School of Electrical Engineering, Computing and Mathematical Sciences

Optimization of Road Network Operations under uncertain events

Chuanye Gu

This thesis is presented for the Degree of
Doctor of Philosophy
of
Curtin University

September 2022

Declaration

To the best of my knowledge and belief, this thesis contains no material previously published by any other person except where due acknowledgment has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Signature: .....

Date: 18-Sep-2022

Abstract

In this thesis, we study the freeway traffic flow dynamics to reduce traffic congestion so as to optimize road management by implementing effective control strategies. Two types of methods are used in this study, including the model-based methods and data-driven methods. The first one is the macroscopic traffic flow model (cell transmission model (CTM)), which is used to describe the dynamical system of the freeway. Based on this model, an optimization program of the freeway system with consideration of partial information, for example, the demand mean and variance, is first proposed, and an effective solution technique is developed. Then, an optimal control model with ramp metering (RM) is proposed to solve the flow holding-back problem caused by that the 'min' operators are directly relaxed by inequality constraints. Theoretical analysis is established and a customized solution method is designed to solve this model. Finally, a smooth optimization model with a combination of RM and variable speed limits (VSL) is developed with consideration of the capacity drop. Convergency analysis is also established and an efficient solution method is designed to optimize this model. The second one is a data-driven Deep Koopman model, which applies the Koopman operator theory to characterize the dynamical freeway system with RM. A real-time controller is obtained by solving the designed optimal control problem. The main contributions are presented in two aspects as follows:

(1) The proposed models have theoretical guarantees.

(i) With available partial information of random traffic demands, the distributionally robust chance constrained optimization model is rewritten as a semi-definite programming. This transformation has the theoretical guarantee. Furthermore, an optimal control is obtained by solving the approximated problem with the objective of total delay of main road and on-ramps.

(ii) Two novel smooth optimization models on the basis of the CTM are proposed to solve the flow holding-back problem resulted from the impractical fundamental diagram if the 'min' operators are relaxed into inequality constraints directly. The former only considers the freeway RM, and the latter takes both

RM and VSL into account under the capacity drop. Theoretical analyses have been carried out for both proposed models. Theoretical results show that the solution obtained from the proposed model converges that of the primal CTM when ϵ approaches to zero.

(iii) A data-driven model on the basis of the Koopman operator is utilized to formulate the freeway dynamical system due to its interpretability and linearity. To learn a finite-dimensional approximation of the Koopman operator, a Deep Koopman model is proposed using neural networks with an encoder-decoder structure. The LSTM units are used to process the sequential traffic data in the encoder.

(2) The numerical results obtained from the study verify the efficiency and performance of the proposed models.

(i) The numerical results of the deterministic optimization model, robust optimization model and the proposed model, are showed and compared. These results indicate that the proposed model is efficient to control the total delay of the system when the considered dynamical system encounters uncertainties in a series of scenarios. The result also shows that the proposed model is more efficient in reducing traffic congestion of the system.

(ii) The proposed smooth optimal control models can effectively handle the flow holding-back problem under practical and synthetic applications. Furthermore, both RM and VSL are efficient control strategies of mitigating traffic congestion. Finally, the performance is the best when RM and VSL are implemented cooperatively.

(iii) A real-time control of RM is provided based on the Deep Koopman model. Compared to the baseline methods, the proposed Deep Koopman model shows a better precision in the predictions of the states. The performance of the real-time control of RM between the Deep Koopman model and other methods is assessed. The Deep Koopman model outperforms other baselines on all metrics.

List of publications during PhD candidature

- C. Gu, C. Wu, Y. Wu and B. Wiwatanapataphee, "Distributionally robust ramp metering under traffic demand uncertainty," Transportmetrica B: Transport Dynamics, vol. 10, no. 1, pp. 652-666, 2022.
- C. Gu, C. Wu, K. L. Teo, Y. Wu and S. Wang, "A Smoothing Method for Ramp Metering," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 8, pp. 13358-13371, Aug. 2022.

Acknowledgements

Time flies like an arrow, time flies like a flash, three years of doctoral study career is coming to an end. These three years will surely become an important experience and valuable wealth in my future life. During the three years as a doctoral student in the Department of Mathematics and Statistics, Curtin University, I have gained a lot in study, life and work.

