulation-based models assume that the flow entering a reservoir from an upstream perimeter boundary instantly reaches the opposing downstream end with infinite speed. This later assumption is the cause of numerical viscosity observed in accumulation-based models.

Dynamic network loading (DNL) and dynamic traffic assignment (DTA) are two

important problems in transportation with major applications in network modelling, policy evaluation, and traffic operation and management. DTA assigns traffic to different parts of traffic network by creating a balance between demand and supply while DNL models evolution of traffic within the traffic network. DTA and DNL models developed in MFD literature are mostly based on conventional accumulation-based MFD models [24, 25, 26].

By partitioning a heterogeneous urban network into several homogeneous regions using clustering techniques similar to [2], proper MFDs for each region can be defined. To model multi-reservoir urban networks based on MFD, accumulation dynamics of each region is disaggregated into internal and external trips while also disaggregating based on final trip destination. Then, various traffic control strategies for large-scale urban networks can be developed. One control strategy is perimeter control, which meters the flow that transfers between any two regions in a multi-region urban network [4, 27, 28, 29, 16, 18, 30]. It operates by keeping accumulation around a desired level. Control strategies are also modelled in a model predictive control (MPC) framework for real-time applications [15, 31, 18, 24, 32, 33].

This dissertation studies modelling and control of a large-scale urban traffic network based on a parsimonious traffic model. Recently, the United Nations announced seventeen goals to reshape the future for sustainable development. This dissertation is consistent with goal number eleven, which recognizes the importance of moving towards more sustainable cities and societies through efficient use of existing infrastructure.

In this introductory chapter, Section 1.2 presents the motivation for this research. Section 1.3 illustrates the objectives of the current study. Section 1.4 introduces the contributions and methodologies, and Section 1.5 gives the outline of the thesis.

1.2 Motivations of the present study

MFD-based traffic models discussed in the literature do not appropriately represent the evolution of vehicles within traffic networks. In fact, important components of traffic networks are not captured by these models. Accumulation-based MFD models ignore regional travel time and assume that flow instantly propagates within regions [23, 16, 34]. In the absence of regional travel time and under fast-varying demand conditions, the shape and scatter of MFD is affected [35], and numerical viscosity in the accumulation graphs are observed as highlighted by [36]. These models also fail to capture multiple kinematic traffic waves that exist in traffic networks. In addition, they do not differentiate between transferring and queuing vehicles; by ignoring queuing dynamics, the real state of a traffic network is misrepresented. Although these simplifying assumptions are desirable from a modelling and computational perspective, they are not realistic and can be easily violated as traffic congestion propagates throughout urban networks and influences travelers' route choice due to changing internal traffic conditions or OD demand [36]. The need for a more realistic MFD model is indicated in several studies in the literature, and recently, research has been initiated in this area. Aggregated traffic models can be used to their fullest if important traffic occurrences, such as traffic shockwave, queue buildup, dispersion, and spillback, can be appropriately modelled.

DTA and DNL models that use conventional accumulation-based models do not precisely propagate traffic within traffic networks, especially under fast-varying demand conditions. DNL models affect the functionality of traffic strategies that are constructed based upon that. Further, aggregated MFD models, due to their low computational requirements, are widely adopted for network-wide control of large-scale urban networks. However, such manipulation of traffic networks changes traffic states as a result of control and the reaction of road users to the implemented control. For instance, applying a travel information provision may influence route choice behaviour and redistribute trips in space (i.e., less congested roads), in time (i.e., shift from peak to off-peak

hours), or to other travel modes (e.g., carpooling, transit). In saturated networks, road users' responses may be significant and lead to critical bias in devising various traffic management strategies. Therefore, the role of road users in traffic networks should be predicted and incorporated into the design of control strategies; otherwise, the developed control scheme is insufficient.

MFD is an elegant and simple tool to be adopted for modelling of large-scale urban networks and has gained huge attention in recent years for control purposes. While various control strategies are developed in literature to address different problems of traffic networks, the research in this area is still at its infancy. The critical role of road users in traffic network are not considered when developing control and management strategies. Further, while the focus is always on improving efficiency and mobility, fairness and equity measures are absent from the current literature. After reviewing the MFD literature, the following questions raised:

- How can the likely responses of road users be considered while developing network-wide control solutions?
- How can the measures of equity and fairness be incorporated when developing network-wide control solutions?
- How can the conventional MFD dynamics be extended to capture multiple traffic phenomenon?
- How can the extended MFD be adopted for dynamically loading traffic networks?
- How can CAV dynamics be modeled using MFD and what are their likely network-wide effect on traffic networks dynamics?

This dissertation contributes to the existing body of research by trying to find appropriate answers to the above questions.

1.3 Objectives of the present study

This dissertation enhances MFD-based modelling and control by capturing aspects that were not addressed in previous studies. The interaction between control and route choice at the regional level for a large-scale urban network is modeled. The proposed control scheme incorporates road users' route choice behaviour into control decisions and finds a mutual solution between control and user equilibrium. While many control models were presented in the literature on MFD, the role of road users was mostly ignored. Further, fairness and equity measure are hugely neglected when developing control solutions in literature. To address the issue, a fair control scheme for a large-scale urban network that maximizes the utility of road users in different regions is developed. Lastly, an enhanced MFD dynamics in terms of efficiency and realism by capturing important traffic phenomena that are not captured by previous MFD models. Also, the network-wide effect of introducing connected/autonomous vehicles (CAV)s into urban networks is investigated; thus, the main objectives of this dissertation are listed as follows:

- Incorporate within-day short-term responses of road users to control instructions on a network-wide level to develop a globally optimal traffic control solution.
- Develop a network-wide control solution for a large-scale urban network that proactively responds to traffic congestion in real-time.
- Introduce a fairness measure for evaluating utility of road users in different regions of a large-scale traffic network.
- Develop a control strategy to simultaneously address efficiency and fairness measures in devising control solutions for large-scale traffic networks.
- Model more realistically traffic propagation based on MFD by capturing important traffic phenomena, such as kinematic waves, queueing, and congestion.

- Develop a computationally efficient scheme for dynamically loading a traffic network on a large scale.
- Investigate the impact of CAVs on network traffic conditions and the evolution of flow, especially during transition periods from conventional vehicles to CAVs.
- Investigate the effect of information provisions on the shape and scatter of MFD.

1.4 Contributions and proposed methodologies

This dissertation consists of three sub-problems, illustrated in three chapters. Contributions of this thesis for each sub-problem are outlined as follows:

Chapter 2 proposes a network-wide anticipatory control (AC) framework that incorporates drivers' route choice behaviour. Control is modelled in an MPC framework for real-time applications, while user equilibrium is established through a logit route choice model and solved using a method of successive averages (MSA). The following are contributions of this chapter:

- Integrating the rerouting responses of road users into a global control framework using an iterative optimization and assignment procedure.
- Developing a two-level control scheme that consists of control and user equilibrium modules.
- Establishing user equilibrium at the regional level through a logit route choice model and solving the model by adopting the method of successive averages (MSA).

Chapter 3 introduces a proportionally fair perimeter control (PFPC) model for a multi-region urban network in an MPC framework for real-time applications. Queuing dynamics at the regional perimeters are incorporated into the MFD dynamics using a point queue (PQ) model. As an objective for the proportionally fair model, a novel

utility measure based on average regional speed is introduced. Contributions of this chapter are listed below:

- Enhancing MFD by integrating queueing dynamics using a PQ model.
- Proposing a utility measure to distribute traffic fairly within a network.
- Introducing a heterogeneity coefficient based on MFD queuing dynamics.
- Developing a network-wide perimeter control scheme using the proportional fairness concept.

Chapter 4 proposes a multi-reservoir dynamic network loading (DNL) model that consists of a link and a node model. Using variational theory (VT), an LWR model is incorporated into the MFD dynamics to enhance traffic propagation at reservoirs by capturing important traffic phenomena, such as kinematic waves, queueing, and congestion, while the node model functions as a medium for transferring flow to other parts of the multi-reservoir network. We use a numerical scheme to solve the DNL model. Contributions of this chapter are as follows:

- Capturing proper wave propagation and estimating rigorous inner reservoir travel time for an urban arterial with a piecewise linear MFD.
- Developing algorithms based on VT to solve the KWT solution for an MFD-based traffic network.
- Developing a computationally efficient solution to dynamically load a traffic network without requiring space-discretization.
- Incorporating the dynamics of CAVs into the MFD dynamics using an MFD aggregated traffic model.

1.5 Outline of the thesis

This thesis consists of 5 chapters that are laid out as follows:

Chapter 2 introduces an anticipatory control framework for network-wide control of an urban network. This chapter introduces accumulation-based MFD models and develops a perimeter control scheme and user equilibrium model based on the introduced MFD model.

Chapter 3, first introduces the accumulation-based MFD model that considers queueing dynamics. In the second part, I illustrate the proportional fairness concept and propose an objective function for the control model that considers utility of road users for travelling in a heterogeneous traffic network separated into several homogeneous regions.

Chapter 4 provides a comprehensive review of accumulation-based MFD models with specific attention to the capability of these models for representing precise inner regional travelling time and a proper description of wave propagation. I present an enhanced traffic model based on VT and develop the solution algorithm. This chapter also covers the incorporation of CAVs into the MFD dynamics and investigates the network-wide effect on maximum flow, shape, and scatter of MFD.

Chapter 5 summarizes the findings of this thesis and presents the main conclusions. In addition, suggestions for future research are mentioned.

Chapter 2

A network-wide anticipatory control of an urban network using macroscopic fundamental diagram

In this chapter [1], a network-wide anticipatory control (AC) framework, incorporating drivers' route choice behaviour, is proposed. The proposed AC, consisting of two main levels of control and route choice, explicitly accounts for road users' responses to the implemented control. Perimeter control, modeled in model predictive control (MPC) framework for real time applications, optimizes inter-regional transferring flow at network level. User equilibrium is stablished through a logit route choice model and solved using method of successive averages (MSA). The modelling approach adopted is based on macroscopic fundamental diagram (MFD), which provides a unimodal low-scatter relationship between density and outflow of any region within a network. Numerical results indicate that the proposed AC outperforms no control and basic control cases and shows promising results in alleviating congestion and deriving the network to near system optimum traffic condition, under various demand patterns and levels.

2.1 Introduction

In recent decades, traffic congestion has become a major issue in traffic networks due to the ever-increasing trend of travel demand and the limited supply of road network

[1]The essential contents of this chapter have been published in: Moshahedi, N., Kattan, L. and Tay, R., 2021. A network-wide anticipatory control of an urban network using macroscopic fundamental diagram. Transportmetrica B: Transport Dynamics, 9 (1), pp.415-436. With permission from Taylor & Francis [see Appendix].

facilities. Traffic management schemes and control measures, such as signal control, ramp metering, variable speed limit, congestion pricing and route guidance, were shown to be effective in alleviating urban traffic congestion [see [37] for a review]. It has been demonstrated that incorporating drivers' routing responses into the control schemes, enhances the efficacy of controllers in improving mobility and mitigating congestion, especially in saturated traffic conditions [38, 39, 40, 41, 42, 43].

Integrating user equilibrium into control problem is often referred to as anticipatory control (AC) of traffic network. AC accounts for drivers' routing decisions by predicting their responses to control, and including the anticipated likely responses in the control decision, with an aim to reach global optimality [44]. Manipulation of urban networks by implementing any type of control, changes traffic state not only due to the control itself but also due to the drivers' routing responses to the implemented control. For instance, applying a travel information provision may influence route choice behaviour and redistribute trips in space (i.e., less congested roads), in time (i.e., shift from peak hours to off-peak), or to the other travel modes (e.g., carpooling, transit). In saturated networks, rerouting effect may be significant, and lead to critical bias in devising various traffic management strategies. The effect of control on routing decisions was first addressed by [45]. This seminal work led to a tail of research on problems such as urban traffic signal control [46, 47] and road pricing [48, 38]. More recently, this approach was adopted in devising control schemes based on dynamic user equilibrium [39, 40, 43]. Anticipatory traffic control approaches can be modelled in the following forms: bi-level optimization problem (BLP) with control model at the upper level and route choice model at the lower level [40] (in BLP, upper level problem refers to the outer optimization problem and lower level problem refers to the inner optimization problem), optimization problem with user equilibrium as a constraint [41, 42], and a separate control and route choice models related through iterative optimization and assignment (IOA) [44]. In game theory terms, BLP is equal to the Stackelberg game in which road authority determines control measures with anticipation of the likely re-

sponse of road users, and IOA is equal to a Nash-Cournot game in which road authority and road users make decision independently and iteratively given the latest information received from the other entity until convergence criterion is satisfied. The difference between solutions found using the two methods is insignificant. For a medium-sized problem, IOA converges within reasonable computational time and effort; therefore, IOA is suitable for real-time applications [49].

The practical relevance of explicitly considering route choice in control problem lies in the rerouting effect induced by implementation of control. In other words, AC facilitates evaluation of road users' responses to various short-term and long-term control strategies. Further, it can be used to examine the control performance under system optimal control conditions, where control and route guidance are integrated [50]. [51] investigated the effect of the type of control on route choice and concluded that route choice is strongly affected by the control used in the traffic network. Hence, ignoring such responses may not only limit the potential usefulness of the developed strategy, but may also lead to deterioration rather than the intended amelioration of traffic conditions when implemented in real-world traffic networks. Despite the merits associated with AC models, integrating route models into control framework adds extra complexity to the problem, making tractability issue a real concern, even in case of static traffic assignment and a simple network with separable travel time function [41, 43]. Addressing the analytical tractability issues suggests the need to incorporate travellers' responses using aggregated traffic models.

Developing a network-wide control solution based on a detailed microscopic or mesoscopic traffic model is not operational for real-time applications. These models are based on highly detailed link-level information, which is difficult to acquire and analyze in real-time; thereby imposing a huge computational burden on the system. Furthermore, the unpredictability of road users' behaviour adds more complexity to the problem. To solve the network-wide control problem, parsimonious or approximate aggregated traffic models proved to be more computationally efficient compared to their detailed micro-

scopic counterparts. The readers are referred to [3] for a review of parsimonious traffic models. Macroscopic fundamental diagram (MFD) has emerged as a promising traffic modelling tool in recent years. It provides a unimodal, demand-insensitive, low-scatter relationship between network density or accumulation (number of vehicles in a network) and network space-mean flow or outflow (trip completion rate), in roughly homogeneous urban networks. It was initially proposed by [5] and was widely recognized due to the seminal work of [4]. [6] derived analytical theories for MFD, and [1] verified its existence in Yokohamas (Japan) traffic network by integrating the highly scattered plots of flow versus density obtained using a combination of fixed detectors and GPS-equipped taxis. It was demonstrated that a well-defined, low-scatter MFD can only be identified for networks with a spatially homogeneous distribution of density. Heterogeneity undermines the existence of a well-defined MFD, and reduces its outflow [11, 12, 13]. MFD was further investigated using empirical or simulated data in other studies [11, 2, 12], and several analytical [6, 9] and experimental [1, 52] methods were suggested in literature for its efficient estimation. To facilitate modelling and control of a heterogeneous network, it can be partitioned into several smaller homogeneous sub-regions with well-defined MFD, using spatial properties of congestion [2]. Perimeter controllers, metering the amount of flow that is allowed to transfer between any two neighboring regions, were developed for control of urban networks [4, 27, 28, 29, 16, 18, 30]. Model predictive control (MPC) was utilized for modelling control problems for real-time control purposes [15, 31, 18, 24, 32, 33].

Recently, research efforts have been devoted to developing route choice models based on MFD dynamics. [53] approximated user equilibrium by developing a region-based route choice model. They proved that an aggregated MFD is capable of representing system dynamics nearly as well as a detailed microscopic one, based on the aggregated traffic information with reasonable computational burden that is suitable for real-time applications. [24] advanced a user equilibrium and a system optimum route choice models based on MFD. For a mixed urban and freeway network under perimeter con-

trol, a simple DTA model based on instantaneous travel time was modeled by [15]. [32] adopted Logit route choice model to describe road users' routing in a traffic network with combined perimeter and route guidance control while considering complying vehicles to the route guidance control recommendations. [54] used C-Logit route choice model to simulate road users' adaptiveness to traffic condition in a network with perimeter and boundary flow controllers. [24] developed a region-based route guidance model that considered a sub-regional level route choice model as a plant to capture network's dynamic properties such as regional mean trip length. [55] developed a hierarchical traffic control scheme that consisted of a centralized higher-level regional route guidance and a local lower-level sub-regional path assignment; the scheme recommended a sequence of sub-regions from an origin to a destination that met the higher-level control goal. [56] proposed a demand management strategy that optimizes network performance by manipulating road users' departure times. In this study, the plant model is based on trip-based MFD with explicit consideration of arrival time, trip length and earliness and lateness fees for each traveller while the optimization model is based on accumulation-based MFD to deal with analytical tractability issues. Although a handful of previous studies in MFD control literature considered route choice, consideration of the rerouting effect induced as a result of the implemented control and its impact on other parts of the traffic network is seriously lacking. One leading efforts in this direction is the work by [57] who developed a modeling framework to assess the impact of perimeter gating control on outside network in terms of queuing, emission and total time spent of the network. For a single region with a by-pass route as an alternative, a simple proportional integral (PI) type controller was modeled to keep the accumulation around the critical level while the demand elasticity resulting from the implemented control and rerouting was considered in dynamic user equilibrium model in the closed-loop control framework. In their study, they used a simple PI-type controller while considering only a single region network.

This chapter extends the literature on MFD-based control by integrating the rerout-

ing responses of road users into the global control framework through an iterative optimization and assignment procedure. In the proposed AC, control and route choice models interact through an iterative procedure which transfers the route choice and control information between the two parties. The urban network under study is divided into several regions dynamics of which is described through MFD. Perimeter control modeled in MPC framework to handle demand uncertainty and accumulation measurement error is devised to control the inflow to the CBD region. MPC plant is modeled at the same abstract level as the prediction model and is assumed to have access to the real demand and accumulation values. To balance the trade-off between computational burden and accuracy of the model and to keep the problem tractable for real-time control application, both control and route choice are developed at the regional level and related through an iterative procedure. Performance of the proposed AC is compared with no control, basic control and system optimum which is considered as a benchmark of an ideal traffic state representing the best network performance that can be achieved when traffic managers have full control over travellers' routing.

Current work focuses on within-day traffic assignment, which investigates within-day short-term response of drivers to traffic condition. It is thus applied to traffic network after the initial period of introduction of control strategy, when day-to-day transient effects have settled. Logit model adopted for describing road users' routing responses accounts for travellers' knowledge of travel time. Today's availability of real-time traffic information from Google traffic or other navigation devices assures and hastens the establishment of some form of traffic equilibrium [58]; thereby making the convergence of the proposed AC more realistic. In fact, travellers are affected by the delays resulting from vehicle queuing at the perimeter boundaries and real-time traffic information considers these delays in the suggested routes.

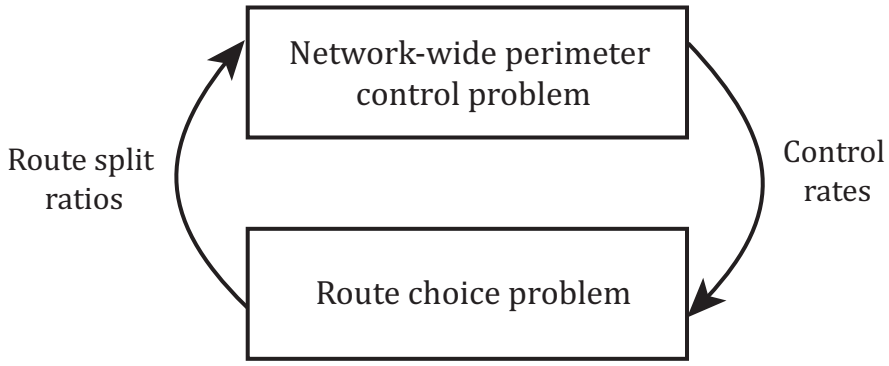


Figure 2.1: General scheme of interaction between control and route choice

2.2 Methodology

To represent the inter-dependency between control and route choice, it is essential to explore the way they interact with one another. Traffic can be considered as a mutually consistent system; travel time is determined by congestion patterns, and congestion patterns are affected by travel time [44]. When perimeter control is introduced to a network, congestion patterns and delays are induced as a result of restricting the inflow to the controlled area of the network and the created queue at the perimeter boundary. Part of the restrained inflow reroutes to less congested alternative routes, increasing travel time at those routes as a consequence. The resulting changes in traffic flow patterns and travel times make another round of control necessary in order to keep the network traffic state around the ultimate control goal. The interaction between two involved parties continues until a mutual solution is found. The inter-dependency between user equilibrium function, F, and travel time function, E, in a controlled network with λ as a routing variable and u as a perimeter control variable, corresponds to a fixed-point problem as follows:

$$\lambda = F(E(\lambda, u))$$

The fixed-point problem is described in a general framework, as shown in Figure 2.1.

Perimeter control values are determined based on the traffic flow patterns received from the route choice model, and routing decisions are made based on the control values estimated at the perimeter control model. The detailed description of the control and route choice models are elaborated in next sections.

2.2.1 Formulation of the upper-level MFD-based perimeter control problem

Consider $G = (R, W)$ as a connected urban network, where $R = \{1, ..., r_m\}$ is a set of regions, W is a set of OD region pairs, and P_w is a set of routes between OD pairs $w \in W$. Consider G as a heterogeneous urban network partitioned into several homogeneous regions. Dynamics of any region $i \in R$ is described by its MFD $= G_i(n_i)$ (veh/s), indicating the number of trips completed within a region as a function of its accumulation n_i (veh).

To model the MFD dynamics, n_i is disaggregated according to the final trip destinations, such that $n_i = n_{ii} + \sum_{j \neq i, j \in R} n_{ij}$. n_{ii} denotes the accumulation of region i with the final trip destination within the same region (internal trip) while n_{ij} denotes the accumulation of region i with the final destination in another region j (external trip). The general mass conservation equations of the multi-region urban network as described by [18] are as follows:

$$\dot{n}_{ii}(t) = d_{ii}(t) + \sum_{l \in H_i} u_{li}(t) f^i_{li}(t) - f_{ii}(t) \qquad \forall i \in R, t = 1, 2, ... \tag{2.1}$$

$$\dot{n}_{ij}(t) = d_{ij}(t) + \sum_{l \in H_i, l \neq j} u_{li}(t) f^i_{lj}(t) - \sum_{r \in H_i} u_{ir}(t) f^r_{ij}(t) \qquad \forall i, j \neq i \in R, t = 1, 2, ...$$

Where d_{ii} (veh/s) denotes newly generated internal demand in region i, and d_{ij} denotes newly generated external demand in region i with a final destination in another region j. $f_{ii} = (n_{ii}/n_i)G_i(n_i)$ (veh/s) denotes internal flow in region i, and $f_{ij} = (n_{ij}/n_i)G_i(n_i)$ denotes the flow traveling from region i to final destination at region j; region i and region j are not necessarily neighboring regions. u_{li} denotes the perimeter

control value, controlling the amount of flow allowed to transfer between region i and its neighbor region $l \in H_i$. f_{ij}^r corresponds to the flow traveling from region i to region j passing through the immediate next region $r \in H_i$, where $f_{ij}^r = \theta_{ij}^r . f_{ij}$. Regional split ratio θ_{ij}^r denotes the percentage of flow transferring from region i to region j through the immediate next region $r \in H_i$. This variable is integrated into the state estimation equations in order to account for route choice decisions and is estimated from the traffic equilibrium model discussed in the next section.

The upper-level perimeter control problem is constructed in model predictive control (MPC) [59, 60] framework. In MPC, at each time step, the perimeter control values are optimized over a prediction horizon; however, only the first sample of the control is applied, the horizon is shifted forward, and control rates are optimized using the updated routing responses over the new shifted control horizon. The whole process is repeated continuously until the end of simulation time. MPC is suitable for adopting in the current control model since it can handle uncertainty, disturbances, and model mismatch errors that exist in the model as a result of data collection noise and demand prediction error. Further, it can easily handle constraints and the nonlinear dynamics of a control problem. MPC for the MFD-based perimeter control problem was first proposed by [30] for a simple two-region urban network, and was later used in more complicated perimeter control problems in MFD literature [15, 31, 18]. This approach effectively handles traffic condition in a congested network compared to simple control approaches such as bang-bang controller [4] that causes uneven distribution of vehicles, increases heterogeneity, and reduces the total network outflow when used in a multi-region urban network; and PI-type controller [29, 16] with unsatisfactory performance in high demand and saturated traffic condition.

The higher-level MPC perimeter controller is formulated as follows:

$$\min_{u_{ij}} \quad T_c \sum_{t=1}^{N_p} \sum_{i \in R} n_i(t) \tag{2.2}$$

(2.1)

$$0 \leq n_i(t) \leq n_{i,jam} \qquad \forall i \in R, t = 1, .., N_p \tag{2.3}$$

$$u_{min} \leq u_{ij}(t) \leq u_{max} \qquad \forall i \in R, j \in H_i, t = 1, ..., N_p \tag{2.4}$$

$$\theta_{ij}^r(t) = \sum_{k \in P_{ij}} \lambda_{ij}^k(0) \cdot \delta_{ij}^{r,k}; \qquad \forall r \in H_i, t = 1, ..., N_p \tag{2.5}$$

$$d_{ij}(t) = d_{ij,t} \qquad \forall i, j \in R, t = 1, ..., N_p \tag{2.6}$$

$$n_{ij}(0) = n_{ij,0} \qquad \forall i, j \in R \tag{2.7}$$

$$\lambda_{ij}^k(0) = \lambda_{ij,0}^k \qquad \forall i, j \in R, \forall k \in P_{ij} \tag{2.8}$$

Eq (2.2) minimizes the total accumulation of the network [this value is equal to the total travel time of the network as discussed by [29]] with respect to perimeter control values u_{ij}, over a prediction horizon N_p with T_c as a control sampling time. Mass conservation Eq (2.1) is added to estimate the dynamics of the network as explained earlier in this section. Eq (2.3) ensures that regions' accumulations do not exceed the jam accumulation level $n_{i,jam}$, and thus prohibits the network from moving towards gridlock. Eq (2.4) constrains the perimeter control values between a lower limit u_{min} and an upper limit u_{max}. Eq (2.6), (2.7) and (2.8) initialize demand, accumulation, and regional split ratio, respectively.

Regional and route split ratios considered in this chapter define the percentage of inflow traveling on any route within the network; however, there is a slight difference in their definition. Route split ratio λ_{ij}^k, defined similar to conventional link-based traffic assignment, is associated with a given path k and thus, denotes the percentage of inflow traveling from origin region i to destination region j through path k, where k constitutes

a set of consecutive regions. On the other hand, regional split ratio θ_{ij}^{r} aggregates route split ratio values based on a common traversed region, $r \in H_i$, which is the immediate neighbor region to the origin region i. The control model requires regional split ratio θ_{ij}^{r} as an input while the traffic assignment model estimates route split ratio λ_{ij}^{k} as its output; hence, Eq (2.5) is added to convert route split ratios to regional split ratios by summing over all the existing routes between region i and j that passes through the next immediate region $r \in H_i$. This is schematically shown in Figure 2.3. Consider route split ratios λ_{15}^{1} corresponding to route 1 $(1-2-3-5)$ and λ_{15}^{2} corresponding to route 2 $(1-2-4-5)$. Since both paths traverse region 2 immediately after origin region 1, the regional split ratio for trips from origin region 1 to destination region 5 passing through immediate region 2, θ_{15}^{2}, is estimated by aggregating route split ratios λ_{15}^{1} and λ_{15}^{2}. Region-route incidence matrix, $\delta_{ij}^{r,k} = 1$, if in a trip from origin region i to destination region j using path $k \in P_{ij}$, region r is the immediate region that is traversed after origin region i; $\delta_{ij}^{r,k} = 0$ otherwise.

2.2.2 Formulation of the lower-level MFD-based route choice problem

To establish user equilibrium, an MFD-based regional route choice model is proposed. Due to unavailability of link-level traffic information and analytical tractability issues, a region-based model determining the sequence of regions taken for a trip from an origin region to a destination region is advanced. For any region i, there exists a separable travel time function estimated as $\tau_i(n_i) = n_i/G_i(n_i)$, provided that regional trip length is assumed constant (the assumption of constant trip length is conventional assumption of accumulation-based model). [1] confirmed this assumption by exploring the relationship between production and outflow of a region, and finding that the relationship indicates a relatively constant value throughout a day; however, in later works, the assumption of constant average trip length was revisited and replaced by continuous modelling of an area within a region [53], or dividing the region into several sub-regions and integrating their dynamics into the route choice model [24, 18, 55]. In this chapter, I assume that

the regional trip length is constant and is not affected by drivers' route choice decisions; however, the proposed framework can be extended to consider variable regional trip lengths.

In order to track the vehicle movements on different routes within the network, accumulation of each region is decomposed according to its origin, destination, and the route on which traverse from a particular region to a destination. Let $n_{ij}^{r,k}$ be the accumulation of vehicles in region r that travels from origin region i to destination region j through path $k \in P_{ij}$. The dynamics of region r based on its location on path k is expressed as follows [24]:

$$
\begin{aligned}
&\dot{n}_{ii}^{i,k}(t) = d_{ii}(t) - f_{ii}(t) \qquad \forall i \in R, t = 1, ..., N_p \\
&\dot{n}_{ij}^{i,k}(t) = d_{ij}^{k}(t) - u_{ik^{+(i)}}(t)\hat{f}_{ij}^{i,k}(t) \qquad \forall i, j \neq i \in R, t = 1, ..., N_p \\
&\dot{n}_{ij}^{j,k}(t) = u_{k^{-(j)}j}(t)\hat{f}_{ij}^{k^{-(j)},k}(t) - f_{ij}^{j,k}(t) \qquad \forall i, j \neq i \in R, t = 1, ..., N_p \\
&\dot{n}_{ij}^{r,k}(t) = u_{k^{-(r)}r}(t)\hat{f}_{ij}^{k^{-(r)},k}(t) - u_{rk^{+(r)}}(t)\hat{f}_{ij}^{r,k}(t) \qquad \forall i, j \neq i, r \neq i \in R, t = 1, ..., N_p
\end{aligned}
\tag{2.9}
$$

The first equation in this set estimates state dynamics of internal trips, and the other three equations are used for estimating state dynamics of external trips. For each route, accumulation of origin and destination regions are estimated using second and third equations, respectively. The last equation is applied when the intended region is none of the origin or destination regions. d_{ij}^{k} denotes part of the demand between origin region i and destination region j that is assigned to path $k \in P_{ij}$, such that $d_{ij}^{k} = \lambda_{ij}^{k} d_{ij}$. Note that the route split ratio λ_{ij}^{k} obtained from the route choice model is different from the regional split ratio θ_{ij}^{r} estimated in Section (2.2.1).

Transferring flow $f_{i,j}^{r,k}$ from region r to its succeeding neighboring region $k^{(+r)}$ on path $k \in P_{ij}$ is restricted by the receiving capacity of the succeeding region; therefore,

the real transferring flow $\hat{f}_{ij}^{r,k}$ can be computed as a minimum of two terms as follows:

$$\hat{f}_{ij}^{r,k}(t) = \min(f_{ij}^{r,k}(t), C_{rk^{+(r)}}(n_{k^{+(r)}}(t)) \frac{n_{ij}^{r,k}(t)}{\sum_{i\in R}\sum_{j\neq i}\sum_{k\in P_{ij}} n_{ij}^{r,k}(t)}) \quad \forall i,j,r \in R, r \in H_i, k \in P_{ij} \tag{2.10}$$

Where $C_{rk^{(+r)}}(n_{k^{+(r)}}(t))$ denotes receiving flow capacity of region $k^{(+r)}$ from region r, estimated as a piecewise function relating to accumulation as follows [18, 61]:

$$C_{rk^{(+r)}}(n_{k^{+(r)}}) = \begin{cases} C_{rk^{(+r)}}^{\max} & \text{if} \quad 0 \leq n_{k^{+(r)}} \leq \chi n_{k^{+(r)},jam} \\ \frac{C_{rk^{(+r)}}^{\max}}{1-\chi}(1 - \frac{n_{k^{+(r)}}}{n_{k^{+(r)},jam}}) & \text{if} \quad \chi n_{k^{+(r)},jam} \leq n_{k^{+(r)}} \leq n_{k^{+(r)},jam} \end{cases} \tag{2.11}$$

$C_{rk^{(+r)}}^{\max}$ denotes the maximum value of region's receiving capacity, computed as a minimum of region's supply function and boundary capacity. χ is the parameter that defines critical accumulation of receiving capacity function, associated with a point that capacity starts to decrease from the constant value. While [62] and [31] considered this point equivalent to the MFD critical accumulation, [61] identified this assumption too restrictive and claimed that turning point of receiving capacity is higher than the critical accumulation value.