On the occasion of the completion of this thesis, I would like to express my heartfelt thanks to my supervisor, Prof. Yonghong Wu. This thesis is completed under the guidance and care of Prof. Wu. Thank you very much for your care and encouragement. Throughout the period of my study, Prof. Wu has paid a lot of effort. Prof. Wu has deep academic attainments and extensive knowledge. I get very good inspirations when I discuss academic issues with Prof. Wu.

I would like to thank my co-supervisors, Prof. Benchawan Wiwatanapataphee, for her meticulous care, support and academic guidance throughout the period of my study.

I would like to express my gratitude to my co-supervisors, Prof. Changzhi Wu of Guangzhou University, for his continuing encouragement and wise advice at every step in my research during the past three years.

I am also indebted to thank Prof. Kok Lay Teo, for his help and encouragement.

I am very grateful as well to research colleagues for their support and help.

I thank all of the staff in the School of Electrical Engineering, Computing and Mathematical Sciences for contributing to a friendly working environment.

I sincerely thank the financial support of the Australian Government Scholarship for my PhD study.

Finally, I am very grateful as well to everyone in my family, especially my parents. They have given me great spiritual support and love during the entire period of my PhD candidature in Australia. Every step of my growth has been poured into their countless efforts and care.

Contents

1 Introduction **1**

1.1 Background 1

1.2 Objectives of the thesis 3

1.3 Outline of the thesis 4

2 Literature Review **6**

2.1 General 6

2.2 Traffic flow theory 6

2.2.1 A brief review of the development of traffic flow theory . . 6

2.2.2 Fundamental diagrams of macroscopic traffic flow models . 7

2.2.3 Macroscopic traffic flow models 9

2.3 Koopman operator 14

2.3.1 Overview of the Koopman operator 14

2.3.2 Definition of Koopman operator 16

3 Distributionally Robust Ramp Metering Under Traffic Demand Uncertainty **18**

3.1 Introduction 18

3.2 Modeling Traffic Flow Dynamics 20

3.3 Optimization Model 22

3.3.1 Deterministic Optimization Model 22

3.3.2 Distributionally Robust Chance Constrained Optimization Model 23

3.4 Approximation of Distributionally Robust Chance Constraint . . 26

3.5 Case Study 30

3.5.1 Results from the Deterministic Model 30

3.5.2 Results from the Disributionally Robust Chance Constrained Optimization Model 32

3.6 Conclusion 34

4 A Smoothing Method for Ramp Metering **36**

4.1 Introduction . . . 36
4.2 Optimization problem . . . 38
4.2.1 Primal cell transmission model . . . 40
4.2.2 Existing Solution Methods . . . 41
4.3 Approximate Model . . . 43
4.3.1 Min Approximation . . . 43
4.3.2 Approximate model . . . 45
4.4 Gradients of the Objective and Constraint Functions . . . 51
4.4.1 Gradient formulas . . . 51
4.4.2 Algorithm . . . 55
4.5 Numerical Studies . . . 56
4.5.1 MPC Design . . . 57
4.5.2 Example 1 . . . 58
4.5.3 Example 2 . . . 66
4.6 Conclusion . . . 68

5 Freeway Ramp Metering and Variable Speed Limit Control Based on Cell Transmission Model 70
5.1 Introduction . . . 70
5.2 Modeling Traffic Flow Dynamics . . . 72
5.2.1 The application framework . . . 72
5.2.2 The CTM Model for the Problem . . . 73
5.2.3 Incorporating the RM impact . . . 75
5.2.4 Incorporating the VSL Impact . . . 75
5.2.5 The Integrated Optimal Control Problem . . . 78
5.3 Approximation Approach and Model Construction . . . 79
5.3.1 Smooth approximation . . . 79
5.3.2 Proposed approximation model . . . 81
5.4 Customized Solution Technique . . . 81
5.4.1 Gradients of objective and constraint functions . . . 82
5.4.2 Algorithm . . . 86
5.5 Numerical Application . . . 86
5.5.1 Case one . . . 87
5.5.2 Case two . . . 89
5.6 Conclusion . . . 97

6 Deep Koopman Traffic Modeling for Freeway Ramp Metering 98
6.1 Introduction . . . 98

6.2 The freeway traffic model and ramp metering controls 101
6.3 Deep Koopman Model for the freeway traffic flow dynamics 103
6.3.1 The Koopman Operator with forced system 103
6.3.2 Extended Dynamic Mode Decomposition 104
6.3.3 The Deep Koopman Method 106
6.4 MPC-Deep-Koopman for Freeway Traffic Flow Systems with Ramp Metering 110
6.5 Numerical Studies 113
6.5.1 Experiment Design 113
6.5.2 Evaluation metrics and baselines 115
6.5.3 Example one 115
6.5.4 Example two 122
6.6 Conclusion 124

7 Conclusions and future work **126**
7.1 Conclusions 126
7.2 Future work 127

Bibliography **129**

CHAPTER 1

Introduction

1.1 Background

The road transportation system is an important link connecting cities and an important channel of transportation and logistics, which brings great convenience to our daily production and life and promotes the rapid development of economy and society. With the development of economy, the increase of population and the expansion of urban cities, people's demands for travel have increased sharply, and the number of private vehicles has also been rising. Although increasing traffic facilities can improve traffic conditions to a certain extent, the speed of infrastructure construction has fallen far behind the growth of traffic demand due to the constraints of land and capital. This has led to the increasingly serious contradiction between traffic supply and demand, resulting in a series of traffic problems not only for the urban traffic in big cities but also for freeways. Traffic congestions could trigger serious consequences, such as time delays, air pollution and economic losses [1]. Thus, strengthening the management of transportation and developing advanced transportation system are very important and valuable.

The common occurrence of traffic congestion on road networks is not only due to the lack of road capacity, but also due to unreasonable road network plans and unscientific traffic managements such as the decrease of traffic capacity at intersections and bottlenecks, the unreasonable proportion of diversion, and the lack of implementation of traffic guidance and other management measures.

From the perspective of traffic supply, the main reasons leading to traffic congestions are the lack of static traffic capacity, traffic management facilities and scientific managements caused by the disharmony between urban planning and traffic planning.

From the perspective of traffic demand, traffic congestion is mainly caused by the weak awareness of traffic laws and regulations of urban traffic participants

and the low degree of traffic civilization. Therefore, to alleviate traffic congestion, on the one hand, we should continue to increase and improve road traffic infrastructures, such as increasing the public transport construction, improving the transfer, improving the level of intelligent traffic management, reasonably inducing the traffic travel, etc. On the other hand, traffic participants should fully abide by traffic rules, follow the traffic guidance, strengthen green and low-carbon travel awareness, and make full use of buses and bicycles. New York and Stockholm have adopted traffic information systems to ease congestion. Copenhagen and Paris are tackling traffic congestion by investing more in the public transports and encouraging the green travel.

It is an urgent problem to explore the internal characteristics of traffic flows and the cause of traffic jams and take effective measures. At present, the development of intelligent information technology has brought not only great changes to the society, but also intelligences and conveniences to the field of transportation. Unilateral measures such as strengthening regulations, restricting vehicles, and increasing road constructions do not comprehensively consider people, cars, roads and the environment. The ideal approach is to combine these four elements of transportation, and the Intelligent Transportation System (ITS) gives full play to its advantages. ITS effectively applies the advanced science and technology (such as data communication technology, information technology, electronic control technology, sensor technology, etc.) to the transportation, the service control and the vehicle manufacturing. Enhancing the connection among vehicles and roads as well as commuters may form a comprehensive transportation system that guarantees safety, improves efficiency, saves energy and improves environment. ITS can make maximum use of the potential of the existing transportation infrastructure, so it has become one of the most important approaches to alleviate traffic problems.

With the development of ITS, Connected Vehicle (CV) has gradually been developed and become the core technology of ITS. CV is the specific application of Internet of Things in ITS field. By integrating various communication technologies, CV connects various components inside the vehicle and between the vehicle and the outside world into a network, forming an integrated network integrating internal network, inter-vehicle network and on-board mobile Internet.

In the past decades, on the one hand, from the perspective of models, many effective traffic flow models have been developed (such as the Car-following Model [2–4], the Cellular Automata Model [5–7] and the Macro-Hydrodynamics Model [8–11]) to study the production mechanism of traffic jams and mitigate traffic

congestion. Among them, the macro-hydrodynamics model focuses on the overall traffic flow characteristics, and applies partial differential equations modelling. It is easy to be solved mathematically due to few parameters and characteristics of physical significance.