In presence of perimeter control at boundary of the neighboring regions, the actual flow that transfers from region r to the next region $k^{+(r)}$ is also constrained by the perimeter control rates $u_{rk^{+(r)}}$. [63] claimed that in such condition receiving capacity constraint can be eliminated as perimeter control prohibits the network from reaching high accumulation levels. In this study, however, since we only control the inflow to CBD region and no control between other regions is applied, this constraint cannot be disregarded.

The total regional accumulation at any time step t is derived as follows:

$$n_r(t) = \sum_{i \in R} \sum_{j \in R} \sum_{k \in P_{ij}} n_{ij}^{r,k}(t) \qquad \forall r \in R \tag{2.12}$$

2.3 Framework for network-wide anticipatory control of urban network

The proposed Anticipatory control (AC) scheme consists of two models, which actively interact and influence one another [see Figure 2.2 for schematic view of the proposed AC]. Perimeter control problem, modeled in MPC framework, minimizes the total network time spent over a pre-defined prediction horizon through optimizing the amount of flow that transfers between adjacent regions, given road authorities' control goals and regulations, demand, and the route split ratios received from the route choice model. Based on the perimeter control rates obtained from the controller, the route choice model minimizes drivers' selfish perceived travel times using the logit route choice model. In the route choice model, firstly a set of initial route split ratios are estimated based on a free-flow travel time which is used to update traffic state and travel time. Based on the updated travel time, route split ratios are estimated using logit assignment formula. Lastly, routing ratios are derived as a weighted average of the previously obtained solutions. If convergence criterion is not satisfied, the procedure continues by updating the traffic state at next iteration. Once convergence of MSA is achieved, route split ratios are sent to the control model. In response to the newly established user equilibrium, the perimeter control rates are re-estimated by the controller. The procedure continues until convergence or stopping criterion is satisfied. The algorithm reiterates over all time steps during the control interval.

Modelling the perimeter control problem in MPC framework enables us to handle the noise and measurement errors attributed to the stochastic nature of demand and accumulation states. Thus, at each time step, after convergence is achieved and new control rates and routing ratios are obtained, the traffic state of the network is updated using the plant model as representing reality of the traffic network. It is assumed that

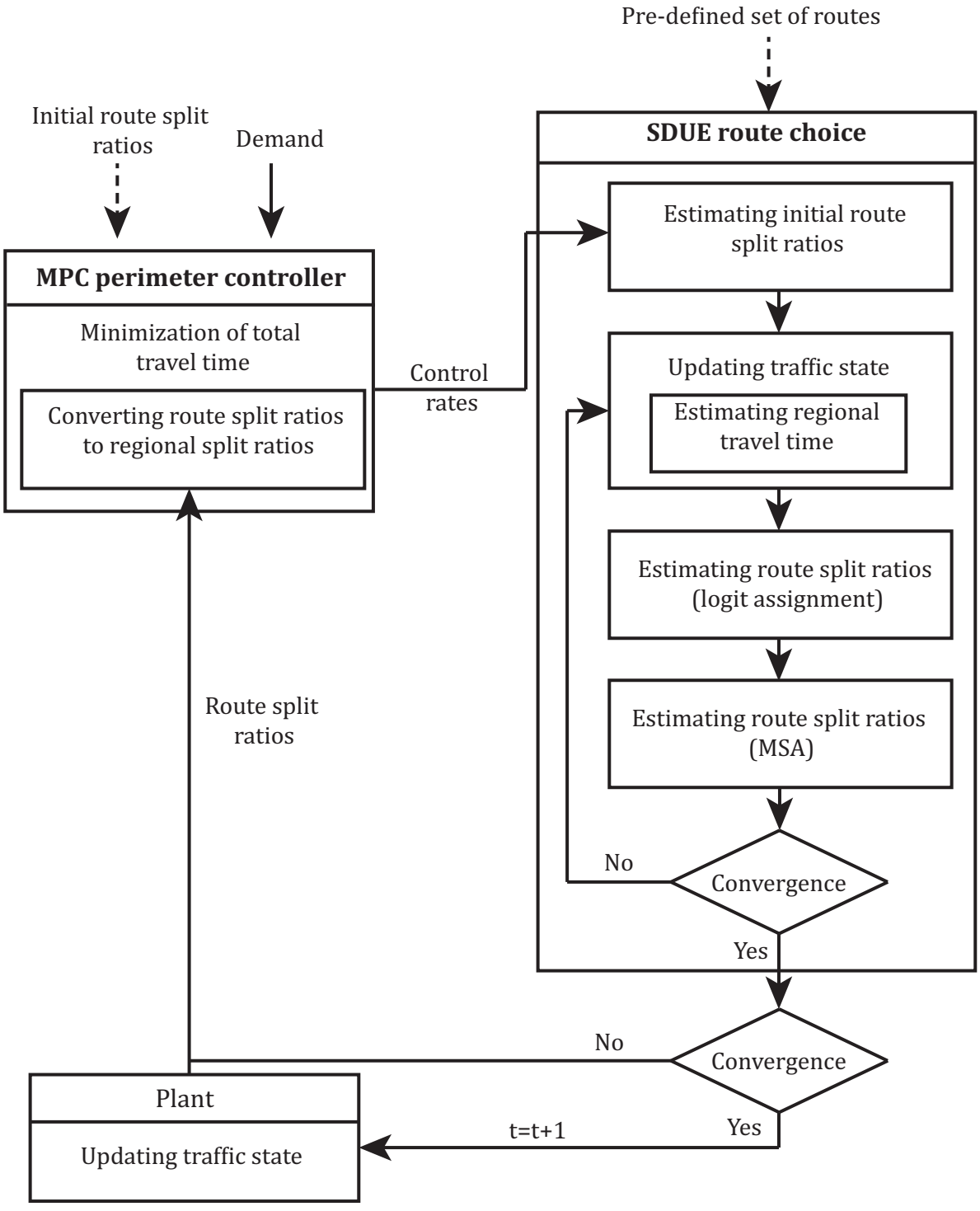


Figure 2.2: Schematic view of the proposed network-wide anticipatory control

plant model has access to real values of demand and accumulation while the prediction model has only access to the average values. Thus, the control schemes are assumed to receive information from a plant model, which is constructed based on a route choice model to simulate the evolution of traffic in the network as the demand profile is loaded; In other words, the state of the system is assumed to be continuously monitored and noisy "observed data" is obtained from the plant model (similar to the data obtained from traffic sensors that is used to estimate the state of the network in real-time). Consistent with [32], regional split ratios estimated using route split ratios are considered constant during the prediction horizon. Predicting them would add extra complexity to the model while a significant change in result is not expected. For future research, it is interesting to test if the solution improves by relaxing the constant route split ratio assumption during the prediction horizon of the MPC model. Performance and convergence of the proposed AC is evaluated using a numerical example in Section 2.4.

2.3.1 Method of successive averages for solving regional route choice model

In this section, we employed the well-known method of successive averages (MSA) to solve the region-based stochastic user equilibrium (SUE) problem embedded in the proposed AC framework. MSA is widely used for solving traffic assignment problems due to its simplicity. It functions based on a pre-determined moving step-size along the descent direction and estimates the route flow at each iteration as an average of the solutions from previous iterations. Since the flow pattern generated at each iteration contributes equally to the solution, MSA suffers from low convergence speed. To speed up the process, we assigned higher weights to the solutions from more recent iterations by setting step size $\epsilon^{(m)} = \frac{2}{m+1}$, with m being number of iterations [64]. [65] proved that under some regularity conditions, any sequence of $\epsilon^{(m)}$ satisfying $\sum_{m=1}^{\infty} \epsilon^{(m)} = \infty$ and $\sum_{m=1}^{\infty} {\epsilon^{(m)}}^2 \leq \infty$ will guarantee convergence of MSA algorithm; hence, the decreasing step size $\epsilon^{(m)} = \frac{2}{m+1}$, forces the algorithm to converge. MSA convergence criterion in this study, requires the maximum difference between variables of two successive iterations

to be less than a pre-defined convergence threshold.

In addition to the SUE model, a stochastic system optimum (SSO) route choice model is formulated for comparison purposes to examine the effectiveness of AC and its ability to reach a near system optimum traffic state. With estimation of mean traveling time $\tau_i(n_i)$ for any region in the network, and assuming a third-order function for MFD, marginal regional travel time required for SSO problem is derived as follows:

$$\begin{aligned}\tilde{\tau}_i(n_i) &= \tau_i(n_i) + n_i.\frac{d\tau_i(n_i)}{dn_i} = \tau_i(n_i) + n_i.\frac{d}{dn_i}(\frac{n_i}{G_i(n_i)}) \\ &= \tau_i(n_i) + n_i.\frac{d}{dn_i}(\frac{n_i}{an_i{}^3 + bn_i{}^2 + cn_i}) = (\tau_i(n_i))^2(-an_i{}^2 + c) \end{aligned} \tag{2.13}$$

In this formulation, $\tilde{\tau}_i(n_i)$ is interpreted as marginal contribution of an additional traveller to the travel time [66]. The MSA algorithm used to solve SUE/SSO problem is demonstrated as follows [65, 66]:

- Step 0- initialization:

  - set iteration number $m = 1$.

  - perform initial stochastic network loading using regional free-flow travel time to estimate route travel time for SUE/SSO using $\mathcal{T}_{ij}^k = \sum_{r\in R}(\tau_r(n_r)+\frac{T}{2}u_{k^{-(r)}r}\hat{f}_{k^{-(r)}r})\delta_{ij}^{r,k}$ / $\tilde{\mathcal{T}}_{ij}^k = \sum_{r\in R}(\tilde{\tau}_r(n_r)+\frac{T}{2})d_{ij}\delta_{ij}^{r,k}$, and the logit assignment formula $\lambda_{ij}^{k\,(1)} = \frac{e^{-\phi\mathcal{T}_{ij}^k}}{\sum_{k\in P_{ij}} e^{-\phi\mathcal{T}_{ij}^k}}; \phi > 0$/ $\lambda_{ij}^{k\,(1)} = d_{ij}\frac{e^{-\phi\tilde{\mathcal{T}}_{ij}^k}}{\sum_{k\in P_{ij}} e^{-\phi\tilde{\mathcal{T}}_{ij}^k}}; \phi > 0$.

- Step 1- update:

  - update MFD dynamics using eq (2.9).

  - estimate total accumulation of each region using eq (2.12).

  - estimate the updated regional travel time/regional marginal travel time for SUE/SSO using $\tau_r(n_r)/\tilde{\tau}_r(n_r)$.

- Step 2- direction finding:

  - perform stochastic network loading using the updated route travel time for SUE/SSO using $\mathcal{T}_{ij}^k/\tilde{\mathcal{T}}_{ij}^k$, and the logit assignment formula ${\lambda^*}_{ij}^{k\ (m)} = \frac{e^{-\phi \mathcal{T}_{ij}^k}}{\sum_{k \in P_{ij}} e^{-\phi \mathcal{T}_{ij}^k}}; \phi > 0 / {\lambda^*}_{ij}^{k\ (m)} = d_{ij} \frac{e^{-\phi \tilde{\mathcal{T}}_{ij}^k}}{\sum_{k \in P_{ij}} e^{-\phi \tilde{\mathcal{T}}_{ij}^k}}; \phi > 0.$

- Step 3- move:

  - update the route ratios $\lambda_{ij}^{k\ (m+1)} = \lambda_{ij}^{k\ (m)} + \epsilon^{(m)}({\lambda^*}_{ij}^{k\ (m)} - \lambda_{ij}^{k\ (m)})$.

- Step 4- convergence criterion:

  - if $\max_{ijk} |\frac{\lambda_{ij}^{k\ (m+1)} - \lambda_{ij}^{k\ (m)}}{\lambda_{ij}^{k\ (m)}}| \leq \kappa$, convergence is achieved. Otherwise $m = m+1$ and return to Step 1.

Stochastic network loading procedure is conducted using a simple logit model with $\phi > 0$ being a random term, implying drivers' level of awareness regarding travel time (higher ϕ corresponds to higher level of knowledge).

2.4 Case study and results

In this section, performance of the proposed anticipatory control (AC) is examined using a numerical example. The urban network under study consists of seven regions as shown in Figure 2.3. Region 4 is the central business district (CBD) of the city. The shortest route between any two non-neighbor regions is usually provided through the CBD. Each region has its own specific MFD in the shape of a third-order function skewed to the right, making the critical accumulation less than half the jam accumulation. The MFD function considered for any region in this network is a scaled version of the one observed in part of downtown Yokohama [1], and is defined as $G_i(n_i) = \alpha_i(4.133 \cdot 10^{-11}(n_i/\beta_i)^3 - 8.282 \cdot 10^{-7}(n_i/\beta_i)^2 + 4.2 \cdot 10^{-4}(n_i/\beta_i))$, where $\alpha = \{1.02, 0.99, 1.04, 1, 0.98, 0.97, 0.96\}$ and $\beta = \{0.96, 1.02, 0.98, 1, 0.99, 0.97, 1.03\}$, with jam accumulation $n_{i,jam} = 10^4 \cdot \beta_i$ (veh),

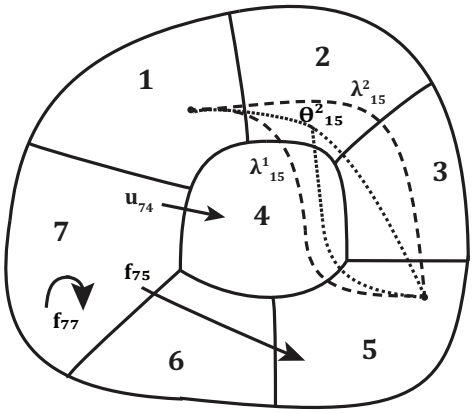


Figure 2.3: Urban network divided into 7 regions

critical accumulation $n_{i,crt} = 3.4 \cdot 10^3 \cdot \beta_i$ (veh) and maximum outflow $G_i(n_{i,crt}) = 6.3.\alpha_i$ (veh/s). Regions' receiving capacity is included by parameters $C_{ik^{(+i)}} = 3.5$ (veh/sec) and $\chi = 0.64$. Accumulation states are assumed to carry measurement noises as they are obtained from fixed and mobile sensors. The error terms are described by normal distribution with zero mean and standard deviation equal to 0.4, i.e., $\varepsilon(n) \sim \mathcal{N}(0, 0.4)$. Choice of noise and error terms are based on an earlier study [67].

The analysis is conducted for four different demand profiles, considering medium and high demand volume, and uniform and radial demand distribution. In all scenarios, travel demand is distributed uniformly within each region. Compared to the uniform demand scenarios, in the radial demand scenarios, CBD region attracts more trips. Demand uncertainty is described by errors terms with normal distribution with zero mean and standard deviation equal to 0.4. , i.e. $\varepsilon(d) \sim \mathcal{N}(0, 0.4)$.

In the analysis, the difference in road users' route choice decisions under controlled and uncontrolled traffic conditions is highlighted. In addition, we investigate the effectiveness of the proposed AC framework in handling congestion and preventing gridlock at network-level compared to the following cases:

(1) No control (NC), in which flow is not controlled at regions' perimeter boundaries, and drivers choose their routes according to the SUE model.

(2) Basic control (BC), in which control rates are estimated using MPC controller

without including route choice in control decisions. Routing is only updated after each control time step according to the SUE model.

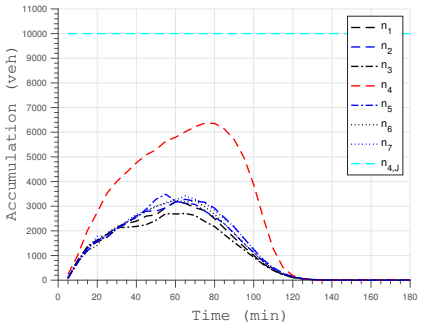
(3) System optimum (SO), in which similar to NC, flow is not controlled at regions' perimeter boundaries, and drivers are assumed to choose their routes according to the SSO model.

Note that there is a difference between inclusion of route choice in BC and AC. While in BC, the route choice is exogenously determined and estimated to reflect the evolution of the traffic state once the BC control rates are injected into the network, in AC, the route choice is endogenous and anticipated as an integral part of AC scheme. SO serves as a benchmark of an ideal control strategy that is used as a target performance indicator of the developed AC. SO can be seen as a cooperative game in which travellers choose their routes such that the total travel time of the network is minimized. SO can be achieved through control and management schemes (e.g., route guidance and an advanced traveller information system with a relatively high compliance rate) that influence travellers' route choice; therefore, maximum performance of the proposed AC is achieved when travellers are distributed in the network in a fashion similar to that of SO.

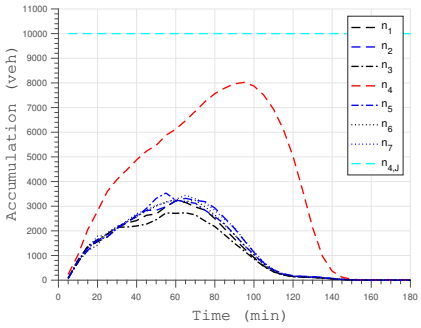
In controlled scenarios, only the inflow from peripheral regions to CBD is metered. All other transferring flows between peripheral regions and from CBD to peripheral regions are not controlled. All regions are empty at the start of the control period. Regional trip length is assumed to be constant. Lower control bound $u_{min} = 0.1$, and upper control bound $u_{max} = 1$ are considered. Path enumeration is avoided by considering the three shortest routes found using Dijkstra's algorithm between any OD region pair in the network. As shown in Figure 2.2, route and regional split ratios are periodically updated based on the logit route choice model and using the newly obtained perimeter control rates based on the pre-defined routes and the associated route-region incidence matrix from Dijkstra's algorithm. The network is simulated and controlled for 3 hours. Route split ratios are updated every 60 sec and control sampling time $T_c = 300$

sec is considered. For the MPC model, a prediction horizon $N_p = 7$, and control horizon $N_c = 2$ are considered based on controller tunning and the computational efficiency results of [32] for a similar 7-region urban network. The MPC control problem is a non-convex non-linear programming problem, and its solution is obtained using the interior-point algorithm in MATLAB, fmincon function. All scenarios are simulated using MATLAB R2016a (9.0.0), on a 64-bit Windows PC with 4-GHZ Intel Core-i7 and 24-GB of RAM.

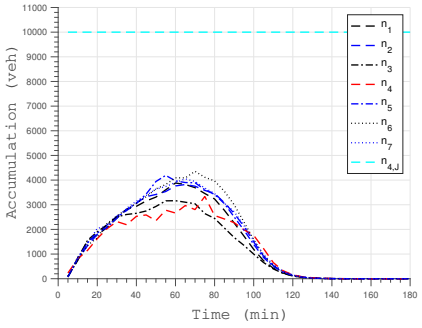
Figures 2.4 and 2.5 show regions' accumulation states under different demand profiles. In the medium uniform demand scenario, all regions are emptied at a same time at t=130 min while in the radial demand scenario, with 5% higher demand towards CBD, regions are emptied at t=155 min. In the later scenario, maximum accumulation level of CBD reaches 8000 veh (80% of jam accumulation), which is approximately 20% higher compared to the uniform demand scenario. As suggested by accumulation graphs of Figure 2.4 (c-f), BC and AC perfectly handle traffic condition in CBD by metering the inflow to this region. As a result of restraining the inflow to CBD and vehicles' rerouting to peripheral regions, accumulation level of these regions has slightly increased. Under BC, regions' accumulation level are close while under AC, higher accumulation level for CBD compared to peripheral regions is obtained. This result might seem counter-intuitive but in fact in order to obtain the lowest total time spent (TTS) for the whole network, it is reasonable for CBD to carry more traffic as it provides the shortest route between any two non-neighbor regions in the network. TTS values of traffic network under different demand and control scenarios during a 3-hour control period is recorded in Table 3.1. These values are average over ten simulation runs. Related to NC, BC and AC improve TTS 3.36% and 8.18% in the medium uniform demand scenario, and 7.09% and 12.56% in the medium radial demand scenario, respectively. Comparing these values with TTS improvement of 9.40% for medium uniform and 17.85% for medium radial scenario under SO, indicates that AC performs well in achieving a near system optimum condition in medium demand scenarios [Figure 2.4 (g,h)].



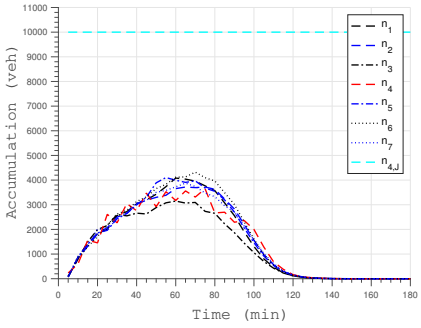
(a) NC (uniform)



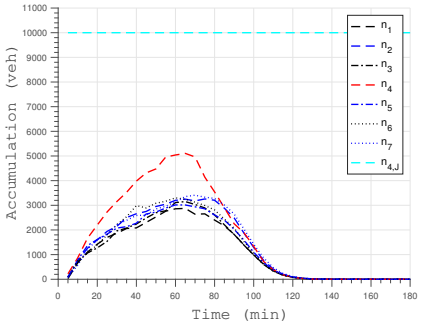
(b) NC (radial)



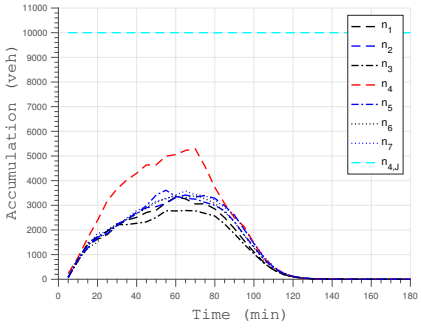
(c) BC (uniform)



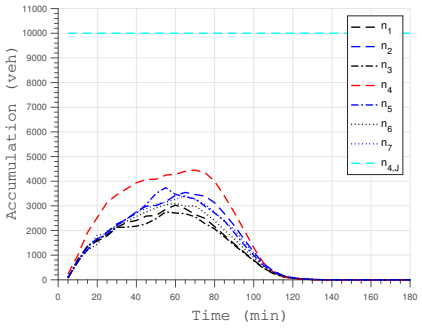
(d) BC (radial)



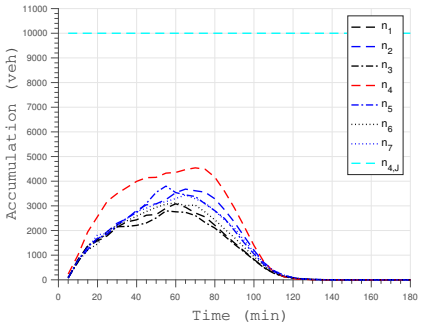
(e) AC (uniform)



(f) AC (radial)

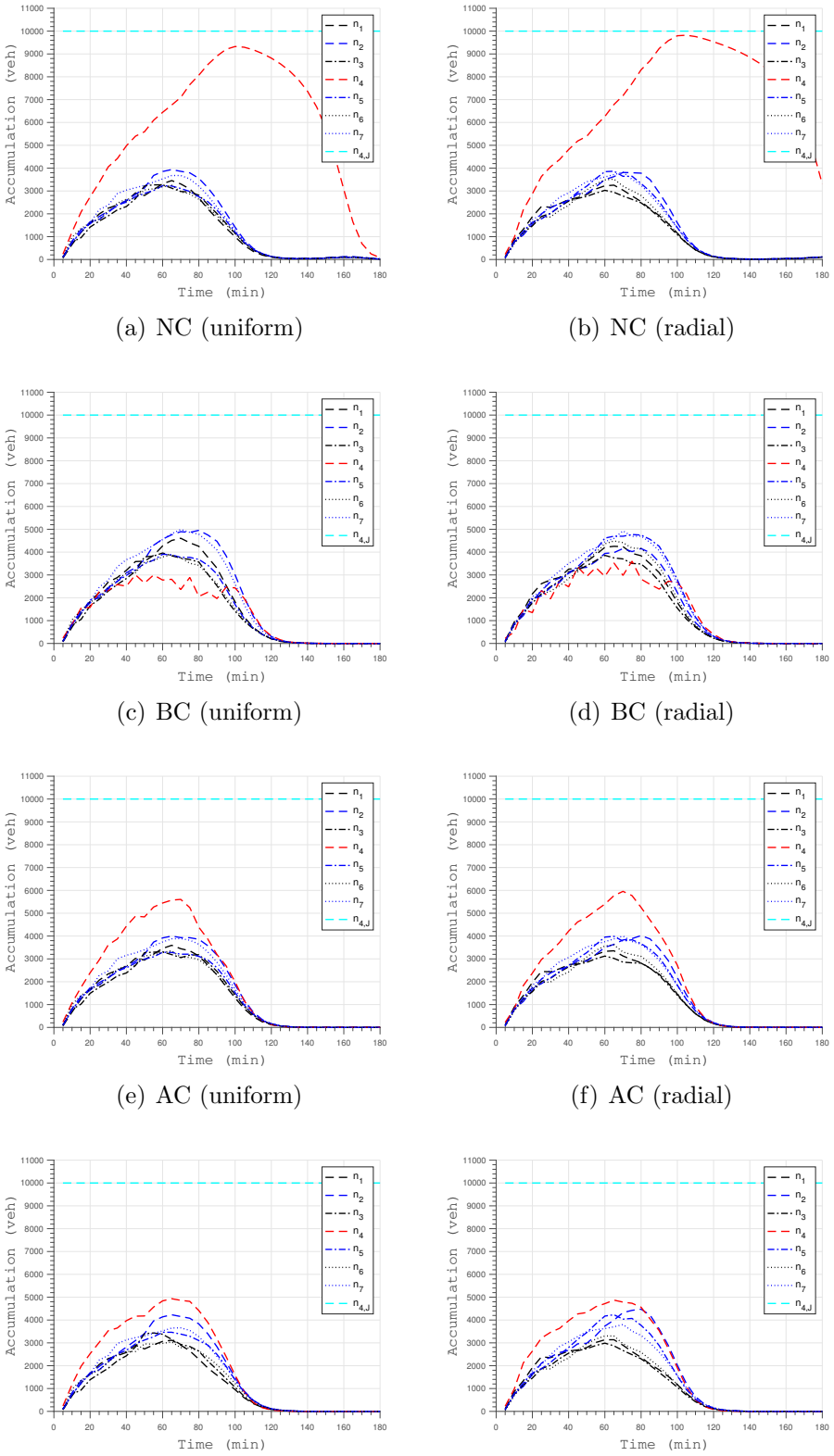


(g) SO (uniform)



(h) SO (radial)

Figure 2.4: Regional accumulations in medium uniform demand scenario (a) NC, (c) BC, (e) AC, (g) SO; and in medium radial demand scenario (b) NC, (d) BC, (f) AC, (h) SO.



(a) NC (uniform)

(b) NC (radial)

(c) BC (uniform)

(d) BC (radial)

(e) AC (uniform)

(f) AC (radial)

(g) SO (uniform)

(h) SO (radial)

Figure 2.5: Regional accumulations in high uniform demand scenario (a) NC, (c) BC, (e) AC, (g) SO; and in high radial demand scenario (b) NC, (d) BC, (f) AC, (h) SO.

In uncontrolled high demand scenarios, CBD mostly operates at high accumulation levels during the 3-hour simulation period [Figure 2.5 (a, b)], which corresponds to low trip completion rates for this region during the same time. In the high uniform demand scenario, peripheral regions are emptied at t=130 min while CBD region is emptied at t=180 min. In the high radial demand scenario, the 5% higher generated demand for CBD, significantly affects the traffic network; CBD region reaches gridlock at t=100 min, and is not emptied at the end of the 3-hour simulation period while peripheral regions are emptied at t=130 min. This result indicates under-utilization of the available capacity at peripheral regions. Similar to medium demand scenarios, herein, BC and AC reduce CBD accumulation level at the cost of increasing number of vehicles at peripheral regions [Figure 2.5 (c-f)]. While both BC and AC are effective in alleviating congestion in CBD, BC keeps its accumulation level unnecessarily lower than peripheral regions at any time instant. Further, the evolution of traffic state under BC is not smooth, which can be attributed to alternating fluctuation between minimum and maximum control rates, similar to bang-bang controller [4]. Under AC, all regions' accumulations are maintained at an acceptable level with CBD experiencing higher accumulation as it is the shortest route linking any two non-neighbor regions in the network. TTS values recorded in Table 3.1, indicate that BC and AC respectively improve TTS 14.85% and 22.05% in the high uniform demand scenario, and 19.21% and 25.94% in the high radial demand scenario. Under SO, a TTS improvement of 26.98% for the uniform demand scenario and 30.67% for the radial demand scenario are observed. Comparing these values with those of AC, indicates efficacy of AC in achieving a near system optimum traffic condition in high demand scenarios [Figure 2.5 (g,h)].

The results suggest significant improvement in TTS under AC especially in radial demand scenarios and highlight the importance of metering the inflow at the borders of CBD region. In congested condition, even use of a simple BC strategy can significantly improve the traffic condition and avoid gridlock. These results confirm the findings

Table 2.1: Total time spent (veh.min) of the network (the percentage improvements in TTS compared to NC are shown in parentheses)

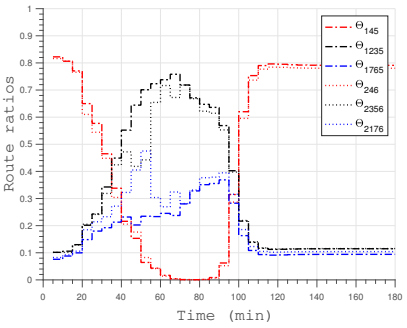
	Demand pattern				CPU time (s)
Scenario	Medium-uniform	Medium-radial	High-uniform	High-radial	Avg.
NC	$1.8023 \cdot 10^6$	$2.0299 \cdot 10^6$	$2.4566 \cdot 10^6$	$2.6712 \cdot 10^6$	0.0383
	-	-	-	-	
BC	$1.7417 \cdot 10^6$	$1.8860 \cdot 10^6$	$2.0918 \cdot 10^6$	$2.1580 \cdot 10^6$	6.0148
	(3.36%)	(7.09%)	(14.85%)	(19.21%)	
AC	$1.6548 \cdot 10^6$	$1.7749 \cdot 10^6$	$1.9148 \cdot 10^6$	$1.9783 \cdot 10^6$	101.5514
	(8.18%)	(12.56%)	(22.05%)	(25.94%)	
SO	$1.6328 \cdot 10^6$	$1.6676 \cdot 10^6$	$1.7937 \cdot 10^6$	$1.8519 \cdot 10^6$	0.0445
	(9.40%)	(17.85%)	(26.98%)	(30.67%)	

from [32] that found perimeter control highly beneficial for networks with high number of destinations in CBD; however, BC tends to implement strict control due to ignoring the fact that part of alleviation process is done by vehicles that reroute to other parts of a network as a result of the perimeter control and the created queue. In fact, perimeter control affects traffic by limiting the inflow to the controlled area and encouraging travellers to reroute by creating queue at the boundaries of the controlled area. Comparison of regional accumulation and TTS value of AC and SO, indicates that the performance of AC is comparable to that of SO, for all demand levels and patterns. The average running time for AC at each time step under different demand profiles as noted in Table 1 is 101.5514 sec, which is less than the assumed control interval (300 sec). This implies that the proposed AC is suitable for real-time applications.

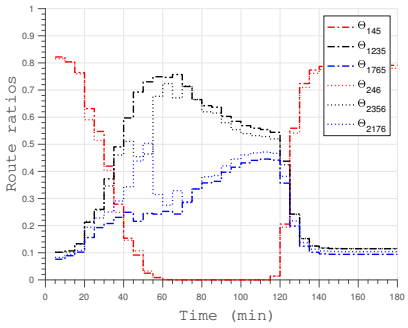
Instances of trip distribution between alternative routes, under different demand profiles, are shown in Figures 2.6 and 2.7. In the uncontrolled medium uniform demand scenario, from t=30 to t=100 min, CBD operates at high accumulation level, which corresponds to high travel time. Travellers, being aware of alternative routes travel time, reroute to less congested routes to reach their destinations (logit parameter $\phi = 0.5$ is considered in all scenarios). θ_{145} and θ_{246} gradually reduce, resulting in an increase in θ_{1235}, θ_{1765}, θ_{2176}, and θ_{2356} values during the same time [Figure 2.6 (a)]. From t=65 to t=85 min, CBD is not available for taking any trip between peripheral regions. Starting

t=85, the decreasing trend of demand and accumulation level of CBD, gradually return the traffic to this region. The similar pattern is observed in medium radial demand scenario; however, due to the higher level of demand for CBD, the region is unavailable for external trips for a longer period, from t=60 to t=115 min [Figure 2.6 (b)]. In uncontrolled high demand scenarios, from t=55 to t=150 min, in the uniform demand scenario, and from t=60 to t=175 min, in the radial demand scenario, routes 1-4-5 and 2-4-6 are not taken for any trip between peripheral regions Figure [2.7 (a,b)]. CBD region operates at extremely high accumulation levels that affects travellers' routing behaviour. Similar to medium demand scenarios, θ_{145} and θ_{246} reduce while θ_{1235}, θ_{1765}, θ_{2176} and θ_{2356} increase. This trend continues until t=150 min for uniform demand and t=175 min for radial demand, when traffic finally starts to return to CBD.