On the other hand, from the data-driven perspective, machine learning techniques (such as Deep learning (DL) [12–15] and Reinforcement learning (RL) [16–18]), and Koopman Operator Theory [19–21] have been applied to transportation systems. Among them, Koopman decomposition can not only analyze the physical properties of transportation system, but also provide the low-dimensional system approximation of the high-dimensional system, so as to achieve pattern reduction. Therefore, it is of great significance to design optimization models to mitigate traffic congestion on roads and networks by using the macroscopic model and the Koopman operator.

1.2 Objectives of the thesis

Using traditional mathematical models and data-driven methods to model and study the evolution dynamics of freeway traffic flows has important guiding significance, which can help us to find the corresponding effective control methods, for example, ramp metering (RM) and variables speed limits (VSL), to alleviate traffic congestion, so as to improve traffic operations.

This thesis mainly focuses on the study of both macroscopic traffic flow models, especially CTM, and data-driven models, particularly Koopman operator. The control of freeway traffic flow dynamics is modeled as an optimal control problem. The main goal of this work is to find the optimal control approaches to mitigate and solve traffic jams on freeways.

The main objectives of this work are stated below:
(1) Propose an optimal control model governing the freeway traffic flow system with partial information (demand mean and variance) to implement RM.
(2) Develop a smooth optimal control model governing the freeway traffic flow dynamics with RM to solve the flow holding-back problem.
(3) Design an optimal control model with consideration of the capacity drop to implement both VSL and RM.
(4) Establish a data-driven model (Koopman operator) to optimize the freeway management by RM.

1.3 Outline of the thesis

In this thesis, different models governing the dynamical freeway traffic flow system taking various scenarios into account have been proposed. This thesis is divided into seven chapters, where the main content in each chapter is stated as follows:

Chapter 1 : The overview of research background and the objective of research have been introduced in section 1.1 - 1.2, respectively.

Chapter 2 : A literature review of the existing work and corresponding results have been presented, including various traffic flow models, control methods.

Chapter 3 : A distributionally robust optimization model is proposed to address the optimal control problem with RM under the random traffic demand flow. Applying the Worst-Case Conditional Value-at-Risk (WCVaR) constraints to approximate the distributionally robust chance constraints, the considered problem is conservatively approximated as a semi-definite programming (SDP), which has high computational efficiency.

Chapter 4 : An optimization model is developed for the RM on the basis of CTM. This problem considered is modeled as an optimal control problem to regulate traffic inflows from the on-ramp. A smooth approximation method is introduced to approximate the 'min' operator and then an efficient computational method is proposed to resolve the problem. Theoretical analysis shows that optimal solutions derived from the approximate problem converge to that of the original CTM.

Chapter 5 : A novel RM and VSL control model is proposed, in which the influences of RM and VSL are incorporated into the CTM as control components. The problem of traffic control in the highway network via RM and VSL is modeled as a optimal control problem with constraints. A smoothing method is introduced to approximate this non-smooth 'min' operators in this model, and an efficient MPC solution technique based on co-state system is used to solve this smooth approximate control problem.

Chapter 6 : A data-driven modelling method using neural networks, denoted by the Deep Koopman model, is developed to learn a finite-dimensional approximation of the Koopman operator. To consider the sequential relations of the

ramps and main roads on the freeway, a long-short term memory (LSTM) network is applied. Furthermore, a MPC controller with the trained Deep Koopman model is proposed for the real-time control of the ramp metering on the freeway.

Chapter 7: A conclusion and discussion on future work are presented.

CHAPTER 2

Literature Review

2.1 General

In this chapter, we present a brief review of the development of traffic flow theory in 2.2.1, the fundamental diagrams of the macroscopic traffic flow models in 2.2.2, the classical macroscopic traffic flow models in 2.2.3, and finally the development and definition of the Koopman operator in 2.3.1 and 2.3.2, respectively.

2.2 Traffic flow theory

2.2.1 A brief review of the development of traffic flow theory

The goal of traffic flow theory research is to establish a traffic flow model that can explain the general characteristics of actual traffic, and reveal the basic laws of traffic flow, so as to better guide the design, management and control of traffic system.