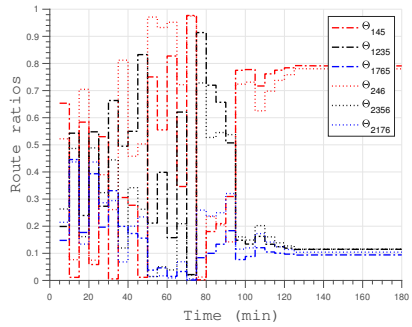
In controlled scenarios, when CBD moves towards congested condition, perimeter controllers meter the inflow at CBD perimeter boundaries as reflected in the control ratios shown in Figure 2.8. Metering the inflow at the regional perimeter boundary and the resulting queue, increase travel cost of taking CBD, and thus similar to NC, rerouting to peripheral regions occurs; however, in controlled scenarios, metering the inflow to CBD resolves congestion and facilitates return of traffic to this region earlier than what is observed in uncontrolled scenarios. In the medium uniform demand scenario, when BC is employed, alternating intervals of route change between CBD and peripheral regions are observed [Figure 2.6 (c)]. The rerouting effect is more extreme in the medium radial demand scenario with higher demand for CBD [Figure 2.6 (d)]. Comparing route split ratios under BC in medium and high demand scenarios indicates that the overall increase in demand intensifies the rerouting effect as shown in Figure 2.7 (c,d). On the other hand, routing pattern under AC mostly shows smooth transition of vehicles from CBD to peripheral regions and vice versa, and sudden rerouting between alternative routes as detected in BC scenarios, is less observed [Figures 2.6 and 2.7 (e, f)]. In addition, routing behaviours in uncontrolled and controlled scenarios are notably different. For instance, in the uncontrolled high uniform demand scenario,



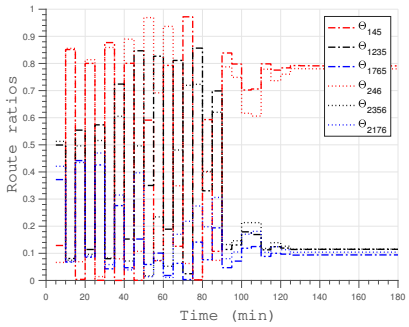
(a) NC (uniform)



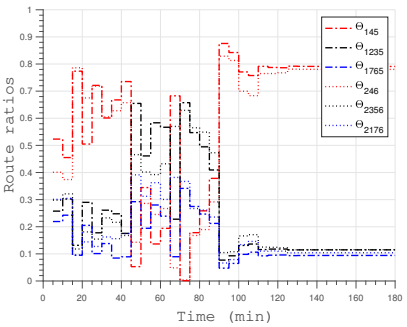
(b) NC (radial)



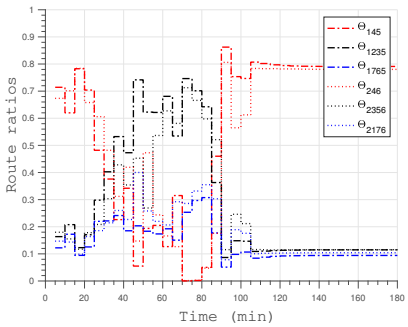
(c) BC (uniform)



(d) BC (radial)

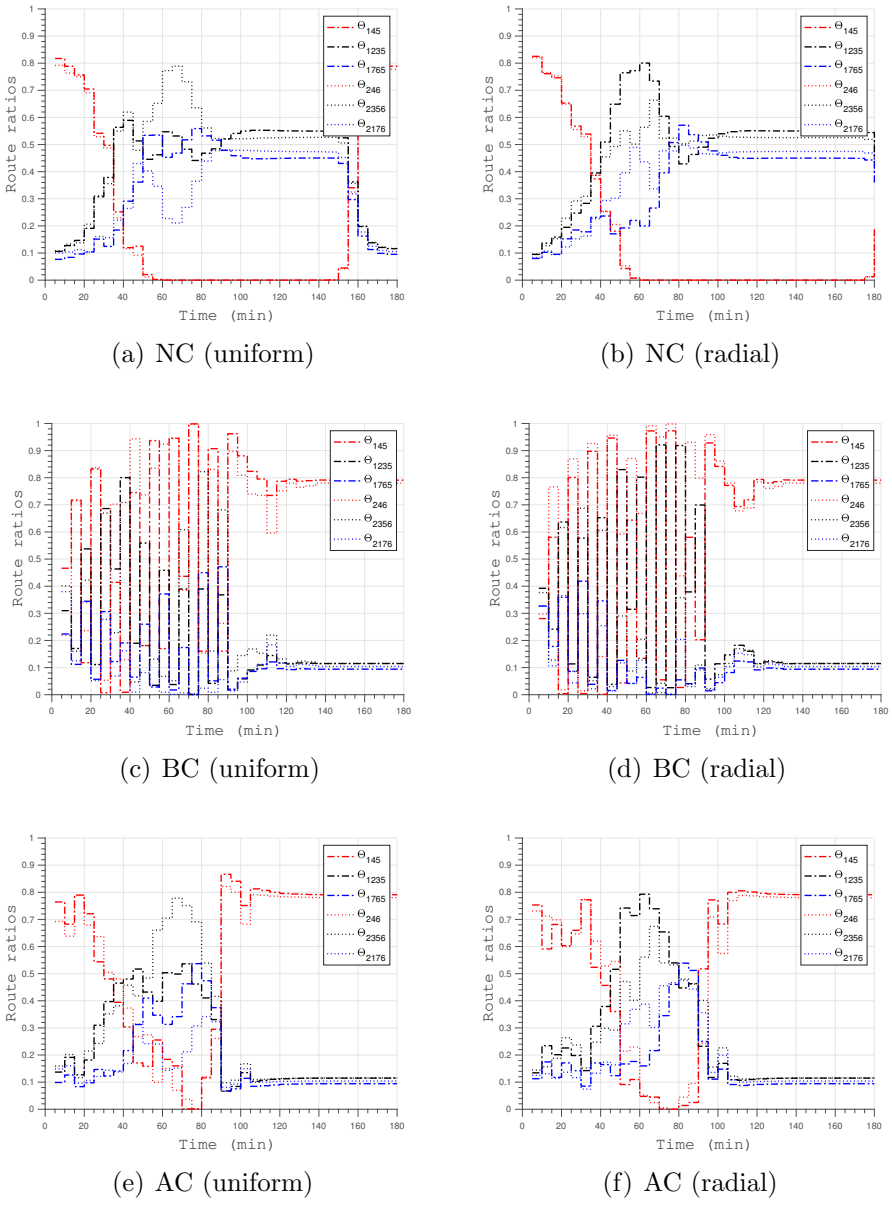


(e) AC (uniform)



(f) AC (radial)

Figure 2.6: Route split ratios in a medium uniform demand scenario (a) NC, (c) BC, (e) AC; and medium radial demand scenario (b) NC, (d) BC, (f) AC (logit parameter $\phi = 0.5$).



(a) NC (uniform)

(b) NC (radial)

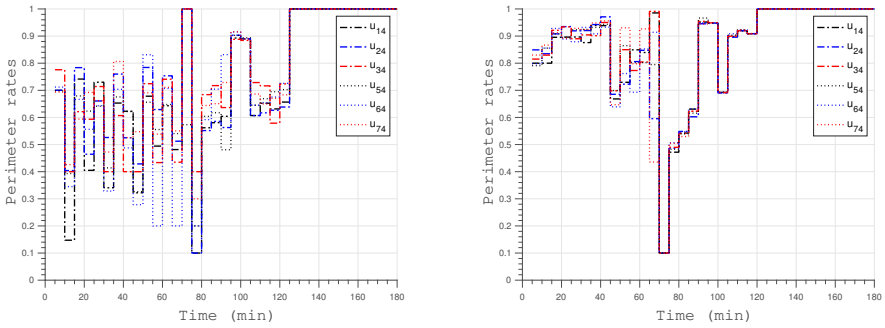
(c) BC (uniform)

(d) BC (radial)

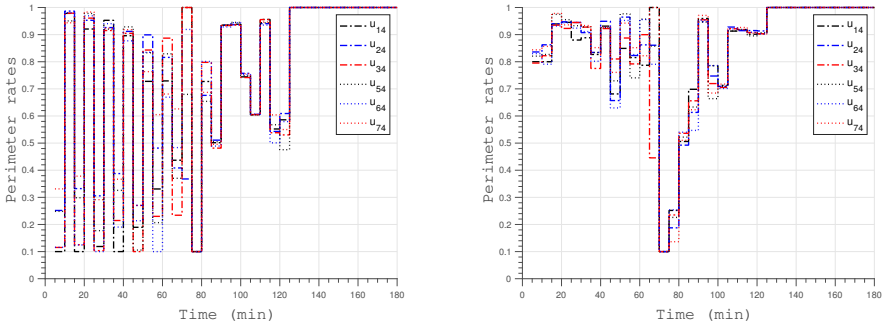
(e) AC (uniform)

(f) AC (radial)

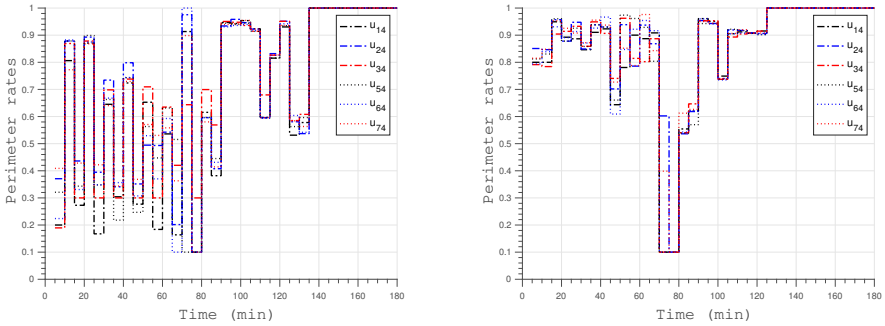
Figure 2.7: Route split ratios in a high uniform demand scenario (a) NC, (c) BC, (e) AC; and high radial demand scenario (b) NC, (d) BC, (f) AC (logit parameter $\phi = 0.5$).



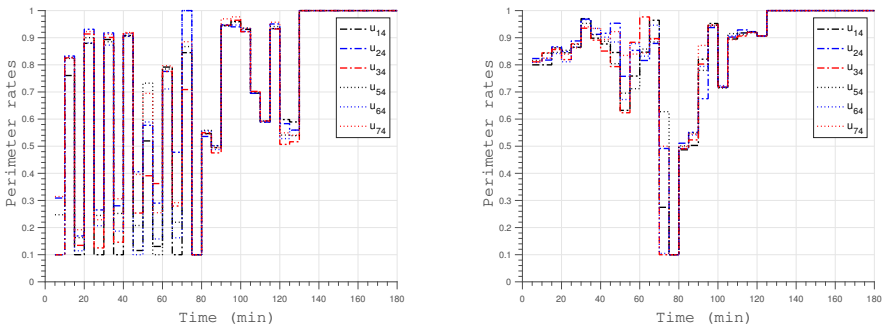
(a) BC (b) AC



(c) BC (d) AC



(e) BC (f) AC



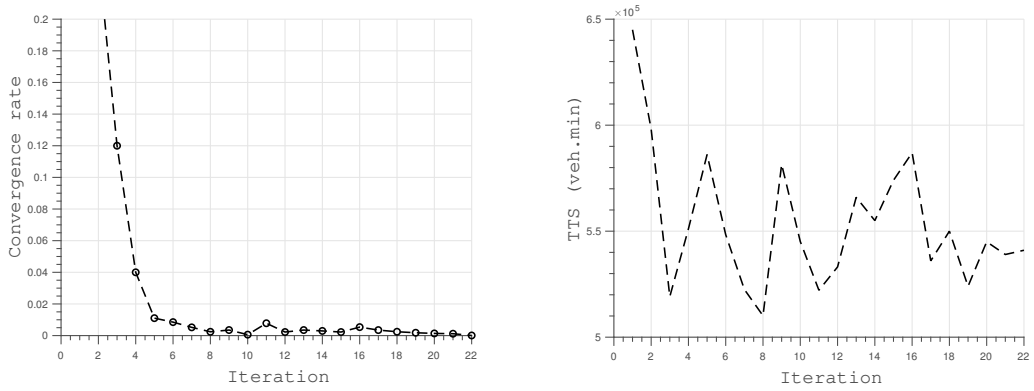
(g) BC (h) AC

Figure 2.8: Perimeter control rates in a medium uniform demand scenario (a) BC, (b) AC; medium radial demand scenario (c) BC, (d) AC; high uniform demand scenario (e) BC, (f) AC; and high radial demand scenario (g) BC, (h) AC.

starting t=60 min, travellers completely switch from route 1-4-5 to routes 1-2-3-5 and 1-7-6-5 until t=150 min, when congestion in the CBD region is alleviated; however, in the AC scenario, during the same time interval, an average of 45% of travel demand between region 1 and 4 is traversed through route 1-4-5. The reason behind the difference in route choice and pattern, under different control scenarios can be explored by analyzing implemented control rates and the way they manipulate network traffic condition.

The perimeter control rates obtained from MPC controller under different demand profiles are shown in Figure 2.8. Under BC, alternating fluctuation in control rates is observed, which increases with increase in demand and congestion level; however, as congestion is mitigated, control rates are maintained around higher values with less variation [Figure 2.8 (a,c,e,g)]. Variation in control rates justifies the fluctuation observed in route split ratios of Figures 2.6 and 2.7 (c,d). BC does not take into consideration the rerouting effect induced from implementation of control; vehicles that reroute to peripheral regions accelerate the congestion alleviation process, and thus, under AC, with consideration of rerouting effect, higher rates for perimeter control as well as a smoother pattern are achieved [Figure 2.8 (b,d,f,h)]. In the high radial demand scenario, AC control rates are between 0.78 and 1 for the first 40 minutes of control period while BC control rates are in the range of 0.17 and 0.9, during the same time period.

In the proposed AC framework, initially control rates are optimized, and based on the estimated control rates, travellers' routing preferences are predicted. Using the current route split ratios, control rates are re-estimated. The iterative procedure continues until a mutually consistent solution, in which the difference between flow patterns of two successive iterations is less than a pre-defined convergence tolerance, is found. [68] showed that the solution obtained using iterative approaches might not be an optimal solution, and total network cost can increase in iterations. For the seven-region urban network of this study, IOA convergence is achieved within finite number of iterations with final TTS less than the initial values, under all demand profiles. An example of



(a) (b)

Figure 2.9: In a high uniform demand scenario at t=8 (a) convergence rate (convergence threshold=0.001), and (b) TTS values.

the evolving congestion pattern and the associated TTS values in high uniform demand scenario at t=8 is given in Figure 2.9, which indicates that convergence is achieved in 22 iterations. On average around 30 iterations are required for AC to converge at different time instants and under different demand profiles.

Consideration of route choice behaviour, while controlling the urban network, is essential. When congestion occurs, vehicles that reroute to the less congested part of the network help to mitigate the congestion. However, as the situation in the controlled region improves, part of the rerouted traffic returns. Consideration of this issue is important in devising effective control schemes. An important issue to study is the sensitivity of the rerouting behaviour to control measures and rerouting effect on the shape and scatter of MFD in large-scale urban networks.

2.5 Conclusion

This chapter presents an anticipative control approach for real-time network-wide control of an urban network. The proposed AC incorporates drivers' route choice behaviour endogenously into the control framework rather than exogenously as an input to the model. Performance of the proposed AC is compared to uncontrolled and basic control

scenarios. Using the numerical results of the 7-region urban network, it is observed that the proposed AC outperforms NC and BC and improves the overall mobility of the network measured in terms of TTS, under different demand profiles. Further, it is demonstrated that integrating route choice into the control framework, moves the network towards system optimum traffic condition, which is used as a benchmark of an ideal control strategy with best optimality.

The perimeter control employed in this study, meters the inflow entering to CBD from its neighboring regions, and thus, enhances traffic condition in the controlled CBD area; however, the resulting queue and waiting time at the regions' perimeter boundaries, has significant impact on drivers' tendency to choose CBD for their trips. Increase in queue length encourages rerouting behaviour; therefore, the proposed AC integrates drivers' routing behavior at the network-wide level using MFD and a point-queue model at regional perimeter boundaries. To reduce the computational burden and make a centralized real-time control possible, both control and traffic assignment are conducted at regional level.

The queue resulting from implementation of perimeter control can adversely affect the existence of a well-defined MFD. The spatio-temporal dynamics of queue and its effect on maximum outflow and heterogeneity of the MFD-based network has to be investigated and included in the control framework. In addition, this chapter assumes that flow propagates instantaneously throughout the network. For future research, the proposed AC framework can be extended to consider regional time-delay [69]. This closer representation of regional delay is expected to improve the estimation of regional travel time and closely describe important phenomenon of traffic networks such as dynamic queuing and congestion; thereby enabling more efficient control schemes.

For field implementation of the developed AC, some limitations exist. Firstly, the concept of perimeter control for urban networks is not properly defined for real traffic systems. While in freeway networks, perimeter control is done using ramp metering, in urban networks many connections between any two regions exist. Therefore, a lower-

level controller should be devised to properly meter the flow at any local controller. Further, it is not clear to what extend controlling the flow between regions is practical since as mentioned hundreds of connecting links can link the two region and control the flow at all the connection points is cumbersome. Lastly, road users' routing is modeled at the regional level, which for real traffic system needs further treatments prior to the implementation.

Chapter 3

Fairness-aware and efficient large-scale urban network control: A macroscopic fundamental diagram approach

This chapter introduces a proportional fair perimeter control (PFPC) model for a multi-region urban network based on the macroscopic fundamental diagram (MFD) modeling approach. The proposed control scheme, modeled in a model predictive control framework, optimizes the flow that is allowed to transfer between any two neighbor regions in a network while considering the equitable distribution of shared resources (i.e., available capacity) and the impacts of such decisions (i.e., delays at queues) in the regions. Regional traffic state dynamics are modeled using an MFD approach, and queuing dynamics at a regional perimeter, caused by a control, is modeled using a point queue model. The utility of each region is expressed based on regional average speed, and a novel perimeter control model that identifies the inflows between any two neighbor regions is formulated. The developed perimeter control model considers road users' trip utility without sacrificing efficiency for fairness. The performance of the proposed model is compared with no control and basic control scenarios. The model enhances both efficiency and fairness goals while degrading queuing dynamics.

3.1 Introduction

Modeling and control of large-scale urban networks based on microscopic traffic models is challenging and impractical due to the high level of detail required and specific features of urban networks, such as short links, traffic signals, and unpredictability of road users' behavior. Thus, the control schemes developed based on microscopic traffic models are inefficient and suffer from tractability issues. Local controllers fall short in mitigating traffic congestion and coordinating with other controllers in the network, whereas centralized controllers would impose a huge computational burden on a system. Macroscopic fundamental diagram (MFD) is developed as an aggregated traffic model that depicts traffic states of an urban network based on a single-state variable and, thus, provides an elegant and reliable tool for modeling and control of large-scale urban networks at a network-wide level.

[4] described dynamics of a homogeneous urban network by considering the network as a single reservoir where outflow was a function of the number of vehicles in the network. [6] derived analytical theories for MFD, and [1] verified the existence of MFD in the Yokohamas, Japan traffic network by integrating the highly scattered plots of flow versus density. MFD was adopted as a modeling approach that underlies different control schemes, such as perimeter control, route guidance, and congestion pricing. Perimeter control was adopted in previous MFD studies to enhance efficiency at the controlled areas of a network through metering the transferring flow between adjacent reservoirs. [4] developed a perimeter control model that metered the inflow to a single-region network to maintain maximum outflow conditions using a simple control rule. [15] advanced a perimeter control framework for a two- region urban network that had a freeway as an alternative route using optimization techniques. The objective function was considered to reduce the total network accumulation, a measure that is equal to the total network travel time, as discussed by [29]. [18] advanced a two-level perimeter control model with the goal of reducing total network travel time while simultaneously

minimizing the spatial heterogeneity of congestion within each region. [70] coupled two competing control objectives for a perimeter control problem to optimize network performance at local- and network-wide levels. The objective function was considered in terms of travel time, which minimized the sum of regions' travel time and intersection delays. [71] developed an anticipatory perimeter control scheme that consisted of an upper-level perimeter control that minimized total travel time and a lower-level user equilibrium that minimized road users' travel time. The approach found a mutual solution that satisfied both controllers' and road users' goals and constraints. Other perimeter control schemes have been discussed in the MFD literature [4, 27, 28, 29, 16, 18].

So far, the majority of MFD-based control models described in the literature have been developed with the main intention of minimizing total network travel time or maximizing outflow or production. None of the published studies, however, accounts for equity and fairness in developing and evaluating control schemes. One leading effort in this direction is a work by [16] who developed a perimeter inflow control approach that considered equity by keeping each region around a desired critical accumulation level while minimizing spatial and temporal heterogeneity of congestion within each region. The homogeneity of each region was maximized by equalizing the distribution of the relative accumulation inside a region in time and space. This study used simple multivariable and integral feedback regulators to control the inflow that transferred between any two regions; however, the impact of these control decisions on equity was not measured. [17] also considered equity by modeling a regional route guidance as a multi-objective optimization problem to minimize total network travel time and regional average speed difference between regions within the network. They discovered that minimizing total network travel time without specific attention to each region yielded high travel time in some regions, resulting in low average regional speed for those regions.

Efficiency, defined in terms of minimizing total time spent of a network, as a con-

trol goal is often achieved at the cost of increased travel time for some travelers. For instance, if a system's objective is to minimize total travel time, saving 1 minute of travel time for one group could result in a 5 minute increase in travel time for another group. However, depending on the prevailing congestion levels, a fair control scheme can be inefficient and result in low trip completion rates for some regions. In an urban traffic network with multiple MFD-based regions, conventional perimeter control methods maintain accumulation around a critical level without differentiating between transferring and queuing vehicles. The control that is forced on a network may cause long queues at regions' perimeters. Strict control and the resulting queue formation not only disturb equity but may also lead to deteriorating traffic performance rather than the intended amelioration. Queueing might block traffic traveling to other regions, causing heterogeneity in the MFD and reduction in regional travel production. While only a few published studies account for the effect of queue at perimeters, it is vital to consider queues when optimizing for either efficiency or fairness.

Perimeter control is similar to the well-known ramp metering (RM) control for freeway traffic networks; both problems operate based on restricting inflow to the protected area of a network. Since equity and fairness measures have been previously investigated in the freeway RM literature, it is beneficial to review a few RM models that have considered these measures. A basic RM reduces congestion and improves efficiency in mainline freeways, while an equitable RM fairly distributes travel time reduction benefits among mainline freeways and on-ramps. The equity and fairness measures considered in the literature are the equal distribution of a weighted sum of delay among on-ramps [72, 73, 74], equal maximum queue length threshold [75, 76], and RMs that are developed based on the Gini coefficient [77, 78, 79]. In these studies, equity is not the only control goal; rather, it is considered along with efficiency measures, such as total travel time or total system delay, and is prioritized using a hierarchical control scheme [74] or a weighted multi-objective optimization problem [72].

These studies emphasize the importance of a trade-off between equity and efficiency.

Introducing fairness and equity measures, such as maximum on-ramp queue threshold, leads to a more equitable but less efficient control [75]. [73], however, proposed an equitable RM and discussed that fairness and efficiency are not necessarily contrasting objectives. Proportional fairness, first introduced in wireless networks [80] and recently in transportation [81, 82], can be adopted to find a trade-off between efficiency and fairness. In fact, proportional fairness is rooted in economics and is a generalization of the Nash bargaining game [83] in which the underlying concept is improvement of the proportion of users' utilities. The underlying concept of proportional fairness is that without sacrificing efficiency, an equitable allocation of shared resources can be achieved by improving the proportion of users' utilities. A proportionally fair allocation is equivalent to maximizing the summation of the proportions of the users' utilities [80, 81, 82]. In other words, if we remove some resources from one user and allocate them to another user, and this move reduces the utility of the first user by 10% but adds to the utility of the other user by 12%, this move is beneficial overall because it increases the total proportion of users' utilities.

In this chapter, a proportionally fair perimeter control (PFPC) model for a multi-region urban network is introduced in a model predictive control (MPC) framework. Regional traffic states are modeled using the MFD approach, and queuing at the perimeters, caused by a control or capacity constraint, is modeled using a point queue (PQ) model (i.e., vertical queue that do not occupy any physical space) and based on the disaggregation method of [28]. The objective function maximizes the weighted utility of road users in terms of regional speed while considering the degrading effect of queuing dynamics on network homogeneity and trip production. Fairness is also considered using maximum queue threshold constraints in the optimization model. To the best of our knowledge, this study is the first attempt to introduce the concept of proportional fairness in traffic control research, especially for large-scale urban networks. The proposed model can appropriately respond to the varying nature of urban traffic states and can be adopted for real-time control applications. This model lies under responsive

rather than pre-timed control models and is capable of dealing with both recurrent and non-recurrent congestion.

With respect to theoretical contributions, this thesis incorporates queuing dynamics into the MFD model. Incorporation of queuing dynamics into MFD is tackled in an earlier research [28] for two region urban networks and herein is extended to the multi-region urban network. In addition, the concept of regional utility is introduced and expressed as a function of regional speeds that can be continuously monitored. Fairness is also incorporated into the distribution of the impacts of perimeter control and, thus, is considered as part of the evolution of queue dynamics. From a methodological standpoint, a fair perimeter control problem in an MPC framework that can proactively deal with traffic conditions is developed. As opposed to traditional perimeter control approaches that solely focus on efficiency, a precise mathematical approach that considers fairness as an integral part of the perimeter control problem for large-scale networks partitioned into MFD-based subregions is proposed. A new objective function that introduces the novel concept of proportional fairness is developed for the first time in the traffic flow control and MFD literature. The objective function incorporates fairness as a precise mathematical formulation as maximizing the weighted sum of the logarithms of a region's utility. Introducing weights is crucial to make the approach responsive to both recurrent and non-recurrent congestion. For instance, in the case of a major incident occurring in a given region, the weights for this zone can be increased to draw attention of the perimeter control to prioritize the utility of this region. Also, equity is considered in the queue constraints. This novel PFPC approach is effective in meeting the fairness objectives without sacrificing efficiency.

3.2 Fair perimeter control approach

The proposed fair perimeter control scheme is modeled in a model predictive control (MPC) [59, 60] framework, as shown in Figure 3.1. At each simulation time step,

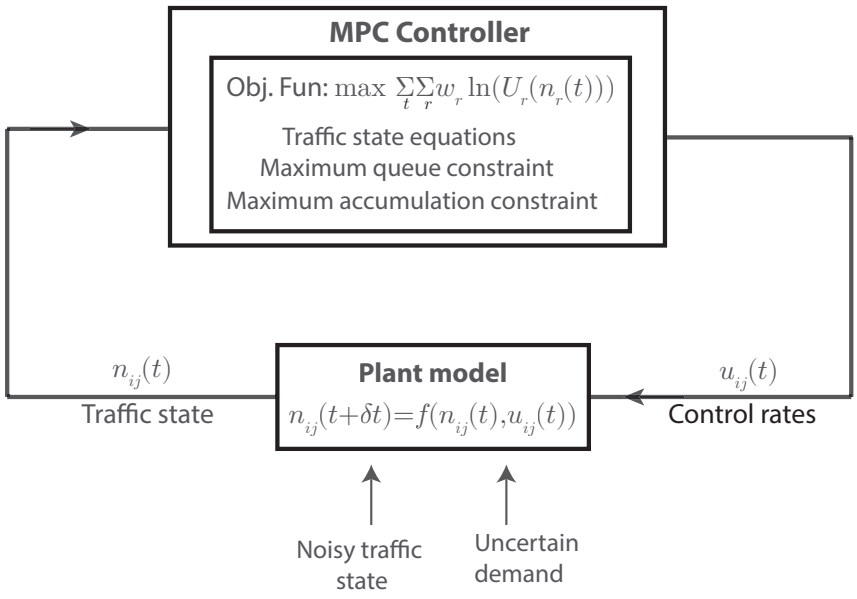


Figure 3.1: Schematic view of the proposed fair perimeter control.

MPC proactively optimizes a pre-defined cost function over a prediction horizon while satisfying the existing constraints. The first sample of the control is applied, the horizon is shifted forward, and control rates are optimized using updated traffic states. The cost function is considered as a weighted sum of the utility functions. Traffic state constraints are constructed based on accumulation-based MFD and consider queues. The maximum queue length is constrained to keep the MFD well-defined.

The plant is modeled at the same abstract level as the prediction model and has access to uncertain demand and noisy accumulation values. MPC can handle uncertainty, disturbances, and mismatch errors that exist in the model as a result of data collection noises and demand prediction errors. Further, it can easily handle constraints and the nonlinear dynamics of a control problem. MPC for the MFD-based perimeter control problem was first proposed by [30] for a simple two-region urban network and was later used for more complicated perimeter control problems [15, 17, 18, 67]. This approach effectively handles traffic conditions in a congested network compared to simple control approaches, such as bang-bang controllers [4] that cause uneven distribution of vehicles,

increase heterogeneity, and reduce total network outflow when used in a multi- region urban network, and PI-type controllers [29, 16] that perform unsatisfactorily in high demand and saturated traffic conditions.

3.3 MFD-based modelling

3.3.1 Dynamics of a single reservoir

Consider a heterogeneous urban network separated into m connected homogeneous regions that have well-defined MFDs. Denote R as a set of regions, W as a set of OD region pairs, and P_w as a set of routes between any OD pair $w \in W$. Dynamics of any region $r \in R$ is described using a well-defined low scatter relationship between accumulation (number of circulating vehicles within a region) n_r (veh) and travel production $P_r(n_r)$ (veh.m/sec) or outflow $G_r(n_r)$ (veh/s), referred to as production-MFD or accumulation-based MFD.

A continuous mathematical form of a single region's traffic state as described by a mass conservation equation is as follows [4]:

$$\dot{n}_r = f_{r,in} - f_{r,out} \qquad \forall r \in R \tag{3.1}$$

In the above equation, each region in an MFD-based traffic network is described as a reservoir, where the dynamics of the reservoir evolve due to the difference between its inflow $f_{r,in}$ (both initiated and transferred trips) (veh/sec), and outflow $f_{r,out}$ (veh/s) (both terminated and transferred trips). Outflow of any region r is approximated using the MFD function $G_r(n_r)$, which indicates the number of trips terminated within a region or transferred to another region as a function of regional accumulation n_r.

The ratio of travelling production to outflow yields average regional trip length $\bar{L} = \frac{P_r(n_r)}{G_r(n_r)}$ (m). As demonstrated by [1], this ratio indicates a relatively constant value throughout a day and, thus, for the sake of simplicity, regional trip length is

usually considered constant in accumulation-based models. The average regional speed is obtained as the ratio of traveling production to accumulation (*i.e.*, $\bar{V}_r(n_r) = \frac{P_r(n_r)}{n_r}$). The average regional travelling time $\tau_r(n_r)$ can be estimated as the ratio of average trip length inside a region (assuming internal and external trips are equal and independent of trip origin, destination, and route choice) to the average regional speed; therefore, the regional travel time is derived as follows [15]:

$$\tau_r(n_r) = \frac{n_r}{G_r(n_r)} \qquad \forall r \in R \tag{3.2}$$

3.3.2 Accumulation-based MFD considering boundary queues

To model the dynamics of boundary queues, created as a result of capacity constraints or control, we use a simple input-output balance differential equation following [28]. Thus, vehicles that reach the borders of two regions are excluded from MFD dynamics and belong to queuing dynamics. Hence, accumulation state n_r can be divided into three dynamical modes: (i) travelling accumulation $n_{r,T}$: vehicles that travel in a region towards a boundary queue with a neighbor region (ii) queuing accumulation $n_{r,Q}$: vehicles that wait in a boundary of a region to enter a neighbor region, and (iii) internal accumulation $n_{r,I}$: vehicles that initialize and terminate their trips within the same region. Therefore, the total accumulation of any region r at any time t can be estimated as the sum of the three above-mentioned accumulations inside a region [28]:

$$n_r(t) = n_{r,T}(t) + n_{r,Q}(t) + n_{r,I}(t) \qquad \forall r \in R \tag{3.3}$$

To estimate the dynamics of queues between any two adjacent regions in a multi-region urban network that has multiple routes, accumulation values have to be further disaggregated based on paths. Extending the two-region MFD dynamics developed by [28] to a general multi-region MFD, and disaggregating based on paths, dynamic state

equations for traveling and queuing vehicles are derived as follows:

$$\dot{n}^p_{r,T}(t) = f^p_{r,in}(t) - \frac{n^p_{r,T}(t)}{n_r(t)} G_r(n_r(t)) \qquad \forall r \in R, p \in P \tag{3.4}$$

$$\dot{n}^p_{r,Q}(t) = \frac{n^p_{r,T}(t)}{n_r(t)} G_r(n_r(t)) - f^p_{r,out}(t) \qquad \forall r \in R, p \in P \tag{3.5}$$

Each path p in an MFD network is defined as a sequence of regions from an origin to a destination in the form $p = \{p_{r_1}, p_{r_2}, ..., p_{r_{m_p}}\}$ where m_p is the number of regions on a given path p. The path-disaggregated values of inflow $f^p_{r,in}$, and outflow $f^p_{r,out}$ of region r, depending on their placement along path p, are calculated as follows:

$$f^p_{r,in}(t) = \begin{cases} d^p_r(t) & \text{if} \quad r = p_{r_1} \\ \hat{f}^p_{p^{(-r)},r}(t) & \text{otherwise} \end{cases} \qquad \forall r \in R, p \in P \tag{3.6}$$

$$f^p_{r,out}(t) = \begin{cases} \frac{n^p_{r,T}(t)}{n_r(t)} G_r(n_r(t)) & \text{if} \quad r = p_{r_{m_p}} \\ \hat{f}^p_{r,p^{(+r)}}(t) & \text{otherwise} \end{cases} \qquad \forall r \in R, p \in P \tag{3.7}$$

Where d^p_r denotes the newly generated demand in region r carrying via path p. $p^{(-r)}$ and $p^{(+r)}$ respectively denote the regions that precede and succeed region r on path p. $\hat{f}^p_{r,p^{(+r)}}$ denotes the restricted transferring flow from region r to the next region $p^{(+r)}$ on path p, estimated as a minimum of two terms: (i) (i) total transferring demand (summation of queuing accumulation at the border of two regions and transferring vehicles joining the boundary queue during simulation time step intervals) and (ii) a fraction of the remaining receiving capacity of the succeeding neighboring region assigned to the demand from region r. The equation is solved as follows during simulation time interval T :

$$\hat{f}^p_{r,p^{(+r)}}(t) = 1/T \min\{n^p_{r,Q}(t) + T\frac{n^p_{r,T}(t)}{n_r(t)} G_r(n_r(t)), \frac{\frac{n^p_{r,T}(t)}{n_r(t)} G_r(n_r(t))}{\sum_{p \in P} \frac{n^p_{r,T}(t)}{n_r(t)} G_r(n_r(t))} C_{rp^{(+r)}}\} \quad \forall r \in R, p \in P \tag{3.8}$$

Where $C_{rp^{(+r)}}$ denotes the receiving capacity of region $p^{(+r)}$ from region r, estimated as follows [18]:

$$C_{rp^{(+r)}} = \begin{cases} C^{\max}_{rp^{(+r)}} & \text{if} \quad 0 \leq n_{p^{+(r)}} \leq \chi.n_{p^{+(r)},jam} \\ C^{\max}_{rp^{(+r)}}\left(\frac{n_{p^{+(r)},jam} - n_{p^{+(r)}}}{(1-\chi)n_{p^{+(r)},jam}}\right) & \text{if} \quad \chi.n_{p^{+(r)},jam} \leq n_{p^{+(r)}} \leq n_{p^{+(r)},jam} \end{cases} \tag{3.9}$$

$C^{\max}_{rp^{(+r)}}$ denotes the maximum value of the receiving capacity of region $p^{(+r)}$, which is the minimum boundary capacity and a region's supplying capacity. χ is the parameter that defines the critical accumulation of the receiving capacity function, associated with the point at which capacity starts to decrease from the constant value. While [62] and [17] considered this point equivalent to the MFD critical accumulation, [84] considered this assumption to be restrictive and claimed that the turning point of the receiving capacity was higher than the critical accumulation value. The capacity that is assigned to any trip on path p is proportional to the ratio of partial accumulation to total accumulation. In the presence of a perimeter control at the boundary of the two neighboring regions, the actual flow that transfers from region r to the next region $p^{+(r)}$ is also constrained by the perimeter control rates $\alpha_{rp^{+(r)}}$. In this situation, Eq (3.5) transforms into the following equation:

$$\dot{n}^{p}_{r,Q}(t) = \frac{n^{p}_{r,T}(t)}{n_r(t)} G_r(n_r(t)) - \alpha_{rp^{+(r)}} f^{p}_{r,out}(t) \qquad \forall r \in R, p \in P \tag{3.10}$$

Total accumulation of any region reads as follows:

$$n_r(t) = \sum_{p \in P} (n^{p}_{r,T}(t) + n^{p}_{r,Q}(t))\gamma^{p}_{r} + n_{r,I}(t) \qquad \forall r \in R \tag{3.11}$$

Where the pre-identified region-route incidence matrix $\gamma^{p}_{r} = 1$, if region r exists on p, and $\gamma^{p}_{r} = 0$, otherwise; thereafter, the total queuing accumulation at the border of any region r with its succeeding neighboring regions is estimated by summing all path-aggregated queuing dynamics in region r that proceeds to region $p^{(+r)} = r' \in H_r$

as a part of the trip on path $p \in P$.

$$\dot{n}_{rr',Q}(t) = \sum_{p \in P} \left(\frac{n^p_{r,T}(t)}{n_r(t)} G_r(n_r(t)) - f^p_{r,out}(t)\right)\gamma^p_{rr'} \qquad \forall r \in R \tag{3.12}$$

The pre-known parameter $\gamma^p_{rr'} = 1$, if region r' is the region following region r on path p (*i.e.* $p^{(+r)} = r'$), and $\gamma^p_{rr'} = 0$, otherwise.

3.3.3 Travel time estimation

In a general multi-region urban network, where each region might have common borders with multiple regions, travel times on different paths passing through a same region is dissimilar due to the difference in queuing delay that occurs at the region's boundary with each of its multiple neighboring regions. In such a network, the traveling time of vehicles in region r passing through path p, τ^p_r constitutes two terms: (i) τ_r, inherent regional travel time as a function of average regional accumulation and (ii) $\tau^p_{r,Q}$, total queuing delay at the border of region r with its succeeding region $p^{(+r)}$. The type of queue considered here is a point queue (PQ) rather than a physical queue. PQ models do not occupy any physical space. Relying on this assumption, the total traveling time of vehicles in region r moving to region $p^{(+r)}$ via path p at time t is outlined as follows:

$$\tau^p_r(n_r(t)) = \tau_r(n_r(t)) + \tau^p_{r,Q}(t) = \frac{n_r(t)}{G_r(n_r(t)))} + \frac{1}{2}\frac{\sum_{p \in P} n^p_{r,Q}(t)\gamma^p_{rr'}}{T.f^p_{r,out}(t)} \qquad \forall r, r' \in R \tag{3.13}$$

The first term in Eq. (3.13) denotes the mean regional travel time estimated earlier using Eq. (3.2) and the second term denotes total queuing delay at the border of two neighboring regions r and $p^{(+r)}$ for all paths crossing these two regions in a sequence of regions from an origin region to a destination region. The second term is estimated by dividing total queuing accumulation by outflow during any simulation step time interval T.

3.4 Fair perimeter control

3.4.1 Utility function

To model the fair perimeter control problem, first I developed a utility function for each region in the network. An objective function was then constructed based on the proposed utility functions. The concept of proportional fairness was initially proposed by [80]. [81, 82] developed a proportionally fair model for the emergency evacuation problem that could address two competing objective functions simultaneously without sacrificing one for the another.

The utility function $U_i(x_i)$ expresses the benefit that user i gains as a result of achieving resource x_i. Provided that $U_i(x_i)$ is a positive, non-decreasing function of x_i, a set of solutions $A^* = \{x_1^*, \ldots, x_i^*\}$ is feasible if it does not violate the capacity constraints.

Lemma 3.1. *Given a set of utility functions* $\{U_1(x_1), U_2(x_2), \ldots, U_n(x_n)\}$ *and a set of weights* $\{w_1, w_2, \ldots, w_n\}$*, a set of feasible solutions* $\{x_1^*, \ldots, x_i^*\}$ *is a weighted proportional fair allocation of resources if it satisfies the following condition:*

$$\sum_{i=1}^{n} w_i \frac{u_i(x_i) - u_i(x_i^*)}{u_i(x_i^*)} \leq 0 \qquad \forall\ A^* = \{x_1^*, \ldots, x_i^*\} \tag{3.14}$$

Where w_i *denotes the weight of utility function* $U_i(x_i)$*.*

Proof. Please refer to [80, 81] for proof of Lemma 3.1. □

Lemma 3.2. *Eq. (3.14) can be achieved by maximizing the weighted sum of the logarithms of theutility functions, i.e.,* $\sum_i w_i \, ln \, U_i(x_i)$*.*

Proof. Please refer to [81] for proof of Lemma 3.2. □

Lemma 3.3. *For a utility function to satisfy Eq. (3.14),* $\sum_r w_r ln \, U_r(x_r)$ *must be a non-negative strictly concave function.*

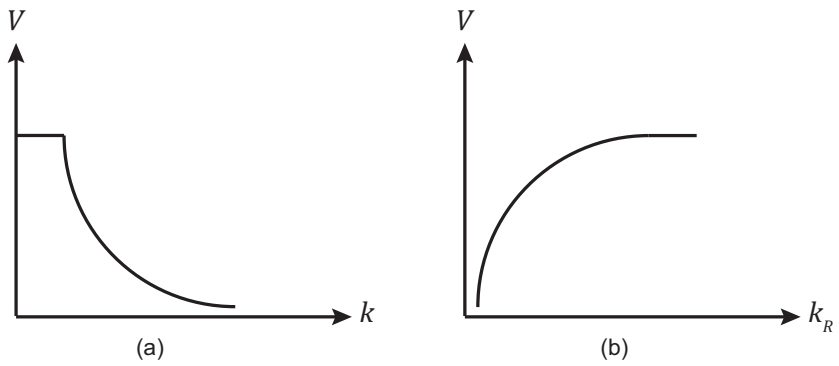


Figure 3.2: (a) speed-density plot, (b) speed-remaining density plot

For the purpose of our problem, the utility function $U_r(n_r)$ can be defined as a function of regional accumulation n_r, for any region $r \in R$. Perimeter control rates determine allocation of resources (i.e., available accumulation) by either permitting or preventing vehicles to proceed to another region.

Because of the properties required for the utility function and based on the nature of the problem, we consider regional speed as an appropriate indicator of a region's utility level. [17] proposed a speed-accumulation relationship for MFD as an exponential function as follows:

$$\bar{V}_r = v_{r,free} \cdot \exp(-\frac{1}{2}(\frac{n_r}{n_{r,crt}})^2) \qquad \forall r \in R \tag{3.15}$$

Where $v_{r,free}$ (m/sec) denotes free-flow speed, and $n_{r,crt}$ (veh) denotes critical accumulation in region r.

In Eq. (3.15), speed is not a non-decreasing function of accumulation (see Fig. 3.2a), which violates one of the conditions for the utility function illustrated earlier in this section. However, speed can be written as a function of the available (remaining) accumulation $n_{r,a} = n_{r,jam} - n_r$ (veh) of any region by substituting n_r with $n_{r,jam} - n_{r,a}$; this substitution results in the following utility function, which is a non-decreasing function of available accumulation $n_{r,a}$ (see Fig. 3.2b) as follows:

$$U_r(n_{r,a}) = v_{r,free} \cdot \exp(-\frac{1}{2}(\frac{n_{r,jam} - n_{r,a}}{n_{r,crt}})^2) \qquad \forall r \in R \tag{3.16}$$

Another issue to consider is the degrading effect of queues on regional production and traveling time and road users' utility levels as a result. Eq. (3.16) does not distinguish transferring vehicles from queuing vehicles; therefore, I introduce the parameter ξ that has values between zero and one as a coefficient for the utility function to capture the effect of queuing vehicles.

$$U_r(n_{r,a}) = \xi.v_{r,free} \cdot \exp(-\frac{1}{2}(\frac{n_{r,jam} - n_{r,a}}{n_{r,crt}})^2) \qquad \forall r \in R \tag{3.17}$$

Where ξ is a linearly decreasing function of the number of queuing vehicles within interval β and 1.

$$\xi = \begin{cases} 1 & \text{if} \quad n_{r,Q} = 0 \\ \eta & \text{if} \quad n_{r,Q} = n_{r,Q}^{max} \end{cases} \tag{3.18}$$

Parameter ξ has an effect similar to the heterogeneity coefficient introduced in [18] and captures the effect of queuing vehicles by depicting the degrading effect of queues on road users' utility functions. Based on the value of parameter η, the utility function can have a variant shape.

Proof. To prove Lemma 3.3 is satisfied for our proposed function, the second derivation of $\sum_{r \in R} ln\ U_r(n_r)$ should always be a negative number.

$$\frac{\partial^2 ln(U_i(x_i))}{\partial x_i^2} = \frac{\partial^2 ln(\xi.v_{r,free} \cdot \exp(-\frac{1}{2}(\frac{n_{r,jam} - n_{r,a}}{n_{r,crt}})^2))}{\partial n_{r,a}^2} = -1 < 0 \tag{3.19}$$

The above Equation is always satisfied; therefore, $ln(U_r(n_r))$ is a strictly concave function. Given a set of positive weight parameters $\{w_1, w_2, \ldots, w_n\}$ and concave function $ln(U_r(n_r))$, the weighted summation of concave functions is concave as well. □

3.4.2 The proportionally fair perimeter control problem

Using the utility function developed in the previous section and based on the MFD dynamic traffic states, a fair control model is proposed in the form of the following optimization problem. The control model is developed in an MPC framework and is optimized over the prediction horizon N_p with control sampling time T_c.

$$\max_{\alpha} \quad T \sum_{t=1}^{N_p} \sum_{r \in R} w_r \; ln \; U_r(n_r(t)) \tag{3.20}$$

$$(3.4), (3.6), (3.7), (3.8), (3.9), (3.10), (3.11)$$

$$0 \leq n_r(t) \leq n_{r,jam} \qquad \forall r \in R, t = 1, .., N_p \tag{3.21}$$

$$n_{r,Q} \leq n_{r,Q}^{max} \qquad \forall r \in R, t = 1, .., N_p \tag{3.22}$$

$$\alpha_{min} \leq \alpha_{rp(+r)}(t) \leq \alpha_{max} \qquad \forall r \in R, t = 1, ..., N_p \tag{3.23}$$

$$q_r^p(t) = q_{r,t}^p \qquad \forall r \in R, t = 1, ..., N_p \tag{3.24}$$

$$n_r(0) = n_{r,0} \qquad \forall r \in R \tag{3.25}$$

Eq. (3.20) is the objective function defined in terms of the weighted sum of the logarithm of a region's utility, which maximizes the total normalized utility of a network with respect to perimeter control values α_{ij}. Weight parameters w_r are included to prioritize important regions of a network by assigning higher weights. For instance, a higher weight can be assigned to downtown areas to reduce traffic and encourage more active modes. Mass conservation equations are added to estimate a network's traffic state in terms of total traveling vehicles, total queuing vehicles and total accumulation in the network at any time instant t. Eq. (3.21) ensures that a region's accumulations do not exceed the jam accumulation level and, thus, prohibits the network from gridlock condition. Eq. (3.22) constrains the number of queuing vehicles to a maximum of $n_{r,Q}^{max}$.

This constraint is important because it ensures that MFD remains well-defined and it contributes to a more equitable solution. If parameter η in Eq. (3.18) is equal to zero, this constraint can be eliminated. Eq. (3.23) constrains the perimeter control values between a lower limit α_{min} and an upper limit α_{max}. Eq.s (3.24) and (3.25) initialize demand and accumulation, respectively.

3.5 Numerical results and discussion

In this section, the performance of the proposed PFPC model is examined using a numerical example for a hypothetical urban network that consists of seven regions as shown in Figure 3.3. Region 4 is the central business district (CBD) of the city. The shortest route between any two non-neighbor regions is usually through the CBD, and thus, usually more vehicles travel in this region compared to other regions in a network. Each region in this network has a specific MFD in the form of an exponential function that relates accumulation to travel production. The MFD function for any region in this network is defined, similar to that of [17] as $P_r = n_r.v_{r,free}.exp(-\frac{1}{2}(\frac{n_r}{n_{r,crt}})^2)$ with critical accumulation $n_{r,crt} = 3.4 \cdot 10^3$ (veh), jam accumulation $n_{r,jam} = 10^4$ (veh), and free flow speed $v_{r,free} = 20$ (m/sec). Regions' receiving capacity is included by parameters $C_{ik^{(+i)}} = 3.5$ (veh/sec) and $\chi = 0.64$. For MFDs to be well-defined, a maximum queue length $n_{Q_{max}} = 600$ (veh) is considered for any region in the network. Weight parameters $w = 1$ are considered for all regions, and parameter $\beta = 0.5$ is assumed. Accumulation states are assumed to have measurement noises as they are obtained from fixed and mobile sensors. The error terms are described by normal distributions that have a zero mean and a standard deviation equal to 0.4, i.e., $\varepsilon(n) \sim \mathcal{N}(0, 0.4)$. Errors terms also describe demand uncertainty, which has a normal distribution that has a zero mean and a standard deviation equal to 0.4. , i.e. $\varepsilon(d) \sim \mathcal{N}(0, 0.4)$.

All regions are empty at the start of the simulation. Regional trip length is assumed to be constant. The lower control bound is $u_{min} = 0.1$, and upper control bound is

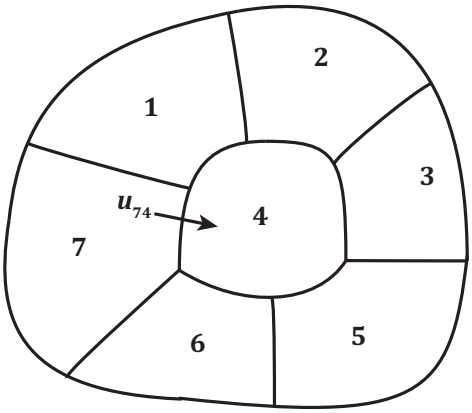


Figure 3.3: Urban network divided into 7 regions

$u_{max} = 1$. Path enumeration is avoided by considering the 3 shortest routes, found using Dijkstra's algorithm, between any OD region pair in the network at the start of the simulation (i.e., we assume a maximum of 3 paths between any OD pair). The network is simulated and controlled for 3 hours using a control sampling time of $T_c = 300$ sec. For the MPC model, a prediction horizon $N_p = 7$, and a control horizon $N_c = 2$ are considered based on controller tunning and the computational efficiency results of [67] for a similar 7-region urban network. The plant model for the MPC model has access to the real values of demand and accumulation, while the control model has only access to the average values. The MPC control problem is a non-convex, non-linear programming problem, and its solution is obtained using the interior-point algorithm in MATLAB (fmincon function). All scenarios are simulated using MATLAB R2016a (9.0.0), on a 64-bit Windows PC with 4-GHZ Intel Core-i7 and 24-GB of RAM.

In the analysis, I investigate the effectiveness of the proposed fair control (FC) scheme in improving fairness measures while maintaining an acceptable level of efficiency. The performance of the model is compared with that of no control (NC) and basic control (BC) scenarios. In the NC scenario, the flow that transfers between neighbor regions is not controlled. Similar to FC, BC is developed in an MPC framework and proactively optimizes the flow that is allowed to transfer between neighboring regions (i.e., CBD and peripheral regions and peripheral regions with each other) over the

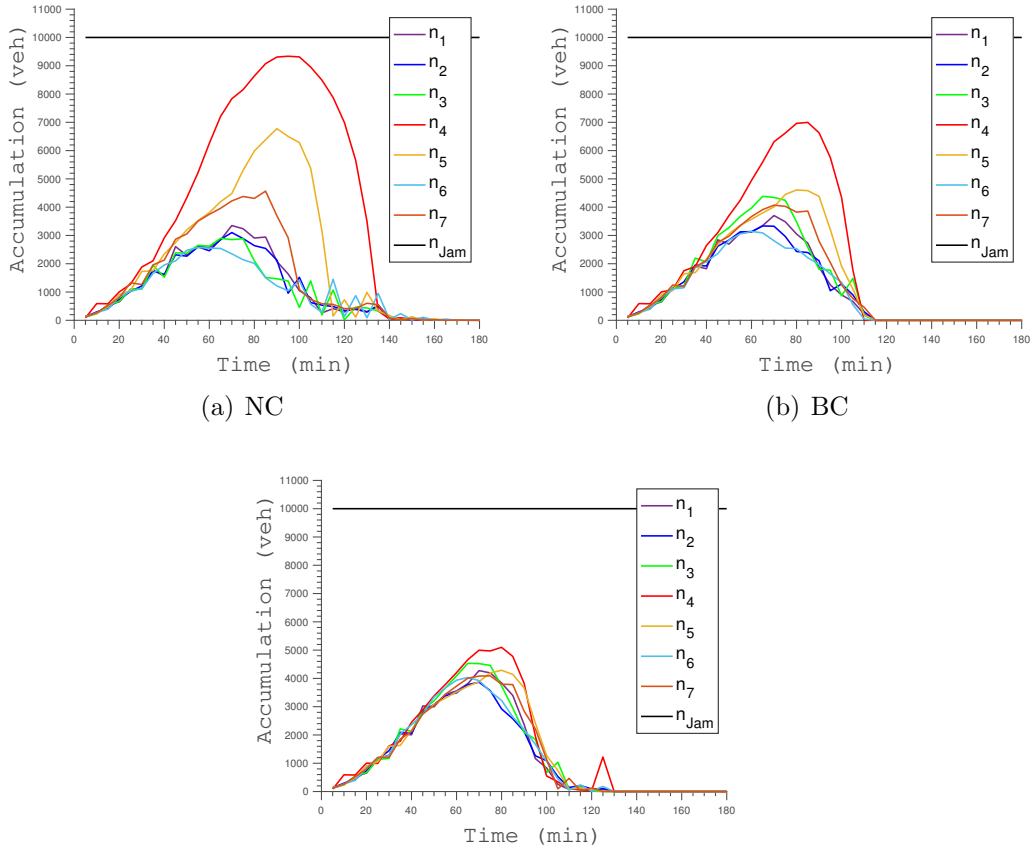
prediction horizon. The objective function is to minimize the total time spent (TTS) of the network, which constitutes traveling time and waiting time in queues. Sections 5.1 and 5.2 respectively evaluate efficiency and fairness for different control scenarios.

3.5.1 Evaluation of efficiency

In the context of MFD-based control, efficiency is measured in terms of TTS of a network. Efficient control strategies manipulate regional accumulation in an attempt to maintain accumulation around a desired level [15, 17, 18, 16, 30]. Efficiency is achieved by minimizing TTS, which, according to [29], is equivalent to the total accumulation of a network.

Figure 3.4 shows regional accumulations under the above-mentioned control scenarios. In NC, CBD operates at high accumulation levels from $t = 50$ min to $t = 140$ min and has a maximum accumulation that reaches 9200 veh (i.e., 92% of a region's jam accumulation value). Further, regions 4 and 7 operate under congested conditions; however, the remaining regions experience low accumulation during the simulation period. In BC, CBD experiences lower accumulation at any time instant during the control period, while maximum accumulation reaches 7000 veh (70% of the jam accumulation value). Regions 4, 5, and 7 experience lower accumulation, but regions 1, 2, and 6 experience higher accumulation compared to NC. In FC, accumulation in regions 3 and 4 is reduced, but accumulation in the other regions increases. Accumulation of regions in FC are maintained fairly close to each other during the entire simulation period; CBD's maximum accumulation never exceeds 4800 veh. In NC, regions are discharged at $t = 160$ min, while in BC and FC, regions are discharged at $t = 120$ min and $t = 130$ min, respectively. Both BC and FC are similarly efficient in improving traffic conditions compared to NC.

The hypothetical network's simulation results under different control scenarios during a 3-hour control period are summarized in 3.1. These values are averages over 10 simulation runs. Compared to NC, BC improves the efficiency measure (i.e., TTS) by



(a) NC (b) BC

(c) FC

Figure 3.4: Regional accumulations in (a) no control (NC), (b) basic control (BC), and (c) fair control (FC)

21.11% and FC by 19.12%. According to these results, the efficiency of BC and FC are not very different, and therefore, both BC and FC are successful in improving efficiency. In addition, BC and FC cause the multi-region urban network to have more homogeneous traffic conditions by producing closer values for a region's accumulation. This is best represented by summing the varying accumulation values in Table 3.1, which indicates a 56.49% reduction using BC compared to 78.94% reduction using FC. Consequently, FC is more successful, and this result can be interpreted as an enhancement in fairness.

Table 3.1: Simulation results under three control scenarios

Traffic measures		No control (NC)	Basic control (BC)	Fair control (FC)
Total travel time		$2.1348 \cdot 10^6$	$1.6842 \cdot 10^6$	$1.7266 \cdot 10^6$
		(−)	(−21.11%)	(−19.12%)
Accumulation	Avg.	$1.3366 \cdot 10^3$	$1.3703 \cdot 10^3$	$1.6943 \cdot 10^3$
		(−)	(−21.11%)	(−19.12%)
	$\sum$ std.	$9.9750 \cdot 10^3$	$4.3400 \cdot 10^3$	$2.1006 \cdot 10^3$
		(−)	(−56.49%)	(−78.94%)
Speed	Avg.	16.6298	17.1847	17.2305
		(−)	(3.34%)	(3.61%)
	$\sum$ std.	25.3487	13.0408	6.6566
		(−)	(−48.55%)	(−73.73%)
Queue	Avg.	0	68.3583	39.7088
		(−)	(−)	(−41.91%)
	$\sum$ std.	0	669.8584	470.4276
		(−)	(−)	(−29.77%)

3.5.2 Evaluation of fairness

Fairness is measured in terms of regional speed and queues. Based on speed, some level of efficiency in all regions within a network is ensured. Also, waiting in a queue is an unpleasant experience for travelers; therefore, a fair control strategy should properly distribute queues among regions while ensuring a minimum average speed in all regions.

The evolution of average speed and queue is depicted in Figure 3.5. Under NC, the average speed in CBD decreases to less than 5 m/sec for 60 minutes from t=60 min to t=120 min and the speed is almost zero from t=80 min to t=100 min, which means that CBD is gridlocked for 10 minutes. However, in NC, regions 1, 2 ,3 and 6 experience high speeds and have an average speed that does not drop below 13 m/sec. Therefore, the average speed in different regions is dramatically different in NC. In BC, variations in speed decrease, and gridlock does not occur. The average regional speed drops to less than 5 m/sec for 25 minutes from t=70 min to t=95 min. In contrast, in FC, the minimum speed in CBD does not fall below 8 m/sec, and the average speed drops to less than 10 m/sec for 25 minutes from t=60 min to t=85 min. In addition, the average

speed of each region is fairly close during the simulation period.

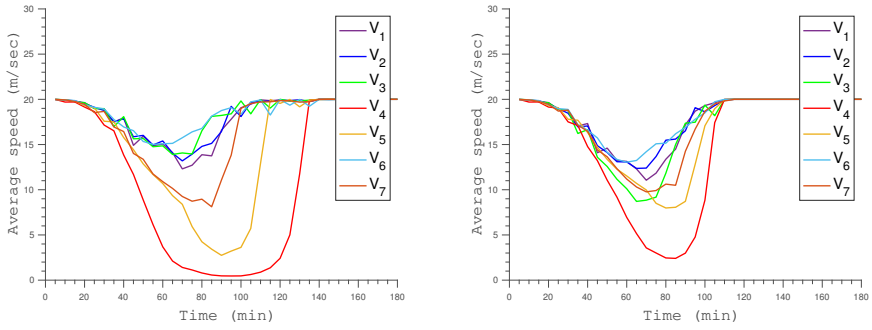
Table 3.1 shows the average and sum of the standard deviations of queues and speed in the multi-region urban network. Compared to NC, the average regional speed increases 3.34% under BC and 3.61% under FC. Further, compared to NC, the sum of the standard deviations of speeds decreases 48.55% under BC and 73.74% under FC.

Figure 3.5 (d, e) shows queuing dynamics in the control scenarios. In NC, queues do not exist in any region, but in BC, an average of 68.3583 vehicles are in queues at any time instant. In contrast, in FC, the average queue decreases by 41.91%, and the sum of the standard deviations decreases by 29.77%. Compared to BC, FC suppresses the number and variations of queuing vehicles.

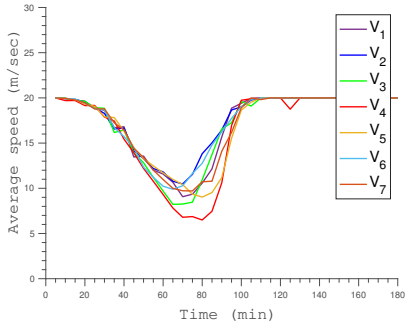
The average and standard deviation values of accumulation, speed, and queues recorded in Table 3.1 confirm the accumulation plots in Figure 3.4. The results indicate that FC improves fairness in an urban network by increasing homogeneity while maintaining an acceptable level of efficiency. The results suggest significant improvement in fairness measures in FC, while efficiency is not significantly affected. BC is also an improvement over NC; however, this control strategy creates long queues near the perimeters of regions by implementing strict control. Heterogeneity in the spatial distribution of vehicle density resulting from queues degrades trip production in a traffic network. In the model, this issue is resolved by introducing parameter ξ as a heterogeneity coefficient; therefore, homogeneity increases in our developed scheme rather than the conventional BC.

3.6 Conclusion

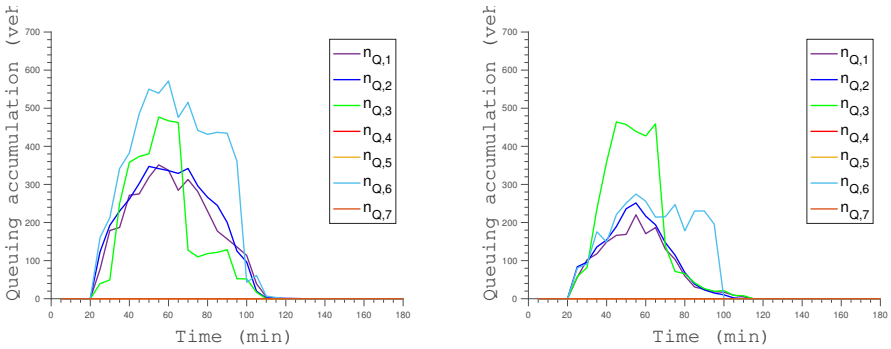
In this chapter, a proportionally fair perimeter control scheme based on a balanced distribution of regional average speed for a network-wide control of an urban network is developed. For state prediction, MFD dynamic state equations are extended to include queuing dynamics using a PQ model. A heterogeneity coefficient based on



(a) NC (b) BC



(c) FC



(d) BC (e) FC

Figure 3.5: Regional average speed in (a) no control (NC), (b) basic control (BC), (c) fair control (FC); regional queues in (d) basic control (BC) and (e) Fair control (FC)

queues is introduced to the utility function to integrate the degrading effect of queues on road users' utility. Consideration of queues in dynamic state equations is vital when optimizing using perimeter control; otherwise, the results can be misleading. Perimeter control, if not implemented properly, creates long queues at regional perimeters that can lead to deteriorating traffic conditions rather than the intended amelioration.

The advantage of the developed scheme is that it improves fairness without sacrificing efficiency. Performance, in terms of efficiency and fairness, of the developed model is compared with no control and basic control scenarios. Efficiency is defined in terms of total time spent of a network, and fairness is defined based on a balanced distribution of speed. BC slightly outperforms FC in terms of efficiency, but in terms of fairness, FC consistently outperforms BC. In addition, the fair perimeter control approach creates homogeneous traffic conditions in a network.

The type of queue considered here is a point queue (PQ) rather than a physical queue. Congestion propagation and shockwave phenomena cannot be described using a PQ model. The kinematic wave theory and its discretized counterpart, the cell transmission model (CTM), can be integrated into an MFD framework to estimate more precisely travel time and track congestion propagation within a region; thereafter, fair perimeter control schemes can be developed. For future research, the concept of fairness can be adopted in developing other control schemes, such as route guidance. This advisory control scheme heavily relies on road users' compliance to the routing instructions; a fair perimeter control strategy that considers road users' utility is expected to increase compliance to the routing instructions and improve a network's traffic conditions.

Chapter 4

A macroscopic dynamic network loading model using variational theory in a connected and autonomous vehicle environment

In this chapter, a multi-reservoir dynamic network loading (MRDNL) model for a large-scale urban road network is developed. The proposed framework consists of a hyper-link and a hyper-node model. The hyper-link model is constructed using variational theory (VT) by introducing a set of constraints that construct an upper bound for upstream and downstream cumulative accumulations. Using VT, LWR model is incorporated into MFD dynamics to enhance traffic propagation at reservoirs. The hyper-link model captures important traffic phenomena, such as kinematic waves, queueing, and congestion, while the hyper-node model functions as a medium for transferring flow to other parts of the multi-reservoir urban network. A numerical scheme is advanced for solving the proposed MRDNL model and its performance is evaluated using a hypothetical traffic network. Compared to previous MFD-based DNL models, our approach is more computationally efficient since it does not require discretization and offers more realistic solutions by considering multiple kinematic waves and including the dynamics of urban traffic signals. This study, further, incorporated connected and autonomous vehicles (CAVs) into MFD dynamics and evaluated the network-wide effect of introduction of CAVs in urban traffic networks. For a network with 100% CAV market penetration rate, an enhancement of 33% in network outflow is found.

4.1 Introduction

The dynamic traffic assignment (DTA) and dynamic network loading (DNL) problems are the main basis of fundamental problems in the traffic research field and have major applications in network modelling, policy evaluation, and traffic operation and management. Rigorous modelling of DNL models has major implication on the functionality of traffic strategies constructed using microscopic, mesoscopic, or macroscopic traffic models. Microscopic representations of DNL models for large-scale networks are scarce in the literature due to the computational burden and tractability issues. These models also require a significant amount of data, which is another reason for their limited application in realistic traffic networks, where the main sources of information are Eulerian (e.g., fixed traffic sensor) and/or Lagrangian (e.g., mobile sensors such as GPS-equipped taxies) data such as mean flow, occupancy, travel data, space mean speed, density, and trajectory. Conversely, macroscopic models use aggregated traffic data, which makes them efficient tools for real-time modelling and controlling traffic networks, especially for urban networks that have specific structures, a high level of detail and complexity, road users displaying stochastic behaviour, and traffic signals. Aggregated traffic models can be used to their fullest if important traffic phenomena, such as traffic shockwave, queue buildup, dispersion, and spillback can be appropriately modelled.

[4] introduced macroscopic fundamental diagram (MFD) as a tool to monitor and control urban networks using a single mean traffic state variable. [6] derived analytical theories for MFD, and [1] empirically verified the theories using a combination of fixed detector and GPS data from downtown Yokohama. MFD establishes a reproducible relationship between network average density and flow in a roughly homogeneous network with slow-varying demand. Homogeneity is necessary for a well-defined, low-scatter MFD, especially in congested traffic conditions. Heterogeneous networks can be handled by decomposing the traffic network into several homogeneous reservoirs [2]. The simplicity of MFD and its low computational requirement make it an elegant

and practical tool to control and improve the efficiency of large-scale urban networks. Recently, a few studies have made important contributions to enhancing MFD dynamics by producing more realistic models that consider important traffic phenomena, such as (i) heterogeneity and hysteresis effect [11, 35, 85, 18], (ii) state delay [69, 26], (iii) queue dynamics [86, 87], and (iv) wave propagation [36, 88].

To estimate MFD analytically, [6] employed variational theory (VT) [7, 8] for an arterial comprising of a set of consecutive links and signalized intersections. VT rephrased the MFD estimation problem as a shortest path problem and used a moving observer method to estimate the cost associated with different paths, based on which, a set of practical cuts forming an upper bound for MFD was defined. [9] modified and extended the original VT method [6] by considering variabilities in network topology, signal settings, and turning movements. [10] advanced an analytical method to derive tight estimations of MFD for networks of both regular and irregular topology and signal settings. They introduced sufficient variational graph (SVG), which gathered all paths required to estimate MFD, and claimed that SVG was sufficient to provide a tight representation of MFD. They also introduced a method of cuts that required less computational effort compared to that of the SVG method. The set of cuts provided a tight estimation for regular networks while producing only an upper bound for irregular networks.

Recent research has been on the MFD-based modelling of DNL and DTA problems for large-scale traffic networks. [53] developed an approximated DTA that had implications for variant trip length inherent in network heterogeneity. They proved that MFD, using aggregated traffic information that has a reasonable computational burden, was capable of representing system dynamics nearly as well as detailed microscopic models. [24] advanced a user equilibrium (UE) and a system optimum (SO) model that consider the route choice effect on MFD in the case of heterogeneous urban networks. [71] developed a stochastic user equilibrium route choice model at the regional level using a logit route choice model in an anticipatory control framework to capture road users' routing

responses to the perimeter control instructions. [15] and [25] developed DTAs for urban networks that have freeways as alternative routes. MFD-based route guidance strategies were devised as advisory control schemes and were shown to be efficient at controlling urban networks. In these models, where needed, a lower level of control was added to control each reservoir locally [17, 55, 67]. Capacity constraint and queueing were also discussed in some MFD-based UE models in the literature. [87] combined Vickrey's model of bottlenecks with MFD to describe boundary queueing dynamics, and thus, formulated the UE model as a PDE for the morning commute problem. [26] developed a whole link dynamic user equilibrium (DUE) model that had simultaneous route and departure time choice, where travel time was included as an endogenous time-varying function. In addition, boundary capacity constraint was considered in their model; however, it was postulated that in the case of limited boundary capacity, vehicles rerouted instead of waiting in queue. Similarly, in [89], queueing dynamics was not carried to the next time step, assuming that when vehicles cannot proceed to the succeeding reservoir due to constrained receiving capacity, they go back to their origin and start their trip from scratch.

DNL and DTA models discussed in the MFD literature are conventionally modelled using accumulation-based models that do not appropriately represent traffic dynamics under fast-varying demand conditions. The models ignore inner-reservoir travel time and assume that the flow entering a reservoir from an upstream perimeter boundary propagates instantaneously downstream at infinite speed. [23] recognized the issue and claimed that, similar to link-based traffic models, in reservoir-based traffic models, propagation speed should not be faster than free-flow speed; otherwise, immediate reaction of outflow to demand surge leads to inconsistent wave propagation within a reservoir. Further, the time-varying nature of travel demand affected the shape and scatter of MFD [35]. The effect was sudden and was magnified in accumulation-based models in the absence of reaction time; the effect was also the cause of the numerical viscosity observed in the accumulation graphs as highlighted by [36]. To circumvent the issue and

improve representation of travel time evolution, especially in the case of fast-varying demand, inner-reservoir travel time was incorporated into MFD dynamics as a time delay in control input [16, 34] or into MFD output function [69, 26]. Another solution to the problem is the seminal kinematic wave (KW) model [90, 91] that describes the spatial-temporal evolution of vehicle density and propagation based on the fundamental diagram and conservation rule that was incorporated into the MFD model for a single reservoir [36, 92], and extended to a multi-class traffic dynamic model [93].

To estimate the KW solution, several approaches are developed in the literature: first-order numerical schemes such as time-space discretized cell transmission models (CTM) [94, 95, 96], approximated dynamic programming [7], and exact solution scheme [97], which models the problem as a Hamilton Jacobi partial differential equation (HJ-PDE) and solves it using the Lax-Hopf formula. Other works defined the problem using an optimization framework and maximized the value of sending flow and receiving capacity at downstream and upstream perimeter boundaries, respectively [98]. In the MFD literature, [62] divided each reservoir into sub-networks (i.e., cells) and solved the reservoir routing problem using a CTM-like approach [95]; this approach used a proper boundary constraint to indicate propagation of flow. [36] and [92] captured traffic waves within a single reservoir that contained multiple routes using a CTM-like model and solved it numerically using a first order Gudonov scheme. The proposed CTM-like approaches require intense discretization of space, which make them computationally difficult. In these studies, a simplified triangular MFD with only two speeds (free flow and backward wave speed) was used. None of the published studies, however, completely account for the effect of variability in wave formation of a more general concave MFD. As demonstrated in [99], travel time at critical points in a link estimated using a concave FD is potentially 50% higher than what is predicted using a linear hypocritical branch; for a network comprised of a set of links, miscalculation of travel time is non-negligible.

Introducing connected and autonomous vehicles (CAVs) into a traffic network can significantly affect the traffic conditions. CAVs behave differently compared to conven-

tional vehicles; for example, they have quicker reaction times compared to conventional vehicles, and this characteristic allows them to maintain a shorter distance from the vehicle ahead. Shorter vehicle-to-vehicle spacing means more vehicles can be accommodated in traffic networks, which enhances capacity. CAVs also reduce travel times, fuel consumption, and parking facility requirements. An important issue to study is the effect of CAVs on network traffic conditions, especially in large-scale urban networks that struggle with congestion. In [100], the effect of autonomous vehicles on MFD is investigated using microsimulation. More research needs to be conducted to further incorporate CAVs as a part of MFD modelling.

A proper description of wave propagation is necessary to develop a well-posed traffic model and accurate estimations of traffic states. DNL and DTA models that are constructed based on proper reservoir models are capable of producing a more elegant and reliable simulation of vehicular movement within a traffic network. The motivation of current study is to propose such DNL model appropriate for network-wide simulation of vehicle movement within traffic network with high precision and reasonable computational effort. While there are many MFD-based DTA and DNL models in the literature, very few of them account for the existence of miscellaneous waves that describe important traffic phenomena, such as shockwave, queue buildup, and dispersion within a reservoir. Existing studies in the MFD literature use algorithms that require intense discretization of time and space. In this chapter, a macroscopic DNL model for a multi-reservoir urban network consisting of a hyper-link (i.e., consecutive links with an equal number of lanes separated by signalized intersections) and a hyper-node (i.e., an intersection of hyper-links at reservoir boundaries) is proposed. The hyper-link is modelled by constructing a set of constraints based on VT and is rephrased as finding the maximum value of sending flow and receiving capacity of a reservoir at the upstream and downstream ends. The hyper-node model operates as an interface for transferring flow to the other parts of a multi-reservoir urban network. Using the LWR model in an MFD context offers precise representation of inner-reservoir wave propagation and

travel time.

The theoretical contribution of this chapter is that our model captures proper wave propagation and estimates rigorous inner reservoir travel time for an urban arterial with piecewise linear MFD. Further, the hyper-node model, proposed to estimate the transferring flow between neighbouring reservoirs, amends the overestimation of receiving capacity that occurs in earlier research in the MFD literature. In terms of a methodological contribution, this chapter uses VT to solve the KWT solution of the hyper-link model. Further, a computationally efficient solution for a network-wide DNL that does not require space-discretization is proposed. An important issue that should be studied is the impact of CAVs on network traffic conditions and the evolution of flow, especially during transition period from conventional vehicles to CAVs. As mentioned by [101], there is a lack of research that investigates the regional network-wide effects of CAVs. The current chapter fills this gap by dynamically loading traffic networks under various CAV penetration rates.

4.2 Methodology

This chapter investigates the flow propagation problem for a multi-reservoir urban network using a dynamic network loading (DNL) and a route choice model. Similar to [92], this chapter assumes that the traffic dynamics of a large-scale urban network, divided into several reservoirs, is captured using the interface between a link model and a node model. Thus, our proposed DNL is composed of a hyper-link model, which maps the propagation of flow in a single reservoir, and a hyper-node model, which determines the transition of flow between reservoirs in the multi-reservoir urban network.

Figure 4.1 provides a schematic view of the multi-reservoir DTA that exhibits components of the proposed model. The hyper-link model is included to compute the sending flow and receiving capacity of reservoirs based on the piecewise linear MFD that is constructed using VT and according to the percentage of CAVs in the network as CAVs

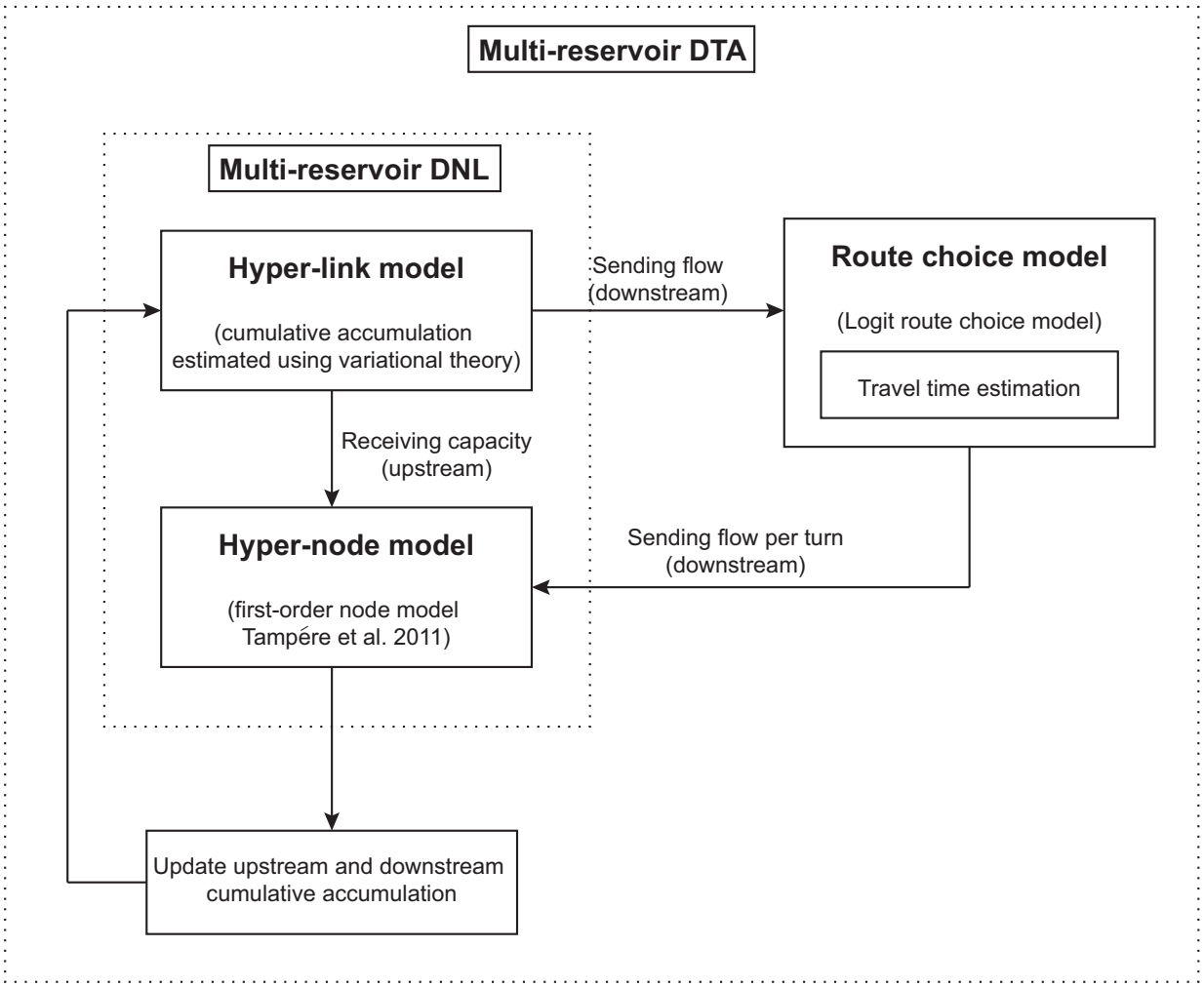


Figure 4.1: Schematic view of the DTA model consisting of route choice and DNL models

affect the shape of MFD and its maximum flow capacity. The hyper-node model distributes the reservoirs' receiving capacity over all demanding reservoirs. It is based on the first-order node model of [102] and is incorporated to amend the over-estimation of reservoir capacity observed in earlier research. Sending flow values computed using the hyper-link model are transferred to the route choice model to distribute the reservoir total demand over its succeeding neighbouring reservoirs. The logit route choice model, which considers road users' knowledge of travel time, simulates road users' routing behaviour and models the heterogeneity of users in a mixed traffic network consisting of conventional vehicles and CAVs. The travel time estimation component embedded in the route choice model computes travel time based on current traffic conditions of a network in terms of upstream and downstream reservoir cumulative accumulation rather than the mean travel time as assumed in earlier studies. As demonstrated in [103], MFD mean travel time properly reproduces the low frequency variations of travel times due to changes in traffic conditions while losing high-frequency variations due to offsets between signals for periodic arterials. This approach fails to approximate travel time for non-periodic arterials. The travel time computed in this chapter is instantaneous travel time, which suppress the problem with MFD mean travel time.

4.3 MFD-based modelling

4.3.1 Accumulation-based models

Consider a homogeneous urban network as a single reservoir comprising of m vehicle classes. Traffic dynamics, in terms of the accumulation of each vehicle class n^m as described by a simple conservation law, is as follows [4, 1, 16, 36]:

$$\frac{dn^m(t)}{d_t} = f_{in}^m(t) - f_{out}^m(t) + \psi^m(t) \quad \Leftrightarrow \quad n^m(t) = N_{in}^m(t) - N_{out}^m(t) + \Psi^m(t) \tag{4.1}$$

Where $f_{in}^m(t)$ (veh/sec), $f_{out}^m(t)$ (veh/sec), $N_{in}^m(t)$ (veh), and $N_{out}^m(t)$ (veh) denote

inflow, outflow, cumulative number of arriving and departing vehicles of class m in reservoir at time t, respectively, and ψ (veh/sec) and Ψ (veh) denote internal flow and accumulation of reservoir at time t, respectively. Reservoir outflow is estimated using the MFD function: $f_{out}(t) = G(n(t))$, which provides a well-defined relationship between accumulation n (veh) and outflow G (veh/sec) or production P (veh.m/sec) of a reservoir. The multi-class traffic dynamic model assumes that production of a network is the sum of the production of all vehicle classes.

Eq. (4.1), referred to as accumulation-based MFD, assumes a constant value for the ratio of reservoir production P (veh.m/sec) to outflow G (veh/sec): $\bar{L} = \frac{P(n)}{G(n)}$, where $\bar{L}$ (m) corresponds to average trip length of reservoir. While assuming a constant average trip length is desirable from a modelling and computational perspective, in reality, the average trip length varies as traffic congestion propagates throughout an urban network and influences travelers' route choice due to changes in internal traffic conditions or OD demand [36]. Variable average trip length was addressed in later studies [53]. The accumulation-based model also ignores inner-reservoir travel time and assumes that the flow entering a reservoir from an upstream perimeter boundary instantly reaches the opposing downstream end with infinite speed. [23] recognized the issue and stated that similar to link-based traffic models, in reservoir-based traffic models, propagation speed should not be faster than free-flow speed; otherwise, immediate reaction of outflow to demand surge leads to inconsistent wave propagation within the reservoir. To resolve the issue, inner-reservoir travel time is incorporated as delay into the control input [34, 16] or traffic state dynamics, i.e., $f_{out}(t) = G(n(t - \tau(t)))$ [69] or $N_{in}(t) = N_{out}(t + \tau(t))$. N_{in} is the cumulative inflow accumulation and N_{out} is the cumulative outflow accumulation. While including delay dilutes the viscosity issue, proper wave propagation is not captured in these models because travel time τ is assumed constant.

To describe the spatial-temporal evolution of vehicle density and capture proper wave propagation between two boundaries, the seminal kinematic wave (KW) model [90, 91], which explicitly considers internal travel time and traffic waves based on the

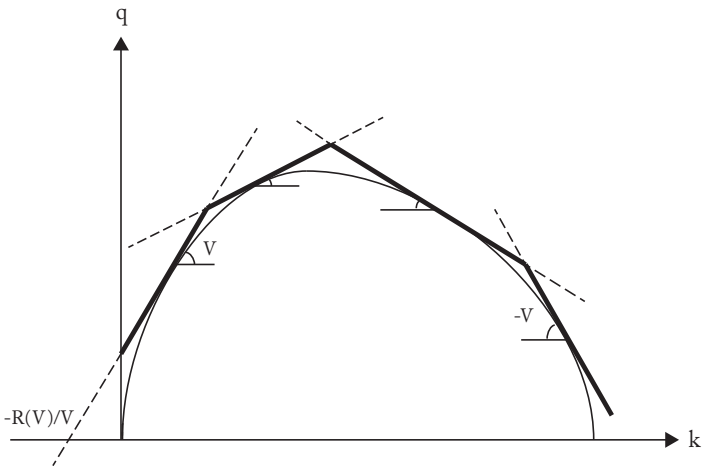


Figure 4.2: MFD defined by cuts $Q = min_V(kV + R(V))$

fundamental diagram and conservation rule, is utilized for a reservoir with a single traffic mode [36, 92] and later extended for a multi-class traffic network as follows [93]:

$$\frac{\partial k^m(x,t)}{\partial t} + \frac{\partial q^m(x,t)}{\partial x} = \psi^m(x,t) \qquad \Leftrightarrow \qquad \frac{\partial n^m(x,t)}{\partial t} + \frac{\partial p^m(x,t)}{\partial x} = \Psi^m(x,t) \quad (4.2)$$

In the above Eq., k (veh/m), q (veh/sec), n (veh), and p (veh.m/sec) denote density, flow, accumulation, and production of the reservoir, respectively, and x (m) and t (sec) denote distance and time traveled within the reservoir, respectively. The PDE-LWR model of Eq. (4.2) for a homogeneous urban network can be solved using VT. [6] analytically approximated MFD for a reservoir using the VT concept introduced by [7, 8] that models the problem in the form of a least cost path problem. Therefore, the traffic state at any instant can be simply computed using VT and is based on a set of constraints that creates an upper bound for cumulative accumulation functions.

4.3.2 Modelling based on variational theory

Consider a single reservoir in the form of an urban arterial of length L with ρ lanes and $j = 1, 2, ..., \nu$ successive links of length l_j divided by $\nu - 1$ signal-controlled intersections with signal setting times: red r_j, green g_j, offset o_j, and cycle $c_j = r_j + g_j$. Similar to

[4] and [6], we assume that the reservoir is in the form of a hyper-link. The dynamics of each link are described by a triangular FD with parameters: free-flow speed γ, wave speed ω, and jam density k_{jam}. Assuming that for each link l_j, q_j denotes flow and k_j denotes density, $q = \frac{\sum_j l_j q_j}{L}$ yields mean flow and $k = \frac{\sum_j l_j k_j}{L}$ yields mean density of the whole arterial. The maximum capacity of an arterial is $q_c = \frac{\rho\omega\gamma k_{jam}}{\omega+\gamma}$, during the green interval of a traffic light. For the sake of simplicity, lane changing behaviour and turning movements at intersections are not considered. For such an arterial, MFD can be approximated as a lower envelope of a set of lines (cuts) with slope V, y-intercept $R(V)$, and x-intercept $-\frac{R(V)}{V}$, as graphically depicted in Figure 4.2 and mathematically expressed in Eq. (4.3) in terms of estimated outflow Q [6]:

$$Q = inf_V[kV + R(V)]; \qquad V \in [-\omega, \gamma] \tag{4.3}$$

Where speed should be between backward speed $-\omega$ and free-flow speed γ. By mathematically manipulating the above equation, MFD in terms of estimated density K can be formulated as follows:

$$K = sup_V[\frac{Q}{V} - \frac{R(V)}{V}]; \qquad V \in [-\omega, \gamma] \tag{4.4}$$

Similarly, by manipulating Eq. (4.3), the cost function $R(V)$ is as follows:

$$R(V) = sup_V[Q - kV]; \qquad V \in [-\omega, \gamma] \tag{4.5}$$

The relative capacity or cost function $R(V)$ is the core of VT and is estimated based on the least cost problem. It uses the moving observer method and identifies the maximum rate at which an observer moving at speed V can be overtaken by other vehicles in the network. Let R_P denote the mean cost along path P. For all paths with mean speed V connecting two identical points in a time-space graph, $R(V)$ denotes the

minimum cost [6]:

$$R(V) = min_p(R_p|V_p = V) \tag{4.6}$$

Where R_p constitutes a set of paths p that have an identical mean speed V_p and are estimated as the average cost of all the moves in the corresponding time-space diagram over time, i.e., $R_p = \frac{1}{\mathcal{T}_p} \int r(v) d_v$, where $r(v)$ denotes the cost of the link as a function of an observer's speed, and $\mathcal{T}_p$ denotes travel time on route p. The cost function $r(v)$ linearly reduces from the maximum cost $r(-\omega) = \omega k_j$ to the minimum value $r(\gamma) = 0$, while $r(0) = q_c$ during a green signal light and $r(0) = 0$ during a red signal light. It is proved that for any network with a linear cost function, an optimum path always exists [8]; however, evaluating the cost function for all speed values $v \in [-\omega, \gamma]$ is cumbersome. [6] proposed using three families of practical cuts with speeds of $0, \gamma$, and $-\omega$ to find an upper bound for MFD. While the practical cut method of [6] provided only an upper bound for MFD, [10] provided a tight estimation of MFD for networks of both regular and irregular signal settings and topology by integrating forward and backward speeds to construct a sufficient variational graph (SVG) for a periodic infinite arterial only using the three above-mentioned speeds. They reduced the number of cuts by discretizing the mean speed values by only considering the paths from the end of the red signal on the upstream and downstream time-space diagram. Further, [10] suggested pre-scanning the cuts to instantly eliminate non-tight cuts (please refer to [10] for a comprehensive explanation of the analytical MFD estimation methods). Lastly, the estimated MFD in the above methods is piecewise linear due to the finite number of cuts used to estimate MFD as depicted in Figure 4.2.

4.4 Multi-reservoir dynamic network loading

4.4.1 Hyper-link model

To describe the traffic dynamics of a reservoir in the form of a hyper-link, I adopt the VT approach. The traffic state is estimated in terms of cumulative accumulation, and VT is used to construct a set of constraints that establish an upper bound for cumulative accumulation values at the upstream and downstream boundary ends of the hyper-link. Since the approach is the same for forward and backward waves, I restrict my exposition to forward waves and later present the solution algorithm for both waves.

Let $N(x,t)$ (veh) be the Moskowitz function [104] that indicates the cumulative number of vehicles passing point x at time t. Let q_{in} (veh/sec) be the sending flow to the reservoir. Assuming that the reservoir is initially empty at the start of the simulation, the upstream cumulative accumulation at any time, $N(x_0,t)$, can be simply computed by summing all sending flows to the reservoir as follows:

$$N(x_0,t) = \int_0^t q_{in}(\tau)d_\tau \tag{4.7}$$

Assume that a reservoir's dynamics evolve over time intervals of length $\Delta t = \frac{T}{M}$, $I = \{i : 1,2,...,M\}$, where T denotes total simulation time, and Δt, M, and I respectively denote length, number, and a finite set of time intervals. Increasing the number of intervals M enhances the accuracy of the results, but it increases computational effort. Therefore, it is important to choose the number of time intervals that strikes a balance between accuracy and effort. Due to time discretization, a piece-wise constant function for demand is considered, and cumulative accumulation curves are acquired as piece-wise affine functions, accordingly.

For a reservoir with a concave MFD, waves of various speeds from upstream might arrive downstream at different times. To properly estimate a reservoir's traffic state, the wave that dominates at any time instant must be determined by capturing the

effect of boundary point constrains on vehicle accumulation. As shown in Figure 4.3, the solution space for a forward wave is between t_1 associated with the minimum and t_2 associated with the maximum forward speed. As previously discussed, for any speed V, there exists a set of paths that have the equivalent mean speed. The path that has the least cost demonstrates the relative cost function $R(V)$ used to construct MFD, as depicted in the time-space and signal timing diagram (Figure 4.3).

Cumulative accumulation values at the two boundary ends of the reservoir are computed using the values of cumulative curves determined in previous time steps. According to fact 3 of [8], only a straight path from the boundary to the solution point has to be considered, which reduces the number of paths that have to be evaluated; therefore, only a set of paths associated with waves from the solution point to the opposing boundary end are considered for evaluation.

Consider time interval $[t_1, t_2]$ as shown in Figure 4.3, where t_1 and t_2 respectively denote the time that the slowest and the fastest kinematic waves reach downstream. For an arterial of length $L = x_L - x_0$, the time that the forward waves reach downstream end is constrained as follows [98]:

$$t_1 + \frac{L}{min(V)} \leq t + \Delta t \leq t_2 + \frac{L}{max(V)} \tag{4.8}$$

According to [7], a change in accumulation between any two points in the time-space diagram is estimated as a summation of changes in the accumulation rate along the connecting path; hence, a change in accumulation from (x_0, t_i) to $(x_L, t + \Delta t)$ is estimated as follows (substituting $q - kV$ in Eq. (4.3)):

$$\begin{aligned} N(x_L, t+\Delta t) - N(x_0, t_i) &= \int_{t_i}^{t+\Delta t} (Q(k) - kx')d_t = \int_{t_i}^{t+\Delta t} (q - kV)d_t = (t + \Delta t - t_i)R(V) \\ &= (t + \Delta t - t_i)R(\frac{x_L - x_0}{t + \Delta t - t_i}) = (t + \Delta t - t_i)R(\frac{L}{t + \Delta t - t_i}) \qquad \forall i \in I \end{aligned} \tag{4.9}$$

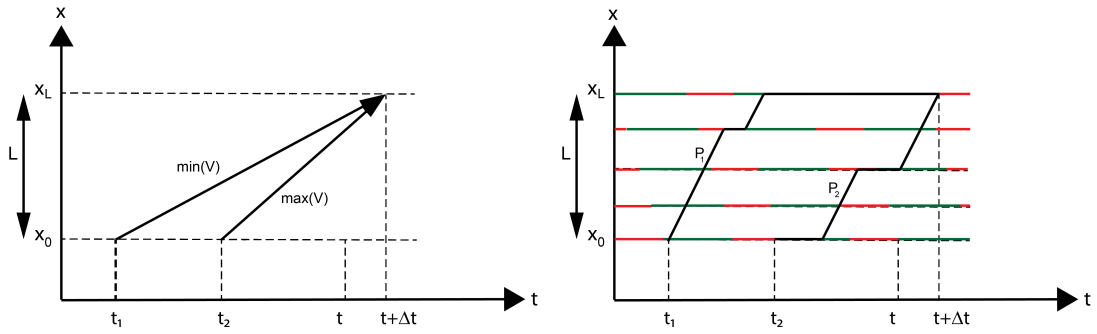


Figure 4.3: (a) Time-space diagram and the solution space between minimum and maximum speeds and (b) time-space and signal timing diagram with instances of practical cut paths.

The downstream cumulative accumulation $N(x_L, t + \Delta t)$ is approximated using VT and the upstream cumulative accumulation of an earlier time instant $N(x_0, t_i)$. According to this formula, an observer moving along an arterial road with speed V cannot be overtaken by more than $\frac{L}{V}R(V)$ vehicles. Since Eq. (4.9) is derived considering only forward wave speeds and ignoring backward speeds and arterial capacity constraints, the above equality constraint transforms into an inequality as expressed in Eq. (4.10). In other words, the right-hand side value constructs an upper bound for the left-hand side value.

$$N(x_L, t + \Delta t) \leq N(x_0, t_i) + (t + \Delta t - t_i)R(\frac{L}{t + \Delta t - t_i}) \qquad \forall i \in I \tag{4.10}$$

Among the tangent cuts that construct the upper bound of MFD, the cut with slope $min(V)$ has the highest cost (provided that cuts are estimated using the method of practical cuts introduced in [10]). In mathematical terms, $R(min(V)) \geq R(V); \forall V \in [min(V), max(V)]$; thereafter, for paths from the solution point to the opposing bound-

ary end emanated over discrete time steps, the following inequality is valid:

$$N(x_L, t+\Delta t) \leq N(x_0, t_i)+(t+\Delta t-t_i)R(V) \leq N(x_0, t_i)+(t+\Delta t-t_i)R(min(V)) \qquad \forall i \in I \tag{4.11}$$

Based on the path discretization method of [10], any valid path from the solution space to the boundary starts at the end of a red signal, and thus, the time windows between the start times of any two valid paths is always an integer multiple of the signal cycle time.

For a reservoir with regular signal settings and topology, maximum flow capacity is equal to q_c during a green signal; therefore, reservoir capacity per unit time is equivalent to $(g/c)q_c$ (veh/sec). This maximum capacity restricts the amount of change in the downstream cumulative accumulation between any two time steps Δt. The difference in cumulative accumulation $N(x_L, t + \Delta t) - N(x_L, t)$ between two paths according to the maximum flow capacity constraint is calculated as follows:

$$N(x_L, t + \Delta t) \leq N(x_L, t) + (g/c)q_c\Delta t \qquad \forall i \in I \tag{4.12}$$

[98] proposed adding an extra constraint for any discretized time interval to estimate the solution of the KW model exactly. Adding this constraint finds an exact solution, and its inclusion is essential to prevent acceleration fans or rarefaction waves from occurring where a sudden increase in demand results in a spike in outflow rather than a continuously increasing outflow as expected with KW models. Consider the time interval $[t_i, t_{i+1}]$ in the solution space $[t_1, t_2]$. Due to time-discretization, the values of cumulative accumulation are constant in any defined time interval. The open interval constraint is established based on the flow of the interval; i.e., $q = \frac{N(x_0,t_{i+1})-N(x_0,t_i)}{\Delta t}$.

Therefore, if $\frac{L}{t_i} \leq V \leq \frac{L}{t_{i+1}}$, the following constraint is satisfied:

$$N(x_L, t + \Delta t) \leq N(x_0, t_i) + q(\frac{L}{V} - t_i) + (t + \Delta t - \frac{L}{V})R(V) \qquad \forall V \in [min(V), max(V)] \tag{4.13}$$

Similar to forward waves, a set of constraints is derived to estimate backward waves V' to establish an upper bound for the upstream cumulative accumulation. The solution for backward waves should satisfy the following constraints:

$$N(x_0, t + \Delta t) \leq N(x_L, t_i) + (t + \Delta t - t_i)R(max(V')) \qquad \forall i \in I \tag{4.14}$$

$$N(x_0, t + \Delta t) \leq N(x_0, t) + (g/c)q_c\Delta t \qquad \forall i \in I \tag{4.15}$$

$$N(x_o, t + \Delta t) \leq N(x_L, t_i) + q(\frac{L}{V'} - t_i) + (t + \Delta t - \frac{L}{V'})R(V') \qquad \forall V' \in [min(V'), max(V')] \tag{4.16}$$

I propose two algorithms to solve the hyper-link model and estimate the flow that can be sent to other reservoirs at the reservoir downstream end (sending flow), and the flow that can be received from other reservoirs at the reservoir upstream end (receiving capacity). Algorithms 1 and 2 contain the procedures for computing sending flow and receiving capacity, respectively. Algorithm 1 divides the time interval $[t+\Delta t-\frac{L}{min(V)}, t+\Delta t - \frac{L}{max(V)}]$, associated with the start times of the minimum and maximum forward waves, into time intervals of length Δt. Three sets of constraints are then imposed on the downstream cumulative accumulation. For waves emanated between any two consecutive cuts, Eq. (4.11) is satisfied. This equation builds the constraints associated with waves emanating from the solution point at discretized time steps. To impose the constraints related to the MFD cuts, the algorithm find outs the time interval from which any cut is emerged (i.e., $t_i \leq \frac{L}{V} \leq t_{i+1}$). Eq. (4.13) is then formulated to establish the related constraint using the flow of this time interval and the cost function of the MFD cut within this time interval. Lastly, the algorithm imposes constraints related to the maximum flow capacity of the hyper-link with traffic signals using Eq. (4.12). An

identical procedure is proposed to estimate the upstream cumulative accumulation and receiving capacity using Algorithm 2.

The hyper-link model is rephrased as the maximum value of $N(x, t + \Delta t) - N(x, t)$ using the set of constraints defined in this section. Based on the estimated Moskowitz functions of upstream and downstream ends of the reservoir, sending flow sf, and receiving capacity rc are computed as follows [105]:

$$sf(t + \Delta t) = N(x_L, t + \Delta t) - N(x_L, t) \tag{4.17}$$

$$rc(t + \Delta t) = N(x_0, t + \Delta t) - N(x_0, t) \tag{4.18}$$

Δt is chosen such that the Courant-Friedrichs-Lewy (CFL) condition is satisfied; i.e., the time step duration should be less than the minimum time required to transfer the hyper-link length, $\Delta t \leq \frac{L}{max(V)}$ for forward waves and $\Delta t \leq |\frac{L}{min(V')}|$ for backward waves. The next section explains how the hyper-node model estimates the flow transferring between reservoirs in a multi-reservoir urban network.

4.4.2 Hyper-node model

In conventional link-based DNLs, nodes operate as a medium for flow exchange between links while assuring adherence to the demand and supply constraints imposed by incoming and outgoing links and the node itself. Nodes can impose constraints inherent to the presence of traffic signals or any other kind of traffic control on transferring flow. Node models can also be adopted to model flow transition between adjacent reservoirs in a multi-reservoir urban network. To distinguish nodes that model flow transition between reservoirs from regular nodes, I refer to them as hyper-nodes throughout this chapter.

Supply capacity or receiving flow capacity from any of a reservoir's neighbouring reservoirs is not fixed. They depend on traffic state dynamics, the remaining capacity of the reservoir itself, and on the sending flow from all neighbouring reservoirs that com-

Algorithm 1: Estimating sending flow

- initialize a set of forward cuts tangent to MFD
  $C = \{V | V \in [min(V), max(V)]\}$
- set integer number of time intervals between the time when minimum and maximum forward speeds emanated
  $M = \lfloor \frac{L(\frac{1}{max(V)} - \frac{1}{min(V)})}{\Delta t} \rfloor$
- estimate downstream cumulative accumulation with regard to time t_1
  $N(x_L, t + \Delta t) = N(x_0, t_1) + (t + \Delta t - t_1)R(min(V))$

for $i = 1 : M$ **do**
  for $V \in C$ **do**
    if $\frac{L}{V_C} \leq t_i < \frac{L}{V_{C+1}}$ **then**
      - update downstream cumulative accumulation with regard to the updated time interval $[t_i, t_i + \Delta t]$
        $N(x_L, t + \Delta t) = min(N(x_L, t + \Delta t), N(x_0, t_i) + (t + \Delta t - t_i)R(V))$
    else
      if $\frac{L}{V_C} - \Delta t \leq t_i \leq \frac{L}{V_C}$ **then**
        - estimate flow within time interval $[t_i, t_i + \Delta t]$
          $q = \frac{N(x_0, t_{i+1}) - N(x_0, t_i)}{\Delta t}$
        - update downstream cumulative accumulation with regard to the updated time interval $[t_i, t_i + \Delta t]$
          $N(x_L, t + \Delta t) =$
          $min(N(x_L, t + \Delta t), N(x_0, t_i) + q(\frac{L}{V} - t_i) + (t + \Delta t - \frac{L}{V})R(V)))$
      else
      end
    end
  end
end

- update downstream cumulative accumulation considering capacity constraint
  $N(x_L, t + \Delta t) = min(N(x_L, t + \Delta t), N(x_L, t) + (g/c)q_c \Delta t)$
- set sending flow
  $sf(t + \Delta t) = N(x_L, t + \Delta t) - N(x_L, t)$

Algorithm 2: Estimating receiving capacity

- initialize a set of backward cuts tangent to MFD
  $C' = \{V' < 0 | V' \in [min(V'), max(V')]\}$
- set integer number of time intervals between the time when minimum and maximum backward speeds emanated
  $M' = \lfloor \frac{L(\frac{1}{min(V')} - \frac{1}{max(V')})}{\Delta t} \rfloor$
- estimate downstream cumulative accumulation with regard to time t_1
  $N(x_0, t + \Delta t) = N(x_L, t_1) + (t + \Delta t - t_1)R(max(V))$

for $i = 1 : M'$ **do**
  for $V' \in C'$ **do**
    if $-\frac{L}{V'_{C'+1}} \leq t_i < -\frac{L}{V'_{C'}}$ **then**
      - update downstream cumulative accumulation with regard to the updated time interval $[t_i, t_i + \Delta t]$
        $N(x_0, t + \Delta t) = min(N(x_0, t + \Delta t), N(x_L, t_i) + (t + \Delta t - t_i)R(V'))$
    else
      if $-\frac{L}{V'_{C'}} - \Delta t \leq t_i \leq -\frac{L}{V'_{C'}}$ **then**
        - estimate flow within time interval $[t_i, t_i + \Delta t]$
          $q = \frac{N(x_L, t_{i+1}) - N(x_L, t_i)}{\Delta t}$
        - update downstream cumulative accumulation with regard to the updated time interval $[t_i, t_i + \Delta t]$
          $N(x_0, t + \Delta t) =$
          $min(N(x_0, t + \Delta t), N(x_L, t_i) + q(-\frac{L}{V'} - t_i) + (t + \Delta t + \frac{L}{V'})R(V')))$
      else
      end
    end
  end
end

- update downstream cumulative accumulation considering capacity constraint
  $N(x_0, t + \Delta t) = min(N(x_0, t + \Delta t), N(x_0, t) + (g/c)q_c \Delta t)$
- set receiving capacity
  $rc(t + \Delta t) = N(x_0, t + \Delta t) - N(x_0, t)$

pete for the same space. [62] introduced the constraining effect of the reservoir supply function on transition flow. [62], [17] and [106] determined the supply function similar to CTM as a non-increasing function of accumulation that has a critical accumulation as its tipping point. Other studies expressed a general function for the receiving capacity rc_{ij} of reservoir i that is assigned to its neighbouring reservoir $j \in \tilde{H}_i$ as a function of accumulation n_i as follows [18**?**]:

$$rc_{ij}(n_i) = \begin{cases} rc_{ij}^{max} & \text{if} \quad 0 \leq n_i < \chi n_{i,jam} \\ \frac{rc_{ij}^{max}}{1-\chi}(1 - \frac{n_i}{n_{i,jam}}) & \text{if} \quad \xi n_{i,jam} \leq n_i \leq n_{i,jam} \end{cases} \tag{4.19}$$

Where rc_{ij}^{max} denotes the maximum value of the reservoir receiving capacity assigned to its neighbour, and $0 < \chi < 1$ is a parameter that specifies the tipping point of the supply function that is associated with the point at which receiving capacity decreases from the maximum value rc_{ij}^{max}. As specified by [**?**], the tipping point of the supply function is higher than the critical accumulation point.

The receiving capacity function presented in previous studies is often overstated because it is estimated solely as a function of the available capacity of a reservoir without considering the other neighbouring reservoirs' claim over the same space. In other words, the available capacity is deemed fully obtainable for each neighbouring reservoir. Overestimating the available capacity in situations with high demand and congestion might lead to gridlock on the urban network. In previous studies, the simulations do not result in gridlock because the models implemented perimeter controls, which prevent reservoirs from reaching high accumulation levels and serve as a natural boundary. [63] demonstrated that, in such circumstances, receiving capacity can be ignored for the sake of brevity. Hence, as opposed to previous studies that assumed transferring flow to a reservoir should be less than its available capacity, I assume that the sum of transferring flows from all the neighbouring reservoirs must not violate the available capacity. In this section, I present a general approach to estimate transition flow between reservoirs

by implementing reduction factors and proportional sharing of capacity assigned to each reservoir. Transition flow between any two neighbouring reservoirs should comply with the following demand and supply constraints:

- Transition flow from any reservoir i to its neighbouring reservoir $j \in \tilde{H}_i$ should be less than the sending flow of reservoir i, i.e., $f_{ij} \leq sf_i; \forall i \in R, j \in \tilde{H}_i$.
- The sum of transition flows to any reservoir j from its neighbouring reservoirs $i \in \tilde{H}_j$ should be less than the supply capacity of reservoir j, i.e. $\sum_{i \in \tilde{H}_j} f_{ij} \leq rc_j; \forall i \in R, j \in \tilde{H}_i$.

To estimate the portion of the receiving capacity of a given reservoir that is assigned to each of its neighbouring reservoirs (i.e., distribute supply over incoming reservoirs), a reduction factor on sending flow per turn can be applied. [107] introduced distribution fractions proportional to the demand, and [108] identified the fractions proportional to capacity and discussed that previous models led to unrealistic results under certain conditions (please see [102] for more information on different reduction factors and their properties). In this chapter, I adopt the approach presented by [102] and model the reduction factor ζ_{ij} imposed on sending flow from reservoir i to its neighbouring reservoir $j \in \tilde{H}_i$, as a function of available receiving capacity rc_j and sending flow sf_{ij}, in the form of a fixed-point problem expressed as follows [102]:

$$\zeta_{ij} = F(sf_{ij}, rc_j(\zeta_{ij})) \qquad \forall i \in \tilde{H}_j, \forall j \in R \tag{4.20}$$

The implicit function expressed above indicates that reduction factors are non-separable, and their estimation is not a trivial task. In link-based DNL problems, using first-order node models for solving problems of the above form is common. In our multi-reservoir network, reservoirs are modelled as hyper-links and, thus, are comparable to a conventional link-based traffic network. Therefore, I solve the above fixed-point formula by adopting an algorithm similar to that of [102] and modify it for our multi-reservoir network using Algorithm 3.

The algorithm initializes by determining sets of incoming and outgoing reservoirs at any hyper-node; thereafter, it initializes incoming reservoirs' sending flow and outgoing reservoirs' receiving capacity. The sending flow of incoming reservoirs estimated using Algorithm 1 is distributed over outgoing reservoirs using the logit route choice model, and the remaining capacity of outgoing reservoirs is considered equal to the receiving capacity estimated using Algorithm 2. Algorithm 3 is developed on the basis of applying a reduction factor on sending flows to balance supply and demand. The reduction factor for any outgoing reservoir is computed as the ratio of outgoing reservoirs' receiving capacity to the sum of the sending flows for the reservoir; i.e., $\zeta_j = \frac{\tilde{C}_j}{\sum_{i \in X} sf_{ij}}; j \in Y$. The minimum value of reduction factors $\zeta = min_j(\zeta_j)$ is considered as the hyper-node reduction factor and is imposed on all sending flows from incoming reservoirs. An incoming reservoir is demand-constrained if the demand inflow is less than the reduced receiving capacity of the target reservoir. In this case, the transition flow is equal to the sending flow, the incoming reservoir is removed from the set of incoming reservoirs, and the outgoing reservoirs' remaining capacity is updated by subtracting the capacity assigned to processed reservoirs. The algorithm then proceeds to the next iteration. In contrast, if none of the incoming reservoirs are demand-constrained, transition flow is equal to the outgoing reservoirs' reduced receiving capacity, and the algorithm terminates. The developed algorithm terminates within a maximum of $J = |\tilde{H}_j|$ iterations, where J is equal to the number of neighbouring reservoirs of reservoir j.

4.4.3 Hyper-link/ hyper-node interface

To solve the multi-reservoir DNL and DTA problems, an algorithm similar to that of link transmission model (LTM) is developed. [105] suggested a sequential three-step algorithm to solve LTM using a link-node interface (please refer to [109], for a comprehensive review of a general link-node interface that combines link and node models). LTM is based on [104]'s simplified kinematic wave theory (SKWT) and uses

Algorithm 3: Estimating transition flow

Initialization:
- initialize reservoirs' sending flow $sf_{ij}, \forall i \in R$ using Algorithm 1
- initialize reservoirs' receiving capacity $rc_i, \forall i \in R$ using Algorithm 2

for *any hyper-node in network* **do**
- set iteration number $it = 1$
- determine the set of incoming reservoirs X and outgoing reservoirs Y
- determine demand inflow from any incoming reservoir $i \in X$ to any outgoing reservoir $j \in Y$ using logit route choice model $sf_{ij} = \frac{exp^{-\theta\tau_j}}{\sum_j exp^{(-\theta\tau_j)}}$
- set remaining capacity of the outgoing reservoirs $\tilde{C}_j = rc_j, \forall j \in Y$
- **while** $X \neq 0$ **do**
  - calculate reduction factor for outgoing reservoirs $\zeta^{(k)} = min_j(\frac{\tilde{C}_j}{\sum_{i \in X} sf_{ij}})$
  - **if** *any of the incoming reservoirs i are demand-constrained, i.e.* $sf_{ij} \leq \zeta^{(k)}\tilde{C}_j$ **then**
    - estimate transition flow $f_{ij} = sf_{ij}$
  - **else**
    - estimate transition flow $f_{ij} = \zeta^{(k)}\tilde{C}_j$
  - **end**
  - update the set of unprocessed reservoirs by removing i from set X $X = X|i$
  - update remaining capacity by subtracting the capacity assigned to processed reservoirs $\tilde{C}_j^{(k+1)} = \tilde{C}_j^{(k)} - \sum_{i \notin X} f_{ij}$
  - set $k = k + 1$
- **end**

end

a triangular shape FD along with cumulative vehicle numbers estimated using vehicle count trajectories in time and space [104, 7, 110]. [111, 112] extended the model to include the case of a general LTM (GLTM) with a concave FD. Compared to an LTM with two branches and only two kinematic waves, GLTM, which has several waves, is more complex; however, GLTM produces a more realistic propagation of traffic, travel delays, and travel time.

In this chapter, a sequential approach based on LTM is presented to model traffic propagation in a multi-reservoir urban network using a general interface of a hyper-link model as discussed in Section 4.4.1 and a hyper-node model as explained in Section 4.4.2. [113] used an iterative approach to satisfy the Courant-Friedrich-Lewy (CFL) condition, which defines an upper bound on length of time steps for updating cumulative accumulations. In their study, the entire simulation period was updated in each iteration. [114] developed an iterative algorithm for LTM based on a space-time discretized space, which, contrary to earlier studies, did not impose an upper bound on time step length; the algorithm significantly reduced the computational requirements. Herein, I use the basic LTM algorithm of [115]. Algorithm 4 illustrates the steps for solving the general hyper-link/hyper-node interface or the MRDNL model proposed in this chapter.

For any time step, the algorithm firstly initializes demand and then estimates the hyper-link sending flow and receiving capacity using Algorithms 1 and 2, respectively. The hyper-node model, computed using Algorithm 3, estimates the transition flow at hyper-node from an incoming reservoir to an outgoing reservoir; thereafter, upstream and downstream cumulative accumulations are derived by adding transition flows estimated during the previous step to the current cumulative accumulation values. After finding the reservoir traffic state in terms of cumulative accumulations, reservoir travel time can be estimated.

Algorithm 4: Multi-reservoir network loading model

At any time step $[t, t + \Delta t]$
- initialize demand
- determine sending flow for reservoirs using hyper-link Algorithm 1
- determine receiving capacity for reservoirs using hyper-node Algorithm 2
- determine transition flows between reservoirs using Algorithm 3
- update upstream cumulative accumulation for reservoirs
  $N_i(x_0, t + \Delta t) = N_i(x_0, t) + \sum_{j \in \tilde{H}_i} f_{ij}$
- update downstream cumulative accumulation for reservoirs
  $N_i(x_L, t + \Delta t) = N_i(x_L, t) + \sum_{j \in \tilde{H}_i} f_{ij}$
- determine inner-reservoir instantaneous travel time τ such that
  $N_i(x_L, t + \Delta t) = N_i(x_0, t + \Delta t - \tau(t))$

4.5 Multi-class dynamic network loading under mixed traffic conditions

Consider a multi-reservoir urban network that encompasses two vehicle classes: conventional vehicles and CAVs (i.e., $m = 2$ in Eq. (4.2)). Different traffic modes affect the traffic network in dissimilar ways. [116] addressed the issue for an MFD-based traffic network by proposing a 3D-MFD using a simulation for a mixed traffic network with conventional vehicles and buses. Such networks with two modes of traffic are usually modelled as 3D-MFD (refer to [93] for a review of these models). However, I modelled our MFD-based traffic network with mixed conventional vehicles and CAVs in a single-class fashion using a transformed 2D-MFD. I postulated that CAVs and conventional vehicles have identical sizes and physical features; therefore, the marginal effect of one unit increase in any vehicle class is equal.

In this chapter, I assume the only features that distinguish CAVs from conventional vehicles are that for CAVs (i) more data on travel times can be obtained from vehicle-to-vehicle (V2V) [117, 118] and vehicle-to-infrastructure (V2I) connections, and (ii) reaction times are shorter due to driver automation. CAVs provide more information on traffic conditions that result in a lower stochasticity in route travel time compared to conventional vehicles. Thus, the logit model is adopted to reflect this difference in travel time data. The logit model that estimates the probability of choosing reservoir

r' after traversing reservoir r'' is shown as follows:

$$Prob(r'' \rightarrow r') = \frac{exp^{-\theta\tau_{r'}}}{\sum_{r\in\tilde{H}_{r''}} exp^{-\theta\tau_r}} \qquad \forall r' \in \tilde{H}_{r''}, r, r', r'' \in R \tag{4.21}$$

Where $\theta > 0$ denotes road users' level of knowledge of traffic conditions. With $\theta \rightarrow \infty$ stochasticity of the model decreases and the obtained solution becomes more deterministic; therefore, for CAVs, since they acquire more information on traffic conditions, higher values for θ are considered as well.

CAVs are also expected to maintain a shorter space headway between themselves and the leading vehicles, and this shorter headway results in a higher flow capacity compared to conventional vehicles. [119] approximated the capacity of a link with mixed conventional vehicles and CAVs using CTM and based on a trapezoidal FD:

$$q = \frac{\gamma}{\gamma((1 - cav\%)rt_{cv} + cav\%rt_{cav}) + \iota} \tag{4.22}$$

In the above equation, $cav\%$ indicates the penetration rate of CAVs in a network. rt_{cav} (sec) and rt_{cv} (sec) respectively denote CAV and conventional vehicles' reaction time, and ι denotes vehicular length, which is considered equal for both vehicle classes. The intention of the current work is not to precisely model CAVs but to examine the network-wide effect on a large-scale urban network under various penetration rates; thus, I use the above-mentioned link flow capacity function in Eq. (4.22).

By modelling a mixed traffic network as a single-level model, a network's dynamics are described using a 2D-MFD. Introducing CAVs into a traffic network and changing the fundamental diagram of the links, the piece-wise linear MFD, approximated using the method of cuts, has to be reconstructed according to the new CAV market penetration rate that includes a modified maximum flow and backward waves. As the CAV penetration rate increases, higher values for maximum flow and backward speeds are achieved. Further, the cost function $R(V)$ associated with speed V also increases. From a mathematical perspective, by increasing the cost function, the interception of

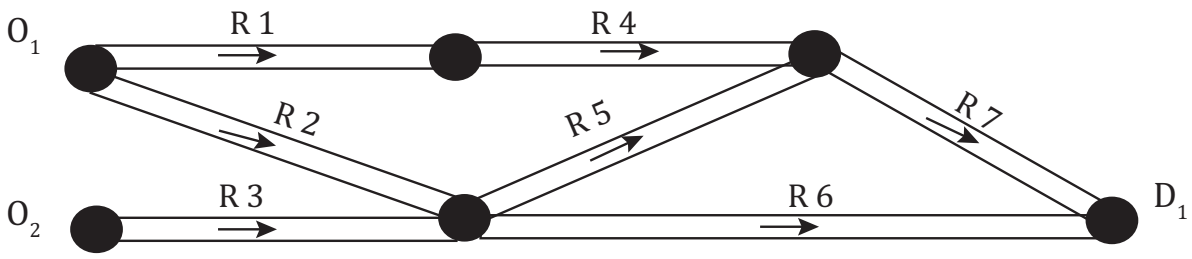


Figure 4.4: Hypothetical urban network divided into seven reservoirs

practical cuts with the y-axis yields higher values; therefore, higher values for flow are achieved at any time.

4.6 Numerical results and discussion

In this section, the developed MRDNL model is tested for a hypothetical urban network consisting of several homogeneous reservoirs. Each reservoir in the network is considered to be a hyper-link composed of several consecutive non-overlapping links separated by traffic signals, and possessing some homogeneity (i.e., links with an identical fundamental diagram and of equal length) and regularity (i.e., identical signal settings: $g_i = g, c_i = c, o_i = i.o$). The urban network is depicted in Figure 4.4 and consists of 7 hyper-links and 6 hyper-nodes. Each link in the reservoir follows a triangular fundamental diagram. The properties of links within each reservoir and traffic signal timings are listed in Table 4.1. For simplicity, signal timings, the number of existing links in reservoirs, link lengths, and fundamental diagrams are considered identical.

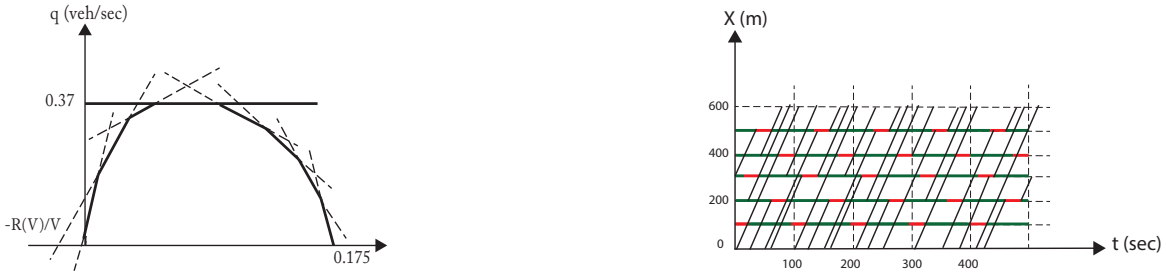
For each reservoir in the urban network, MFD is estimated using VT and the method of cuts introduced by [10] by utilizing the information on link characteristics and signal settings in Table 4.1. [10] introduced simple pre-scanning approaches to eliminate non-tight cuts. For regular traffic networks with regular signal timings, this method

provides a tight estimation, but for irregular traffic networks, it only provides an upper bound. In irregular networks, only SVG with both forward and backward speeds is sufficient to provide a tight estimation. Without losing generality, I assume that reservoirs considered in this work are regular, and thus, I adopt the cut estimation method of [10] to provide a tight MFD estimation. However, the developed DNL model can be used for irregular networks as well. The resulting piecewise linear MFD is depicted in Figure 4.5 (a), and the corresponding time-space diagram for forward waves along with traffic signal timing are depicted in Figure 4.5 (b). In Figure 4.5 (a), the stationary cut $S = q_c g/c$ corresponds to the capacity of a traffic signal during the green light, which demonstrates a reservoir's flow capacity. For reservoirs that have a stationary cut value higher than the intersection point of the forward and backward waves, capacity is equal to the intersection point value.

Propagation of flow within the network is simulated using the developed MRDNL model for $T = 200$ min with time step increments of $\Delta t = 1$ min and a simulation warming period of 50 min. Similar to [92] and [88], CTM model [94, 95] is used as a benchmark analysis model to simulate the network and compare its performance with that of MRDNL. For CTM model, time step increments of $\Delta t = 20$ sec are considered. Then, the size of each cell is estimated as a product of time step increment Δt and the free flow speed γ of a link. Simulations are conducted under two different demand profiles as depicted in Figure 4.6. The network is empty at the start of the simulation at $t = 0$. In the low demand scenario, the network is loaded with low demand. Then, the demand is increased to examine the performance of the MRDNL model under high demand. Demand loading rate is considered equal for both CTM and MRDNL models. $O_1 - D_1$ is the demand loaded in reservoirs 1 and 2, while $O_2 - D_1$ is the demand loaded in reservoir 3. For other reservoirs, demand inflow is distributed based on the turning ratios approximated using the logit route choice model at each hyper-node. All vehicles arrive at the D_1 hyper-node as their final destination. Cumulative curves in Figures 4.7 and 4.8 are obtained assuming that all vehicles are of conventional type. All simulations

Table 4.1: Properties of a multi-reservoir urban network

Reservoir	Number of links	Link length (m)	Free-flow speed (m/sec)	Backward speed (m/sec)	Jam density (veh/m)
1,...,7	9	200	10	5	0.175
Number of traffic signals	cycle (sec)	green (sec)	offset (sec)		
8	100	85	59		



(a) NC (uniform) (b) NC (radial)

Figure 4.5: (a) A piecewise linear MFD estimated using VT and the method of cuts and (b) the associated time-space diagram

are conducted using MATLAB R2019a (9.6.0), on a 64-bit macOS with 2.2-GHZ Intel Core-i7 and 16-GB of RAM.

The resulting cumulative upstream and downstream accumulations in reservoirs obtained using MRDNL and CTM under low and high demand profiles are illustrated in Figures 4.7 and 4.8, respectively. Generally speaking, both models exhibit the same trend, suggesting that the developed MRDNL model is capable of representing traffic dynamics nearly as well as a mesoscopic traffic model. It is observed that under CTM, higher accumulation values for some reservoirs are obtained at any time instant. In the low demand scenario, the downstream cumulative accumulation curve is identical to the upstream cumulative accumulation curve with a time delay equal to the free-flow travel time ($\approx$ 4 min). The reservoirs are in a free-flow traffic condition as can be estimated from accumulation (the vertical distance between upstream and downstream accumulation at any time instant) and travel time (the horizontal distance between upstream and downstream accumulation per any accumulation) values of reservoirs. The precise

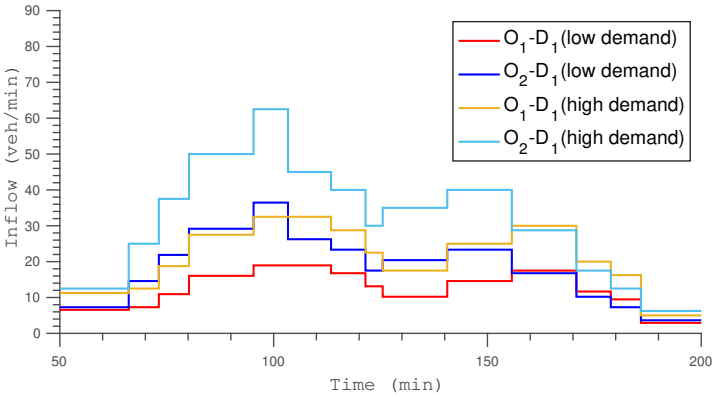


Figure 4.6: Demand profile

values of accumulation and inner-reservoir travel time are shown in Figures 4.7 and 4.8, respectively. Accumulation values shown in Figure 4.7 are less than the critical accumulation value (=116 veh), which also confirms the free-flow traffic condition under the low-demand scenario.

In the high demand scenario, demand for reservoirs 6 and 7 steadily increase during the simulation period. As a result, cumulative demand exceeds cumulative upstream accumulation at any time instant, and thus, bottlenecks form at the downstream ends of reservoirs 2, 3, 4, and 5. The downstream boundary transfer capacity for reservoirs 6 and 7 is assumed to be unconstrained; therefore, accumulations stabilize around the critical accumulation level at $t = 106$ min for reservoir 6 and $t = 97$ min for reservoir 7. Indeed, only an inflow equal to the maximum outflow capacity can enter any reservoir at any time instant.

With high demand entering the traffic network in the high demand scenario, congestion and queueing firstly appear at the downstream ends of reservoirs 4 and 5 at $t = 85$ min. With steady increase of demand for the network during the first 100 minutes of the simulation and the congestion propagating backward at reservoirs 4 and 5, accumulation increases in these reservoirs. At $t = 100$ min, reservoirs 4 and 5 operate under congested traffic condition and at extremely high accumulation levels ($n_4 = 251$ veh and $n_5 = 214$ veh while $n_{4,jam} = n_{5,jam} = 315$ veh). Travel time in reservoirs

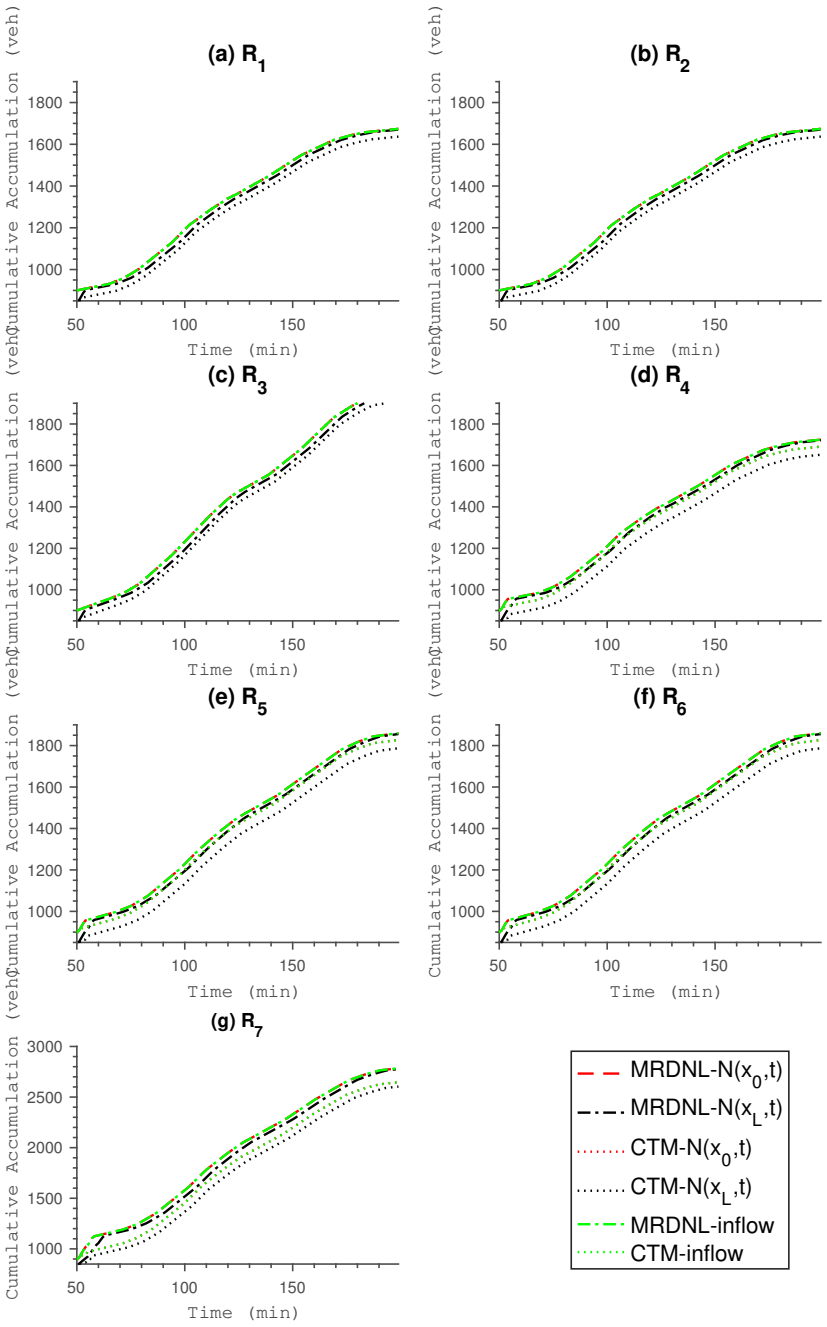


Figure 4.7: Cumulative upstream and downstream accumulations of the reservoirs in the low demand scenario, CAV=0% and logit route choice $\theta = 0.1$

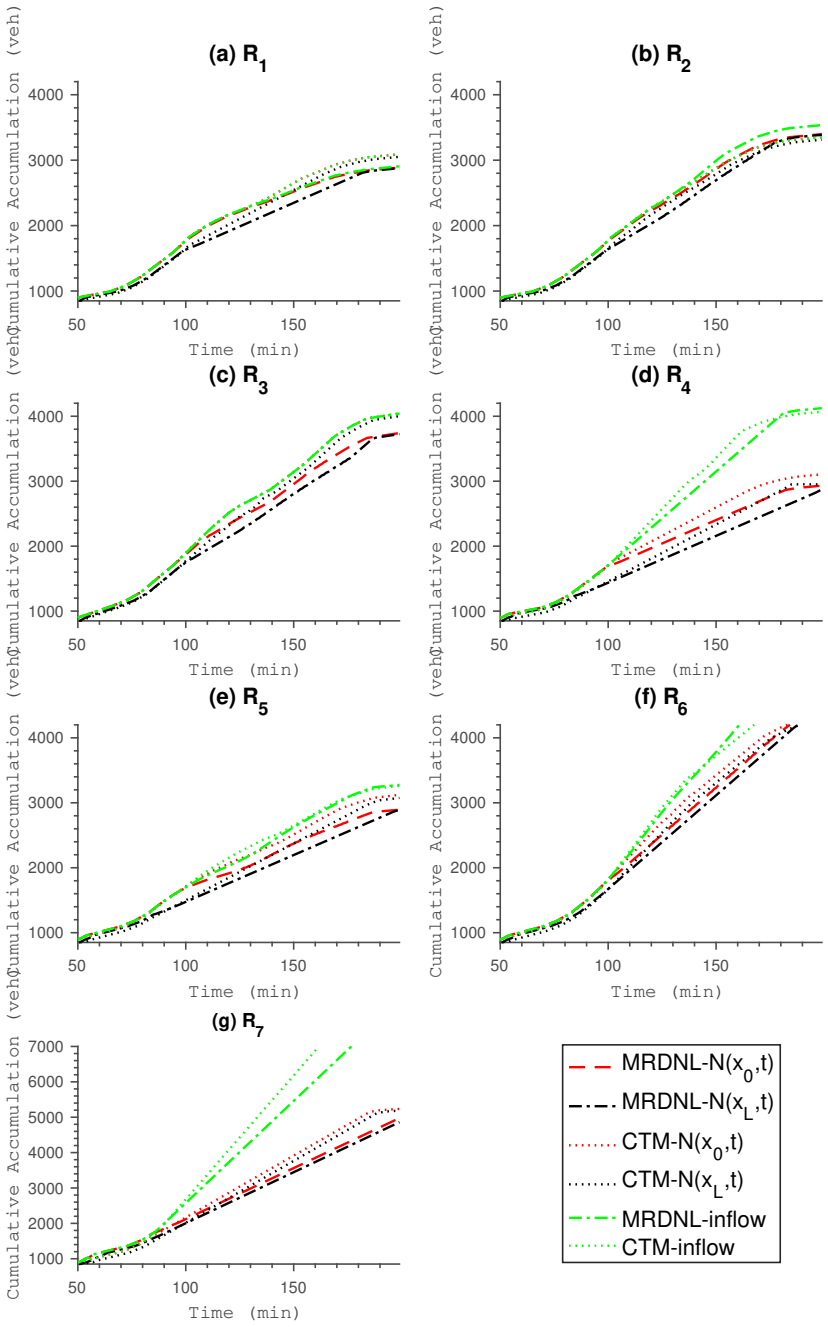
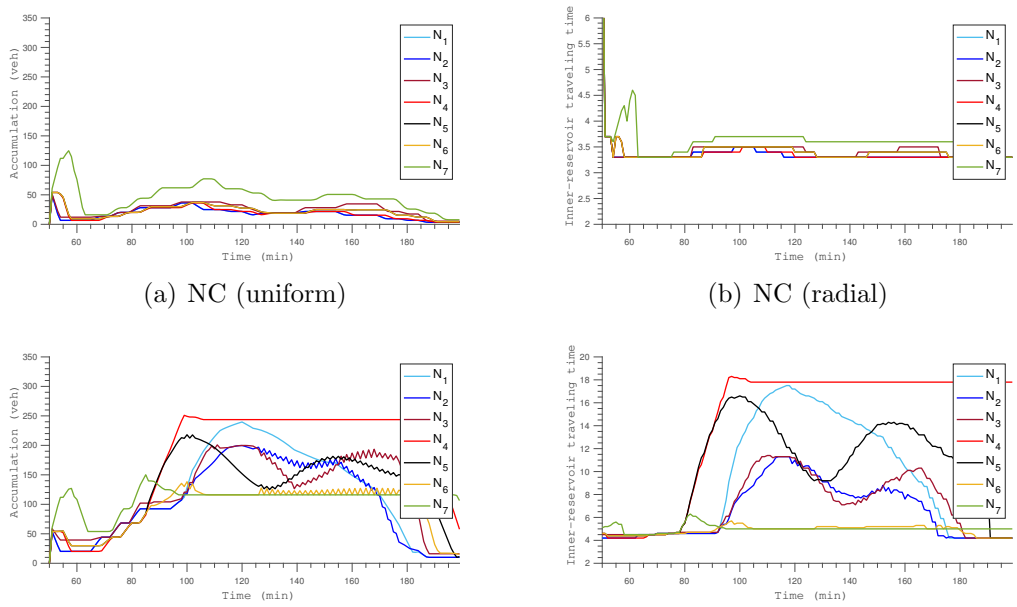


Figure 4.8: Cumulative upstream and downstream accumulations of the reservoirs in the high demand scenario, CAV=0% and logit route choice $\theta = 0.1$



(a) NC (uniform)

(b) NC (radial)

(c) BC (uniform)

(d) BC (radial)

Figure 4.9: Low demand scenario (a) accumulation of the reservoirs and (b) inner-reservoir travel time, high demand scenario (c) accumulation of the reservoirs and (d) inner-reservoir travel time.

4 and 5 at $t = 100$ min is respectively 18 and 16.4 minutes as shown in the travel time plot in Figure 4.8. Reservoir 7 remains a bottleneck, which constrains the outflow from reservoirs 4 and 5 until the end of the simulation period. Accordingly, reservoir 4 operates under congested conditions during the whole simulation period; however, a different traffic condition in reservoir 5 is observed. The reason for the descending trend in the accumulation is that when reservoir 6 forms a bottleneck at the downstream ends of reservoirs 2 and 3 at $t = 100$ min, the inflow from these reservoirs to reservoir 6 is constrained. As a result of restricting the inflow and based on the node model adopted in this chapter, inflow to reservoir 5 is constrained as well; therefore, reservoir 5 returns to a lower accumulation level at $t = 126$ min and remains in this condition until $t = 133$ min. Given that reservoir 7 forms a bottleneck at the downstream end of reservoir 5, and despite the reduced inflow to reservoir 7, accumulation gradually increases in reservoir 5. The reservoir, however, recovers to free flow traffic conditions at the end of

the simulation period at $t = 188$ min.

At $t = 100$ min, cumulative demand for reservoir 4 exceeds the capacity of this reservoir, and thus, a bottleneck forms at the downstream end of reservoir 1, causing congestion. Reservoir 1 remains in this condition until $t = 180$ min, at which point the accumulation begins to reduce as a result of demand reduction in the network. Likewise, reservoirs 2 and 3 also experience high accumulation levels during most of the simulation period. All reservoirs recover to free flow traffic conditions at the end of the simulation period. This result indicates the efficiency of our developed MRDNL model in capturing multiple traffic waves within a network and describing the evolution of traffic dynamics. Furthermore, the viscosity issue observed in accumulation-based models mentioned in [36] does not exist because the reaction of outflow to the surge in demand is not sudden but gradual and happens over an extended time period.

4.6.1 Numerical analysis for the multi-reservoir DNL under mixed traffic conditions

The multi-reservoir urban network in Figure 4.4 is comprised of two vehicle classes, conventional vehicles and CAVs. I assume that CAVs and conventional vehicles have identical sizes and physical features. Further, I assume that vehicles use the same traffic infrastructure, and there is no dedicated lane for CAVs. The reaction time of CAVs $rt_{cav} = 0.5$ sec, and the reaction time of conventional vehicles $rt_{cv} = 1.5$ sec, are considered. The length of all vehicles in both classes is $\iota = 6$ m. I assume that the demand is not affected by the presence of CAVs and their positive effect on network capacity (i.e., demand elasticity is not considered).

The resulting MFDs (outflow vs. accumulation) for reservoirs 1, 2, 3, 4, and 5 obtained from simulation results under various CAV penetration rates are depicted in Figure 4.10, for the low demand scenario ($\theta = 0.1$ for conventional vehicles and $\theta = 0.3$ for CAVs). As mentioned earlier, boundary transfer capacity for reservoirs 6 and 7 is unlimited. These reservoirs operate under uncongested and maximum flow traffic conditions under any demand profile; thus, I am focusing on reservoirs 1 to 5 for the sake of

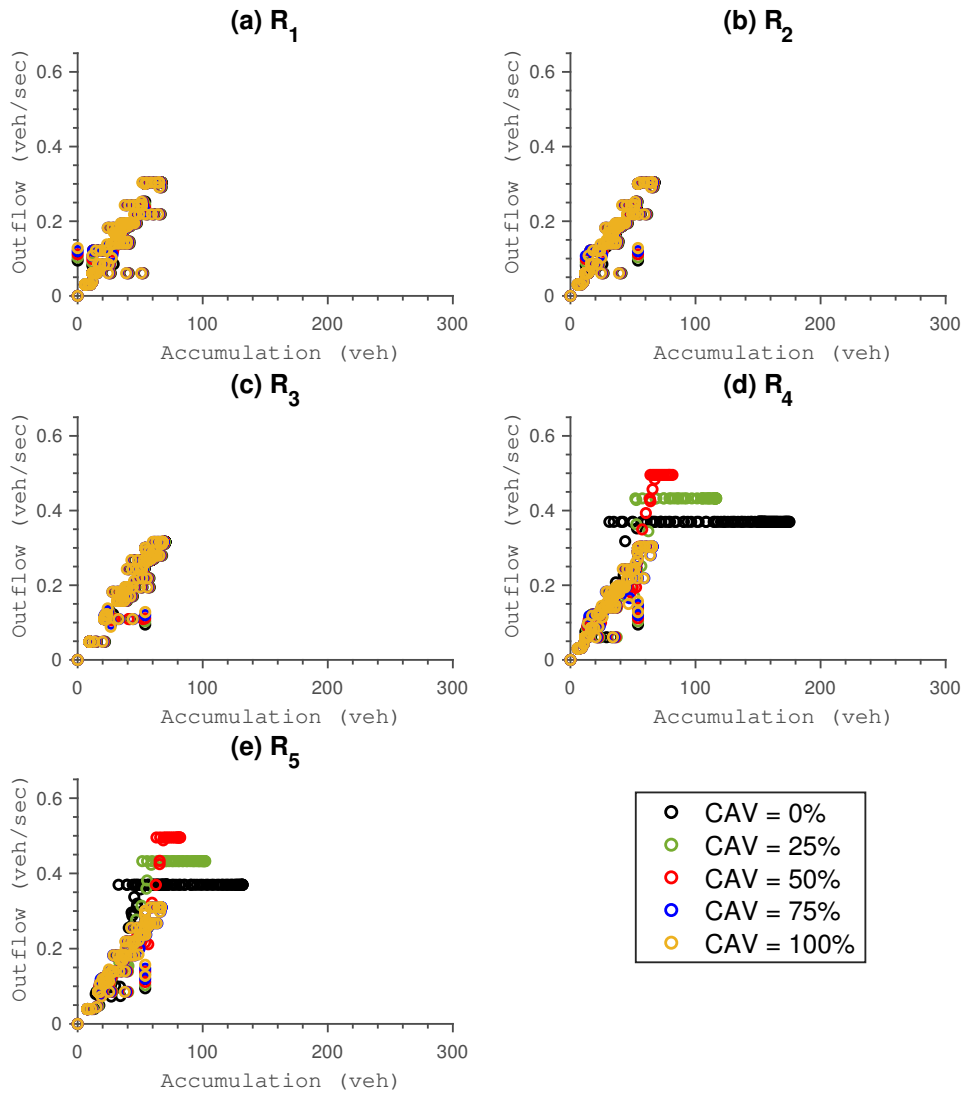


Figure 4.10: Outflow-accumulation relationships in the low demand scenario for different CAV penetration rates (logit route choice parameter $\theta = 0.1$ for conventional vehicles and $\theta = 0.3$ for CAVs) (a) R_1, (b) R_2, (c) R_4

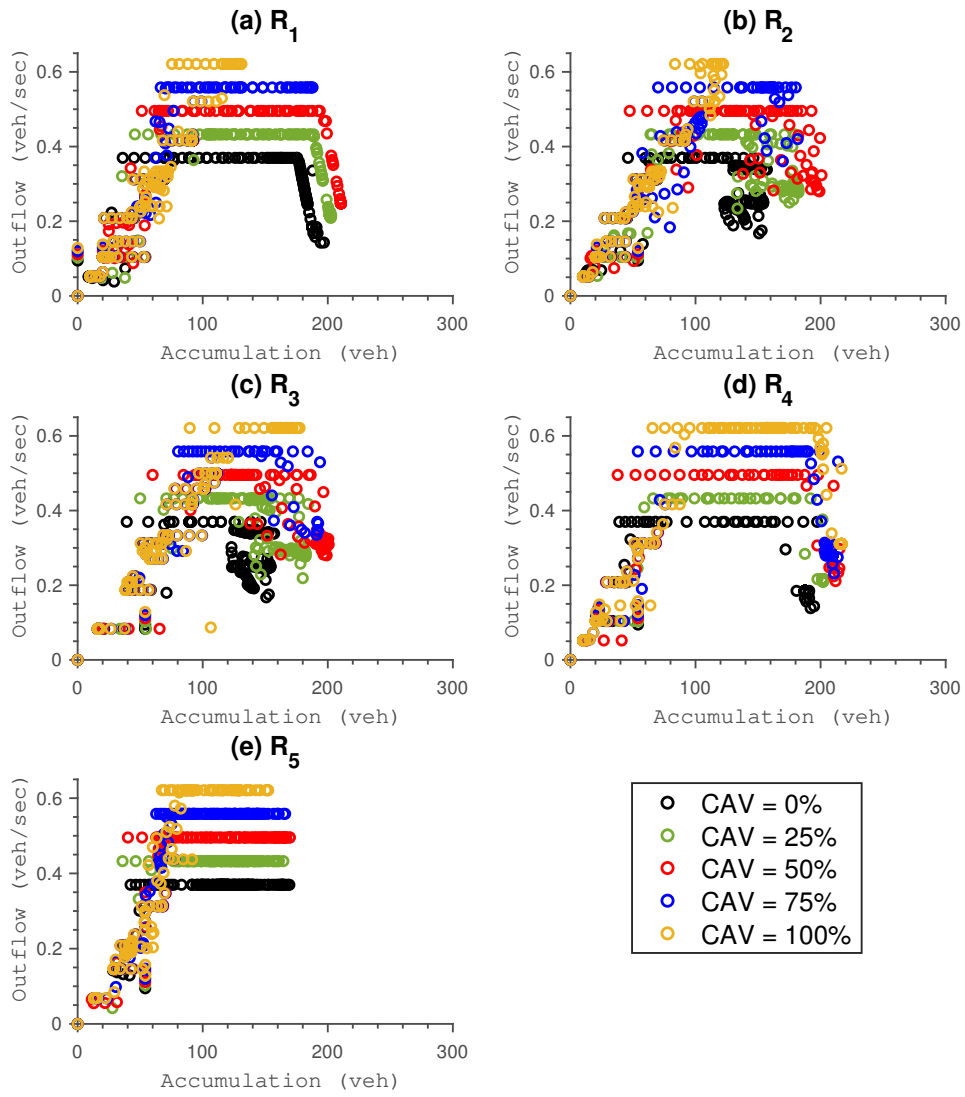


Figure 4.11: Outflow-accumulation relationships in the high demand scenario for different CAV penetration rates (logit route choice parameter $\theta = 0.1$ for conventional vehicles and $\theta = 0.3$ for CAVs) (a) R_1, (b) R_2, (c) R_4

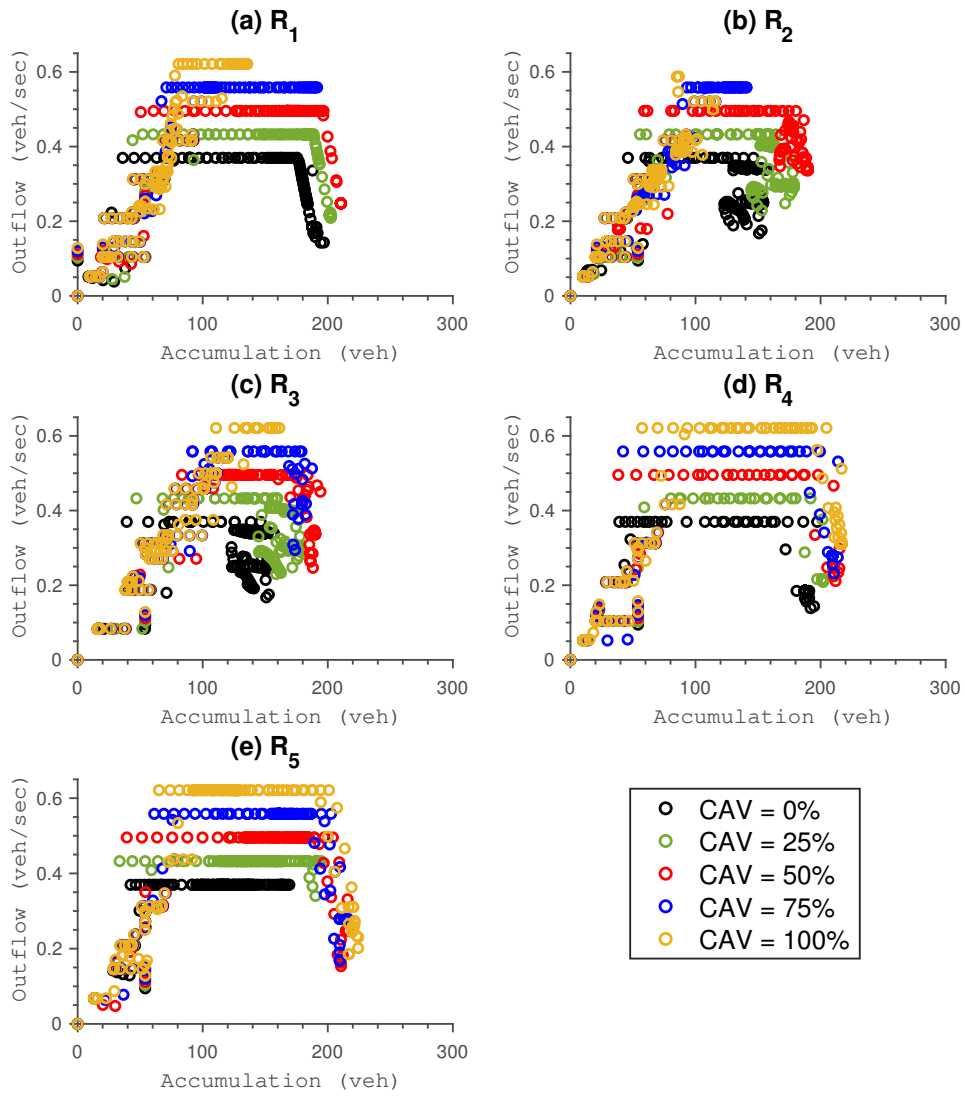
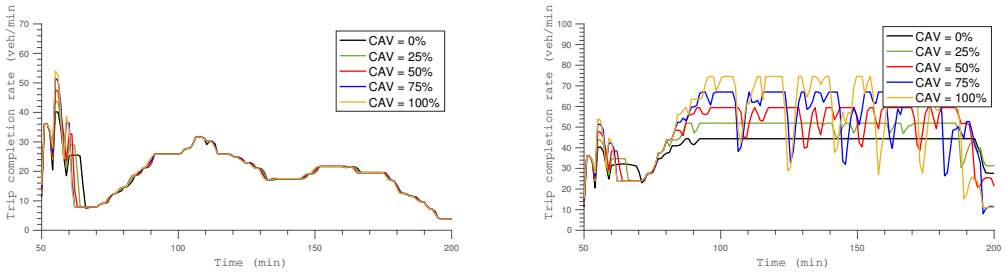


Figure 4.12: Outflow-accumulation relationships in the high demand scenario for different CAV penetration rates (logit route choice parameter $\theta = 0.1$ for both conventional vehicles and CAVs) (a) R_1, (b) R_2, (c) R_4

this analysis. From the figure it is observed that, as the CAV penetration rate increases, no significant change in MFD occurs before the maximum flow point corresponding to MFD with 100% conventional vehicles. When CAV penetration rate is 50% or less, the resulting MFDs for reservoirs 4 and 5 have points in maximum flow traffic conditions; however, as CAV penetration rate increases, reservoirs experience uncongested traffic conditions. While the free flow speeds obtained under various CAV penetration rates remain the same, due to the resulting increase in cost function $R(V)$ as a result of increase in CAV penetration rate, higher outflows for scenarios that have higher CAV penetration rates are expected to achieve. The difference, however, is insignificant. The results indicate that our developed MRDNL model performs exceptionally well in terms of illustrating traffic states in the low demand traffic conditions.

In high demand scenario, as CAV penetration rate increases, higher maximum outflow and throughput for reservoirs are achieved, as can be observed in Figure 4.11. In reservoir 1, when the penetration rate is 50% or less, congested side of the corresponding MFDs are constructed; however, when the penetration rate increases to 75%, no congestion is observed and the reservoir operates in maximum flow traffic conditions. In reservoirs 2 and 3, shift from congested to uncongested traffic conditions occur when all vehicles are CAVs. It is also observed that faster backward speeds for scenarios that have higher CAV penetration rates are achieved (see reservoir 1 to 4). The resulting MFDs are slightly scattered under congested traffic conditions. To examine the effect of information provision on shape and scatter of MFD, simulation is conducted assuming that CAVs have no superiority over conventional vehicles in terms of travel time information (θ=0.1 for both CAVs and conventional vehicles). In other words, this scenario models the impact of automation with no connectivity feature. The MFD plots of Figure 4.12, that are estimated under minimal information provision assumption, are found to be more well-defined and low-scattered compared to that of Figure 4.11, which were slightly scattered, especially in reservoirs 2 and 3. With increase in information provision, route switching behaviours are encouraged; thus, heterogeneity



(a) NC (uniform) (b) NC (radial)

Figure 4.13: Trip completion for the (a) low demand scenario, and (b) high demand scenario.

in spatial distribution of flow, which appears as scatter in MFD plots, increase. The spatial distribution of vehicle density in traffic network can affect the shape and scatter of MFD [12, 85]. Heterogeneous distribution of traffic in congested reservoirs leads to points below the MFD outflow level and thus, creates scatter [16]. Despite the slightly higher level of scatter in high information provision scenario, fewer points in the congested side of MFDs are presented. In reservoir 5, in minimal information provision scenario, congested side of MFD is fully constructed; however, congestion is resolved in high information provision scenario. The results suggest that when higher level of information is provided to vehicles, traffic network performance enhances by producing fewer points in the congested side of its corresponding MFD. In other words, increase in autonomy, enhances traffic condition and shifts the network towards uncongested and maximum flow traffic conditions. The overall performance of traffic network improves in this scenario.

Trip completion for low and high demand scenarios under different market penetration rates are illustrated in Figures 4.13 (a) and 4.13 (b), respectively. The results indicate no noticeable change in network outflow as the CAV market penetration rate increases in low demand traffic conditions while a significant improvement in network outflow is detected in high demand traffic conditions. Table 4.2 shows the total network outflow for any reservoir within the urban network under various CAV market penetration rates and demand scenarios. According to the results, in the low demand

Table 4.2: The outflow and percentage increase in outflows of reservoirs in brackets under mixed conventional and CAV traffic conditions (logit route choice parameter $\theta = 0.1$ for conventional vehicles and $\theta = 0.3$ for CAVs)

		CAV penetration rate				
Demand scenario	Region	0%	25%	50%	75%	100%
Low demand	1	$1.4040e+03$	$1.4040e+03$	$1.4040e+03$	$1.4040e+03$	$1.4040e+03$
		-	(0%)	(0%)	(0%)	(0%)
	2	$1.4040e+03$	$1.4040e+03$	$1.4040e+03$	$1.4040e+03$	$1.4040e+03$
		-	(0%)	(0%)	(0%)	(0%)
	3	$1.8799e+03$	$1.8799e+03$	$1.8799e+03$	$1.8799e+03$	$1.8799e+03$
		-	(0%)	(0%)	(0%)	(0%)
	4	$1.4521e+03$	$1.4521e+03$	$1.4521e+03$	$1.4521e+03$	$1.4521e+03$
		-	(0%)	(0%)	(0%)	(0%)
	5	$1.4586e+03$	$1.5806e+03$	$1.6511e+03$	$1.6883e+03$	$1.6883e+03$
		-	(8.36%)	(13.20%)	(15.75%)	(15.75%)
	6	$1.9180e+03$	$1.7960e+03$	$1.7255e+03$	$1.6883e+03$	$1.6883e+03$
		-	(-6.36%)	(-10.04%)	(-11.98%)	(-11.98%)
	7	$2.9512e+03$	$3.0732e+03$	$3.1437e+03$	$3.1809e+03$	$3.1809e+03$
		-	(4.13%)	(6.52%)	(7.78%)	(7.78%)
	total	$1.2468e+04$	$1.2590e+04$	$1.2660e+04$	$1.2698e+04$	$1.2698e+04$
		-	(0.98%)	(1.54%)	(1.84%)	(1.84%)
High demand	1	$1.6738e+03$	$1.8541e+03$	$1.9948e+03$	$2.1389e+03$	$2.2718e+03$
		-	(10.77%)	(19.18%)	(27.78%)	(35.73%)
	2	$2.1559e+03$	$2.4122e+03$	$2.5461e+03$	$2.5754e+03$	$2.4648e+03$
		-	(11.89%)	(18.10%)	(19.46%)	(14.33%)
	3	$2.3353e+03$	$2.6323e+03$	$2.9383e+03$	$3.0831e+03$	$3.1841e+03$
		-	(12.72%)	(25.82%)	(32.02%)	(36.35%)
	4	$1.6205e+03$	$1.8651e+03$	$2.0387e+03$	$2.1828e+03$	$2.3157e+03$
		-	(15.09%)	(25.81%)	(34.70%)	(42.90%)
	5	$1.6473e+03$	$1.8713e+03$	$2.0490e+03$	$2.1751e+03$	$2.3382e+03$
		-	(13.59%)	(24.39%)	(32.04%)	(41.94%)
	6	$2.8928e+03$	$3.2550e+03$	$3.5172e+03$	$3.5651e+03$	$3.3925e+03$
		-	(12.52%)	(21.58%)	(32.24%)	(17.27%)
	7	$3.2325e+03$	$3.7036e+03$	$4.1188e+03$	$4.3888e+03$	$4.6847e+03$
		-	(14.57%)	(27.42%)	(35.77%)	(44.93%)
	total	$1.5558e+04$	$1.7594e+04$	$1.9203+04$	$2.0109e+04$	$2.0652e+04$
		-	(13.09%)	(23.43%)	(29.25%)	(32.74%)

scenario, increasing the CAV penetration rate does not enhance the total outflow of a network; improvement is less than 2% for a CAV penetration rate of 75%, while no further improvement above this percentage is observed. In the high demand scenario, the outflow of the reservoirs increases as the CAV penetration rate increases; therefore, increasing the CAV penetration rate has a positive effect on a networks' total outflow in a high demand scenario. For CAV penetration rates of 25%, 50%, 75%, and 100%, outflow increases by 13.09%, 23.43%, 29.25%, and 32.74%, respectively. The results from this section highlight the positive role CAVs play in enhancing network flow capacity and preventing congested traffic conditions, especially in high demand scenarios. CAVs have a positive effect on congestion alleviation and they can be adopted to enhance the congested traffic condition.

4.7 Conclusion

This chapter developed a new multi-reservoir dynamic network loading (MRDNL) approach to model more closely traffic dynamics in large-scale urban networks composed of several reservoirs. Based on interactions between hyper-link and hyper-node models, I further studied the proper wave propagation in an urban reservoir and closely estimated the resulting inner-reservoir travel time. These hyper-link and hyper-node models are continually interacting through the conventional LTM framework. The hyper-link model, which is based on a piece-wise linear MFD, introduces a set of constraints based on VT to compute upper bounds for the upstream and downstream cumulative accumulations; thus, adequately capturing multiple existing kinematic waves. While considering the total receiving capacity and the competing demand, the new hyper-node model correctly estimates the transferring flow between neighbouring reservoirs. Unlike existing accumulation-based MFD models that often ignore inner-reservoir travel time, our MRDNL captures time delay in vehicle propagation between the two boundaries of a reservoir and accordingly provides an appropriate estimation of traffic flow. Based on

this traffic network representation, the viscosity issue observed in earlier accumulation-based MFD models is thus prevented. Another significant contribution of this approach is its ability to incorporate CAVs into urban traffic networks to provide insights into their network-wide effect.

The numerical results suggest that the new approach successfully discerns the evolution of flow within reservoirs and captures multiple forward and backward kinematic waves under various demand scenarios. With its avoidance of space discretization requirements, our MRDNL model significantly suppresses large-scale urban networks' computational tractability issues. Therefore, the MRDNL model improves both computational efficiency and realism compared to the previous models. Also, the numerical analysis for incorporating CAVs into the traffic model successfully evaluated the performance of the large-scale urban network in a mixed traffic environment composed of different combinations of CAVs and conventional vehicles. As expected, the presence of CAVs is shown to modify the shape and scatter of MFD. Overall, increasing the CAV penetration rate results in improved network outflow. The results also indicate that when more precise information is shared with CAVs, the congestion level reduces in some reservoirs, while the MFD scatter might increase in others due to frequent rerouting. However, more research based on microsimulation analysis is needed to confirm these findings.

For future research, MRDNL can be extended to include more complex urban network structures with multiple parallel or crossing arterials and intersections that allow turn movements. The current framework considers links with identical fundamental diagrams and is thus not directly applicable to heterogeneous traffic networks with irregular topology and specifications.

Chapter 5

Conclusions and Recommendations

This final chapter summarizes the major achievements of this thesis and offers an outlook on issues that need to be addressed in MFD-based modelling and control of large-scale urban networks. The current study better represents MFD-based traffic networks by capturing important phenomena of traffic networks, such as multiple kinematic waves, queueing, and precise estimations of inner-reservoir travel time. Rigorous modelling of traffic networks has major implications on the efficacy of the developed control models. Further, the crucial role of road users is considered in the models, and anticipatory and fair control approaches are developed with this role in mind. Understanding the effect of CAVs on the shape and scatter of MFD, while suggesting a fairly easy approach for incorporating CAVs into the MFD dynamics and without modelling at the micro-level, is another major achievement of this thesis.

Section 5.1, 5.2, and 5.3 share an overall summary and the key findings of chapters 2, 3, and 4. Section 5.4 suggests directions and recommendations for future research.

5.1 Research contributions and findings on the network-wide anticipatory control of an urban network using macroscopic fundamental diagram

Chapter 2 presents an anticipative control approach for a network-wide control of an urban network using a macroscopic fundamental diagram. The control scheme consists of two levels of control and route choice and is modeled in an MPC framework to

proactively respond to congested traffic conditions. This research contributes to the previous body of knowledge as follows:

- *Modelling a network-wide anticipatory control scheme by incorporating stochastic user equilibrium as a lower level into the control model:* The developed anticipatory control scheme incorporates road users' routing behaviour into the control framework rather than as an input to the model. The developed control solutions are global solutions to the model. The results indicate improvement in overall efficiency of the network measured in terms of TTS. Further, the network moves toward the system optimum traffic conditions, which is used as a benchmark of an ideal control strategy with best optimality. Control strategies developed without consideration of road users yield sub-optimal solutions and force strict control instructions on the network.

- *Modelling a region-based stochastic user equilibrium route choice model:* The stochastic route choice model developed at the regional level illustrates road users route choice behaviour in response to perimeter control instructions inserted into the network and the resulting regional travel time and waiting time. Such solutions are modelled considering road users' perception of travel time captured using a logit route choice model and is solved using an MSA algorithm.

5.2 Research contributions and findings on the fairness-aware and efficient large-scale urban network control: A macroscopic fundamental diagram approach

Chapter 3 introduces a proportionally fair perimeter control scheme for a multi-region urban network based on MFD. The control scheme, modelled in an MPC framework, meters the flow transferring between different neighbour regions while considering the balanced distribution of average speed and queue. Chapter 3 contributes to the previous

body of knowledge as follows:

- *Introducing the concept of fairness for control of a large-scale urban network:* The proportional fairness concept is adopted for control of a large-scale urban network. A novel utility measure based on a balanced distribution of average regional speed and queue among regions is introduced. Performance of the developed model indicates increased fairness without sacrificing efficiency.

5.3 Research contributions and findings on the macroscopic dynamic network loading model using variational theory in a connected and autonomous vehicle environment

In chapter 4, a multi-reservoir dynamic network loading (MRDNL) scheme that consists of a hyper-link and a hyper-node model is developed. The developed MRDNL more closely models traffic dynamics in a large-scale urban network that is composed of several reservoirs. Using variational theory, the LWR model is incorporated into MFD dynamics to enhance traffic propagation at reservoirs. The hyper-link model describes the evolution of vehicles inside the reservoirs, while the hyper-node model transfers the flow to other parts of the network. The hyper-link and hyper-node models are related through the conventional link transmission model (LTM). Further, this chapter incorporates CAVs into MFD dynamics to model an urban network under mixed traffic conditions. This chapter contributes to the previous body of knowledge as follows:

- *Enhancing MFD to capture multiple forward and backward kinematic waves and presenting a precise estimation of inner-regional travel time:* The LWR kinematic wave model is integrated into MFD dynamics to capture multiple forward and backward kinematic waves in an urban network in the form of a hyper-link that has a piece-wise linear MFD. Using variational theory, a set of constraints are developed that construct an upper bound for the cumulative accumulation val-

ues at the upstream and downstream ends of reservoirs. Two algorithms are developed to solve the forward and backward wave models. Unlike the existing accumulation-based MFD models that often ignore inner-reservoir travel time, our MRDNL captures time delay in vehicle propagation between the two boundaries of a reservoir and, accordingly, provides an appropriate estimation of traffic flow. Based on this traffic network representation, the viscosity issue observed in earlier accumulation-based MFD models is thus prevented.

- *Comparing the performance of the developed MRDNL to that of CTM*: To validate the assumption that MRDNL is able to model traffic network precisely, the performance of the developed model is compared to that of CTM. The results indicate that the developed model performs well in terms of propagating traffic within a network, and it requires less computational effort compared to previous models discussed in the literature. Our MRDNL model significantly suppresses large-scale urban networks' computational tractability issues because it avoids space discretization requirements.

- *Developing a simple framework for integrating CAVs into an MFD framework and evaluating network performance under mixed traffic conditions:* The developed model can easily integrate CAVs into MFD traffic dynamics and provides insight into the network-wide effect of CAVs in traffic networks. Increasing CAV penetration rates improves network outflow and production. In addition, the effect of the information accessible to CAVs on networks' traffic conditions is investigated. We found that when more precise information is shared with CAVs, congestion levels decrease in some reservoirs, while the MFD scatter might increase in other reservoirs due to frequent rerouting.

5.4 Recommendations for future work

There are several possible future directions for this research.

The queue resulting from implementing perimeter control can adversely affect the existence of a well-defined MFD. The spatio-temporal dynamics of queues and their effect on maximum outflow and heterogeneity of the MFD-based network have to be investigated and included in the control framework. The type of queue considered in chapters 2 and 3 of this thesis is a point queue model that does not occupy any space and does not consider spillback. Although heterogeneity in travel production resulting from queuing was modelled in chapter 3, this issue needs further investigation by considering physical queues and spillback effect in an MFD-based traffic network.

Another potential extension is to extend the developed MRDNL in chapter 4 to describe more complex urban network structures that have multiple parallel or crossing arterials and intersections that allow turn movements. The current framework considers links with identical fundamental diagrams and is thus not directly applicable to heterogeneous traffic networks with irregular topology and specifications.

The proportional fairness concept introduced in chapter 3 can be adopted to develop other control strategies, such as route guidance. This advisory control scheme heavily relies on road users' compliance to the routing instructions; a fair perimeter control strategy that considers road users' utility is expected to increase compliance to the routing instructions and improve a network's traffic conditions. In addition, multi-level control schemes that target different goals or different regional and sub-regional levels can be developed using MFD concept.

The limitation of this dissertation is unavailability of real traffic data to test the developed control strategies. The applicability of current control strategies can be tested using real traffic networks. Also, data driven approaches can be adopted for enhancing travel time estimation and traffic state prediction.

References

[1] N. Geroliminis and C. F. Daganzo, "Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings," *Transportation Research Part B: Methodological*, vol. 42, no. 9, pp. 759–770, 2008.

[2] Y. Ji and N. Geroliminis, "On the spatial partitioning of urban transportation networks," *Transportation Research Part B: Methodological*, vol. 46, no. 10, pp. 1639–1656, 2012.

[3] C. F. Daganzo, V. V. Gayah, and E. J. Gonzales, "The potential of parsimonious models for understanding large scale transportation systems and answering big picture questions," *EURO Journal on Transportation and Logistics*, vol. 1, no. 1-2, pp. 47–65, 2012.

[4] C. F. Daganzo, "Urban gridlock: Macroscopic modeling and mitigation approaches," *Transportation Research Part B: Methodological*, vol. 41, no. 1, pp. 49–62, 2007.

[5] J. Godfrey, "The mechanism of a road network," *Traffic Engineering & Control*, vol. 8, no. 8, 1969.

[6] C. F. Daganzo and N. Geroliminis, "An analytical approximation for the macroscopic fundamental diagram of urban traffic," *Transportation Research Part B: Methodological*, vol. 42, no. 9, pp. 771–781, 2008.

[7] C. F. Daganzo, "A variational formulation of kinematic waves: basic theory and complex boundary conditions," *Transportation Research Part B: Methodological*, vol. 39, no. 2, pp. 187–196, 2005.

[8] C. F. Daganzo, "A variational formulation of kinematic waves: Solution methods," *Transportation Research Part B: Methodological*, vol. 39, no. 10, pp. 934–950, 2005.

[9] N. Geroliminis and B. Boyacı, "The effect of variability of urban systems characteristics in the network capacity," *Transportation Research Part B: Methodological*, vol. 46, no. 10, pp. 1607–1623, 2012.

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

[11] C. Buisson and C. Ladier, "Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams," *Transportation Research Record*, vol. 2124, no. 1, pp. 127–136, 2009.

[12] A. Mazloumian, N. Geroliminis, and D. Helbing, "The spatial variability of vehicle densities as determinant of urban network capacity," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 368, no. 1928, pp. 4627–4647, 2010.

[13] H. S. Mahmassani, M. Saberi, *et al.*, "Urban network gridlock: Theory, characteristics, and dynamics," *Procedia-Social and Behavioral Sciences*, vol. 80, pp. 79–98, 2013.

[14] J. Haddad and N. Geroliminis, "On the stability of traffic perimeter control in two-region urban cities," *Transportation Research Part B: Methodological*, vol. 46, no. 9, pp. 1159–1176, 2012.

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

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

[17] M. Hajiahmadi, V. L. Knoop, B. De Schutter, and H. Hellendoorn, "Optimal dynamic route guidance: A model predictive approach using the macroscopic fundamental diagram," in *Intelligent Transportation Systems-(ITSC), 2013 16th International IEEE Conference on*, pp. 1022–1028, IEEE, 2013.

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

[19] R. Arnott, "A bathtub model of downtown traffic congestion," *Journal of Urban Economics*, vol. 76, pp. 110–121, 2013.

[20] M. Fosgerau, "Congestion in the bathtub," *Economics of Transportation*, vol. 4, no. 4, pp. 241–255, 2015.

[21] R. Arnott, A. Kokoza, and M. Naji, "Equilibrium traffic dynamics in a bathtub model: A special case," *Economics of transportation*, vol. 7, pp. 38–52, 2016.

[22] R. Arnott and J. Buli, "Solving for equilibrium in the basic bathtub model," *Transportation Research Part B: Methodological*, vol. 109, pp. 150–175, 2018.

[23] G. Mariotte, L. Leclercq, and J. A. Laval, "Macroscopic urban dynamics: Analytical and numerical comparisons of existing models," *Transportation Research Part B: Methodological*, vol. 101, pp. 245–267, 2017.

[24] M. Yildirimoglu, M. Ramezani Ghalenoei, and N. Geroliminis, "Equilibrium analysis and route guidance in large-scale networks with mfd dynamics," *Transportation Research Part C: Emerging Technologies*, vol. 59, no. ARTICLE, pp. 404–420, 2015.

[25] J. A. Laval, L. Leclercq, and N. Chiabaut, "Minimal parameter formulations of the dynamic user equilibrium using macroscopic urban models: Freeway vs city streets

revisited," *Transportation Research Part B: Methodological*, vol. 117, pp. 676–686, 2018.

[26] Y. Huang, J. Xiong, A. Sumalee, N. Zheng, W. Lam, Z. He, and R. Zhong, "A dynamic user equilibrium model for multi-region macroscopic fundamental diagram systems with time-varying delays," *Transportation Research Part B: Methodological*, vol. 131, pp. 1–25, 2020.

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

[28] J. Haddad, "Optimal coupled and decoupled perimeter control in one-region cities," *Control Engineering Practice*, vol. 61, pp. 134–148, 2017.

[29] M. Keyvan-Ekbatani, A. Kouvelas, I. Papamichail, and M. Papageorgiou, "Exploiting the fundamental diagram of urban networks for feedback-based gating," *Transportation Research Part B: Methodological*, vol. 46, no. 10, pp. 1393–1403, 2012.

[30] N. Geroliminis, J. Haddad, and M. Ramezani, "Optimal perimeter control for two urban regions with macroscopic fundamental diagrams: A model predictive approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 348–359, 2012.

[31] M. Hajiahmadi, J. Haddad, B. De Schutter, and N. Geroliminis, "Optimal hybrid perimeter and switching plans control for urban traffic networks," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 2, pp. 464–478, 2014.

[32] I. I. Sirmatel and N. Geroliminis, "Economic model predictive control of large-scale urban road networks via perimeter control and regional route guidance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 4, pp. 1112–1121, 2017.

[33] A. Aalipour, H. Kebriaei, and M. Ramezani, "Analytical optimal solution of perimeter traffic flow control based on mfd dynamics: a pontryagin's maximum principle approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 9, pp. 3224–3234, 2018.

[34] J. Haddad and B. Mirkin, "Adaptive perimeter traffic control of urban road networks based on mfd model with time delays," *International Journal of Robust and Nonlinear Control*, vol. 26, no. 6, pp. 1267–1285, 2016.

[35] C. F. Daganzo, V. V. Gayah, and E. J. Gonzales, "Macroscopic relations of urban traffic variables: Bifurcations, multivaluedness and instability," *Transportation Research Part B: Methodological*, vol. 45, no. 1, pp. 278–288, 2011.

[36] L. Leclercq, C. Parzani, V. L. Knoop, J. Amourette, and S. P. Hoogendoorn, "Macroscopic traffic dynamics with heterogeneous route patterns," *Transportation Research Part C: Emerging Technologies*, vol. 59, pp. 292–307, 2015.

[37] M. Papageorgiou, C. Diakaki, V. Dinopoulou, A. Kotsialos, and Y. Wang, "Review of road traffic control strategies," *Proceedings of the IEEE*, vol. 91, no. 12, pp. 2043–2067, 2003.

[38] X. Zhang and H. Yang, "The optimal cordon-based network congestion pricing problem," *Transportation Research Part B: Methodological*, vol. 38, no. 6, pp. 517–537, 2004.

[39] H. Taale, "Integrated anticipatory control of road networks: A game-theoretical approach," 2008.

[40] K. Han, Y. Sun, H. Liu, T. L. Friesz, and T. Yao, "A bi-level model of dynamic traffic signal control with continuum approximation," *Transportation Research Part C: Emerging Technologies*, vol. 55, pp. 409–431, 2015.

[41] M. Rinaldi and C. M. Tampère, "An extended coordinate descent method for distributed anticipatory network traffic control," *Transportation Research Part B: Methodological*, vol. 80, pp. 107–131, 2015.

[42] B. Zhou, M. Bliemer, H. Yang, and J. He, "A trial-and-error congestion pricing scheme for networks with elastic demand and link capacity constraints," *Transportation Research Part B: Methodological*, vol. 72, pp. 77–92, 2015.

[43] M. Rinaldi, C. M. Tampère, and F. Viti, "On characterizing the relationship between route choice behaviour and optimal traffic control solution space," *Transportation research part B: methodological*, vol. 117, pp. 892–906, 2018.

[44] W. Huang, F. Viti, and C. M. Tampère, "An iterative learning approach for anticipatory traffic signal control on urban networks," *Transportmetrica B: Transport Dynamics*, vol. 5, no. 4, pp. 402–425, 2017.

[45] R. E. Allsop, "Some possibilities for using traffic control to influence trip distribution and route choice," in *Transportation and traffic theory, proceedings*, vol. 6, 1974.

[46] M. Smith, "Traffic control and traffic assignment in a signal-controlled network with queueing," *Transportation and Traffic Theory*, 1987.

[47] H. Yang and S. Yagar, "Traffic assignment and signal control in saturated road networks," *Transportation Research Part A: Policy and Practice*, vol. 29, no. 2, pp. 125–139, 1995.

[48] H. Yan and W. H. Lam, "Optimal road tolls under conditions of queueing and congestion," *Transportation Research Part A: Policy and Practice*, vol. 30, no. 5, pp. 319–332, 1996.

[49] G. E. Cantarella, G. Improta, and A. Sforza, "Iterative procedure for equilibrium

network traffic signal setting," *Transportation Research Part A: General*, vol. 25, no. 5, pp. 241–249, 1991.

[50] C. Meneguzzer, "Review of models combining traffic assignment and signal control," *Journal of Transportation Engineering*, vol. 123, no. 2, pp. 148–155, 1997.

[51] H. J. van Zuylen and H. Taale, "Anticipatory optimization of traffic control," *Transportation research record*, vol. 1725, no. 1, pp. 109–115, 2000.

[52] M. Saberi, H. S. Mahmassani, and A. Zockaie, "Network capacity, traffic instability, and adaptive driving: findings from simulated urban network experiments," *EURO Journal on Transportation and Logistics*, vol. 3, no. 3-4, pp. 289–308, 2014.

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

[54] K. Ampountolas, N. Zheng, and N. Geroliminis, "Macroscopic modelling and robust control of bi-modal multi-region urban road networks," *Transportation Research Part B: Methodological*, vol. 104, pp. 616–637, 2017.

[55] M. Yildirimoglu, I. I. Sirmatel, and N. Geroliminis, "Hierarchical control of heterogeneous large-scale urban road networks via path assignment and regional route guidance," *Transportation Research Part B: Methodological*, vol. 118, pp. 106–123, 2018.

[56] M. Yildirimoglu, M. Ramezani, *et al.*, "Demand management with limited cooperation among travellers: A doubly dynamic approach," *Transportation Research Part B: Methodological*, vol. 132, no. C, pp. 267–284, 2020.

[57] D. Ingole, G. Mariotte, and L. Leclercq, "Perimeter gating control and citywide dynamic user equilibrium: a macroscopic modeling framework," *Transportation research part C: emerging technologies*, vol. 111, pp. 22–49, 2020.

[58] P. Bonsall, "Modelling response to information systems and other intelligent transport system innovations," in *Handbook of Transport Modelling*, Emerald Group Publishing Limited, 2007.

[59] C. E. Garcia, D. M. Prett, and M. Morari, "Model predictive control: Theory and practice—a survey," *Automatica*, vol. 25, no. 3, pp. 335–348, 1989.

[60] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. Scokaert, "Constrained model predictive control: Stability and optimality," *Automatica*, vol. 36, no. 6, pp. 789–814, 2000.

[61] G. Mariotte, L. Leclercq, *et al.*, "Flow exchanges in multi-reservoir systems with spillbacks," *Transportation Research Part B: Methodological*, vol. 122, no. C, pp. 327–349, 2019.

[62] V. L. Knoop and S. P. Hoogendoorn, "Network transmission model: a dynamic traffic model at network level," in *Proceedings of the 93rd Annual Meeting Transportation Research Board*, 2014.

[63] A. Kouvelas, M. Saeedmanesh, and N. Geroliminis, "Enhancing model-based feedback perimeter control with data-driven online adaptive optimization," *Transportation Research Part B: Methodological*, vol. 96, pp. 26–45, 2017.

[64] H. X. Liu, X. He, and B. He, "Method of successive weighted averages (mswa) and self-regulated averaging schemes for solving stochastic user equilibrium problem," *Networks and Spatial Economics*, vol. 9, no. 4, pp. 485–503, 2009.

[65] W. B. Powell and Y. Sheffi, "The convergence of equilibrium algorithms with predetermined step sizes," *Transportation Science*, vol. 16, no. 1, pp. 45–55, 1982.

[66] Y. Sheffi, *Urban transportation networks*, vol. 6. Prentice-Hall, Englewood Cliffs, NJ, 1985.

[67] I. I. Sirmatel and N. Geroliminis, "Economic model predictive control of large-scale urban road networks via perimeter control and regional route guidance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 4, pp. 1112–1121, 2018.

[68] T. J. Dickson, "A note on traffic assignment and signal timings in a signal-controlled road network," *Transportation Research Part B: Methodological*, vol. 15, no. 4, pp. 267–271, 1981.

[69] J. Haddad and Z. Zheng, "Adaptive perimeter control for multi-region accumulation-based models with state delays," *Transportation Research Part B: Methodological*, 2018.

[70] K. Yang, N. Zheng, and M. Menendez, "Multi-scale perimeter control approach in a connected-vehicle environment," *Transportation research procedia*, vol. 23, pp. 101–120, 2017.

[71] N. Moshahedi, L. Kattan, and R. Tay, "A network-wide anticipatory control of an urban network using macroscopic fundamental diagram," *Transportmetrica B: Transport Dynamics*, vol. 9, no. 1, pp. 415–436, 2021.

[72] L. Zhang and D. Levinson, "Balancing efficiency and equity of ramp meters," *Journal of transportation engineering*, vol. 131, no. 6, pp. 477–481, 2005.

[73] L. Zhang and D. Levinson, "Ramp metering and freeway bottleneck capacity," *Transportation Research Part A: Policy and Practice*, vol. 44, no. 4, pp. 218–235, 2010.

[74] G. Zhang and Y. Wang, "Optimizing coordinated ramp metering: A preemptive hierarchical control approach," *Computer-Aided Civil and Infrastructure Engineering*, vol. 28, no. 1, pp. 22–37, 2013.

[75] A. Kotsialos and M. Papageorgiou, "Efficiency and equity properties of freeway network-wide ramp metering with amoc," *Transportation Research Part C: Emerging Technologies*, vol. 12, no. 6, pp. 401–420, 2004.

[76] M. Papageorgiou and A. Kotsialos, "Efficiency versus fairness in network-wide ramp metering," in *Proceedings of the IEEE Intelligent Transportation Systems Conference*, pp. 1189–1194, 2001.

[77] D. Levinson, L. Zhang, S. Das, and A. Sheikh, "Evaluating ramp meters: evidence from the twin cities ramp meter shut-off," in *81st TRB Annual meeting, Wahsington, DC. Under Review Transportation Research Part C*, 2002.

[78] Y. Yin, H. Liu, and H. Benouar, "A note on equity of ramp metering," in *Proceedings. The 7th International IEEE Conference on Intelligent Transportation Systems (IEEE Cat. No. 04TH8749)*, pp. 497–502, IEEE, 2004.

[79] Q. Tian, H.-J. Huang, H. Yang, and Z. Gao, "Efficiency and equity of ramp control and capacity allocation mechanisms in a freeway corridor," *Transportation Research Part C: Emerging Technologies*, vol. 20, no. 1, pp. 126–143, 2012.

[80] F. Kelly, "Charging and rate control for elastic traffic," *European transactions on Telecommunications*, vol. 8, no. 1, pp. 33–37, 1997.

[81] S. Aalami and L. Kattan, "Fair dynamic resource allocation in transit-based evacuation planning," *Transportation research part C: emerging technologies*, vol. 94, pp. 307–322, 2018.

[82] S. Aalami and L. Kattan, "Fair transit trip planning in emergency evacuations: A combinatorial approach," *Transportation Research Part C: Emerging Technologies*, vol. 122, p. 102760, 2021.

[83] J. Nash and F. John, "The bargaining problem," *Econometrica: Journal of the econometric society*, pp. 155–162, 1950.

[84] G. Mariotte and L. Leclercq, "Flow exchanges in multi-reservoir systems with spillbacks," *Transportation Research Part B: Methodological*, vol. 122, pp. 327–349, 2019.

[85] N. Geroliminis and J. Sun, "Properties of a well-defined macroscopic fundamental diagram for urban traffic," *Transportation Research Part B: Methodological*, vol. 45, no. 3, pp. 605–617, 2011.

[86] J. Haddad, "Optimal perimeter control synthesis for two urban regions with aggregate boundary queue dynamics," *Transportation Research Part B: Methodological*, vol. 96, pp. 1–25, 2017.

[87] M. Amirgholy and H. O. Gao, "Modeling the dynamics of congestion in large urban networks using the macroscopic fundamental diagram: User equilibrium, system optimum, and pricing strategies," *Transportation Research Part B: Methodological*, vol. 104, pp. 215–237, 2017.

[88] Q. Ge, D. Fukuda, K. Han, and W. Song, "Reservoir-based surrogate modeling of dynamic user equilibrium," *Transportation Research Part C: Emerging Technologies*, vol. 113, pp. 350–369, 2020.

[89] W. Ni and M. Cassidy, "City-wide traffic control: modeling impacts of cordon queues," *Transportation research part C: emerging technologies*, vol. 113, pp. 164–175, 2020.

[90] M. J. Lighthill and G. B. Whitham, "On kinematic waves ii. a theory of traffic flow on long crowded roads," *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 229, no. 1178, pp. 317–345, 1955.

[91] P. I. Richards, "Shock waves on the highway," *Operations Research*, vol. 4, no. 1, pp. 42–51, 1956.

[92] Q. Ge and D. Fukuda, "A macroscopic dynamic network loading model for multiple-reservoir system," *Transportation Research Part B: Methodological*, vol. 126, pp. 502–527, 2019.

[93] M. Paipuri and L. Leclercq, "Bi-modal macroscopic traffic dynamics in a single region," *Transportation Research Part B: Methodological*, vol. 133, pp. 257–290, 2020.

[94] C. F. Daganzo, "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transportation Research Part B: Methodological*, vol. 28, no. 4, pp. 269–287, 1994.

[95] C. F. Daganzo, "The cell transmission model, part ii: network traffic," *Transportation Research Part B: Methodological*, vol. 29, no. 2, pp. 79–93, 1995.

[96] S. Godunov, "A finite difference method for the computation of discontinuous solutions of the equations of fluid dynamics.," *Sbornik: Mathematics*, vol. 47, no. 8-9, pp. 357–393, 1959.

[97] P.-E. Mazaré, A. H. Dehwah, C. G. Claudel, and A. M. Bayen, "Analytical and grid-free solutions to the lighthill–whitham–richards traffic flow model," *Transportation Research Part B: Methodological*, vol. 45, no. 10, pp. 1727–1748, 2011.

[98] J. P. van der Gun, A. J. Pel, and B. van Arem, "Extending the link transmission model with non-triangular fundamental diagrams and capacity drops," *Transportation Research Part B: Methodological*, vol. 98, pp. 154–178, 2017.

[99] M. C. Bliemer and M. P. Raadsen, "Continuous-time general link transmission model with simplified fanning, part i: Theory and link model formulation," *Transportation Research Part B: Methodological*, vol. 126, pp. 442–470, 2019.

[100] Q. Lu, T. Tettamanti, D. Hörcher, and I. Varga, "The impact of autonomous vehicles on urban traffic network capacity: an experimental analysis by microscopic traffic simulation," *Transportation Letters*, vol. 12, no. 8, pp. 540–549, 2020.

[101] C. L. Melson, M. W. Levin, B. E. Hammit, and S. D. Boyles, "Dynamic traffic assignment of cooperative adaptive cruise control," *Transportation Research Part C: Emerging Technologies*, vol. 90, pp. 114–133, 2018.

[102] C. M. Tampère, R. Corthout, D. Cattrysse, and L. H. Immers, "A generic class of first order node models for dynamic macroscopic simulation of traffic flows," *Transportation Research Part B: Methodological*, vol. 45, no. 1, pp. 289–309, 2011.

[103] E. Hans, N. Chiabaut, and L. Leclercq, "Applying variational theory to travel time estimation on urban arterials," *Transportation Research Part B: Methodological*, vol. 78, pp. 169–181, 2015.

[104] G. F. Newell, "A simplified theory of kinematic waves in highway traffic, part i: General theory," *Transportation Research Part B: Methodological*, vol. 27, no. 4, pp. 281–287, 1993.

[105] I. Yperman, S. Logghe, and B. Immers, "The link transmission model: An efficient implementation of the kinematic wave theory in traffic networks," in *Proceedings of the 10th EWGT Meeting*, pp. 122–127, Poznan Poland, 2005.

[106] A. F. Lentzakis, S. I. Ware, and R. Su, "Region-based dynamic forecast routing for autonomous vehicles," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1464–1469, IEEE, 2016.

[107] W. Jin and H. M. Zhang, "On the distribution schemes for determining flows through a merge," *Transportation Research Part B: Methodological*, vol. 37, no. 6, pp. 521–540, 2003.

[108] D. Ni and J. D. Leonard II, "A simplified kinematic wave model at a merge bottleneck," *Applied Mathematical Modelling*, vol. 29, no. 11, pp. 1054–1072, 2005.

[109] N. Yu, J. Ma, and H. M. Zhang, "A polymorphic dynamic network loading model," *Computer-Aided Civil and Infrastructure Engineering*, vol. 23, no. 2, pp. 86–103, 2008.

[110] C. F. Daganzo, "On the variational theory of traffic flow: well-posedness, duality and applications," *Networks & Heterogeneous Media*, vol. 1, no. 4, pp. 601–619, 2006.

[111] G. Gentile, L. Meschini, and N. Papola, "Macroscopic arc performance models with capacity constraints for within-day dynamic traffic assignment," *Transportation Research Part B: Methodological*, vol. 39, no. 4, pp. 319–338, 2005.

[112] G. Gentile *et al.*, "The general link transmission model for dynamic network loading and a comparison with the due algorithm," *New Developments in Transport Planning: Advances in Dynamic Traffic Assignment*, vol. 178, p. 153, 2010.

[113] G. Bellei, G. , and N. Papola, "A within-day dynamic traffic assignment model for urban road networks," *Transportation Research Part B: Methodological*, vol. 39, no. 1, pp. 1–29, 2005.

[114] W. Himpe, R. Corthout, and M. C. Tampère, "An efficient iterative link transmission model," *Transportation Research Part B: Methodological*, vol. 92, pp. 170–190, 2016.

[115] I. Yperman, S. Logghe, C. M. Tampere, and B. Immers, "Multicommodity link transmission model for dynamic network loading," tech. rep., 2006.

[116] N. Geroliminis, N. Zheng, and K. Ampountolas, "A three-dimensional macroscopic fundamental diagram for mixed bi-modal urban networks," *Transportation Research Part C: Emerging Technologies*, vol. 42, pp. 168–181, 2014.

[117] L. Kattan, M. Moussavi, B. Far, C. Harschnitz, A. Radmanesh, and S. Saidi, "Evaluating the potential benefits of vehicle to vehicle communication (v2v) under incident conditions in the paramics model," in *Proceedings of the 13th International IEEE Conference on Intelligent Transportation, Madeira, Portugal*, vol. 9, 2010.

[118] E. Paikari, L. Kattan, S. Tahmasseby, and B. H. Far, "Modeling and simulation of advisory speed and re-routing strategies in connected vehicles systems for crash risk and travel time reduction," in *2013 26th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, pp. 1–4, IEEE, 2013.

[119] M. W. Levin and S. D. Boyles, "A multiclass cell transmission model for shared human and autonomous vehicle roads," *Transportation Research Part C: Emerging Technologies*, vol. 62, pp. 103–116, 2016.

Appendix A

Copyright permission

RE: ttrb20:A network-wide anticipatory control of an urban network using macroscopic fundamental diagram

Academic UK Non Rightslink <permissionrequest@tandf.co.uk>
Fri 4/16/2021 2:15 AM
To: Nadia Moshahedi <nadia.moshahedi2@ucalgary.ca>

[ΔEXTERNAL]

Dear Nadia Moshahedi,

Material Requested: Nadia Moshahedi, Lina Kattan & Richard Tay (2021) A network-wide anticipatory control of an urban network using macroscopic fundamental diagram, Transportmetrica B: Transport Dynamics, 9:1, 415-436, DOI: 10.1080/21680566.2021.1878966

Thank you for your correspondence requesting permission to reproduce the above content from our Journal in your online thesis and to be posted in the university's repository – University of Calgary.

We will be pleased to grant permission to reproduce your '**Accepted Manuscript'** on the sole condition that you acknowledge the original source of publication.

This is an '**Accepted Manuscript'** of an article published by Taylor & Francis Group in Transportmetrica B: Transport Dynamics on 1 Feb 2021, available online: https://doi.org/10.1080/21680566.2021.1878966

Please note: This **does not allow** the use of the **Version of Record (VoR)** to be posted online, however you may include the VoR as an Appendix to the printed version of your thesis.
(VoR is the final, definitive, citable version of your paper, which has been copyedited, typeset, had metadata applied, and has been allocated a DOI (Digital Object Identifier). This is the published version (PDF) on Taylor & Francis Online)

Using a DOI to link to the VoR on Taylor & Francis Online means that downloads, Altmetric data, and citations can be tracked and collated – data you can use to assess the impact of your work.

This permission does not cover any third party copyrighted work which may appear in the material requested.

Sharing your work - https://authorservices.taylorandfrancis.com/sharing-your-work/

Please do not hesitate to let me know if I can be of any additional assistance.

Kind Regards

Jo Bateman – Permissions & Licensing Executive, Journals
Taylor & Francis Group
3 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN, UK.
Web: www.tandfonline.com
e-mail: joanne.bateman@tandf.co.uk