Traffic flow theory originated in the 1930s. In the 1930s and 1940s, the free flow theory was mainly studied. In this phase, the traffic flow is mainly manifested as the free flow, which is characterized by low vehicle density on the road, large distance between vehicles, and free driving of vehicles. Therefore, the probability theory and mathematical statistics can be utilized to establish a mathematical model to study the internal relationship between traffic flows and speeds.

In the 1950s and 1960s, the non-free flow theory entered the research stage. During this period, with the rapid growth of highway mileage and the sharp increase of road traffic flows in developed countries, the independence of vehicles in traffic flow became smaller and smaller, and various new models emerged one after another. Representative theories include the Car-following Model [22–25],

the traffic wave theory based on fluid dynamics and queuing theory. Lighthill and Whitham [8] proposed the hydrodynamic simulation theory. They assumed that the road was closed with no entrances and exits. Considering that vehicular traffic flow satisfies the law of conservation, the equilibrium velocity was adopted to establish a one-dimensional continuous model, which could describe traffic shock waves. However, due to the adoption of equilibrium velocity, the model could not effectively describe traffic flow phenomena, for instance, capacity drop and stop-and-go.

From 1970s to 1980s, the number of vehicles on highways further increased. Some scholars believed that when the traffic density on the road reached a certain level, the traffic flow could be regarded as the water flow. Based on the fluid dynamics equation of vehicle motion and the flow conservation equation of Lighthill and Whitham [8], a continuous fluid dynamics model was obtained. Since 1990s the Cellular Automata Model has become a hot topic in traffic flow theory.

In short, after decades of development, traffic flow models mainly include microscopic, mesoscopic and macroscopic models. The microscopic models describe individual behaviors between individual vehicles and are discrete in spatial and temporal distribution in research methods. The macroscopic models consider a fleet of vehicles on the road as a continuous fluid and the average behavior of a large number of vehicles rather than the individual behavior of a single vehicle. The research method is continuous both in spatial and temporal distribution. The mesoscopic model is a kind of model between the microscopic and macroscopic model, mainly referring to the gas dynamic model based on probability theory. Among them, the macroscopic model is modeled by partial differential equation, which has few parameters, easy to obtain mathematical solutions, and has the characteristics of physical significance. The basic parameters and fundamental diagrams of traffic flow are the basis for macroscopic traffic flow models.

2.2.2 Fundamental diagrams of macroscopic traffic flow models

Fundamental diagrams of traffic flow refer to the relationship between traffic flow and density, which describes the functional relationship among vehicle density and traffic flow.

The flow is the number of vehicles passing a certain observation point in unit time, usually denoted by $f = N/T$, in which N represents the number of vehicles passing the observation point in time period T.

The speed usually refers to the average velocity of traffic flows, including the temporal average velocity and spatial average velocity. The temporal average velocity is the arithmetic average of velocity of multiple cars passing a point on the road in a period of time, expressed by $v_t = \sum_{i=1}^{N} v_i/N$, where v_i represents the velocity of the ith car passing the point. The spatial average speed is the arithmetic average of all the time that takes for multiple vehicles to pass a certain section of road in a certain time, usually expressed by $v_s = L/(\sum_{i=1}^{N} t_i/N)$, where L denotes the length of observed road and $t_i = L/v_i$ refers to the time that takes for the ith vehicle to pass the section of road.

The density refers to the number of vehicles per unit length of road at a given moment, usually expressed by $\rho = N/L$.

The three parameters of traffic flow, density and speed meet the following basic relation:

$$f = \rho v, \tag{2.1}$$

where the traffic flow f is the product of the density ρ and velocity v. The relation (2.1) connects the three parameters, so as long as any two parameters are known, the third parameter can be determined. This relation can be expressed in the coordinate system by the diagrams, namely, the fundamental diagrams, which includes the measured fundamental diagrams and the theoretical fundamental diagrams.

Based on a large number of measured results, it is generally considered that when the density $\rho \to 0$, the flow $f \to 0$; when $\rho = \rho_{max}$ (jam density), the vehicle can hardly run, so $f = 0$. Between these two extremes, the flow f must has a maximum C (also known as the road capacity). This maximum inflection point should has a critical density ρ_{cr}. When $\rho < \rho_{cr}$, the traffic flow is in the condition of free flow; when $\rho > \rho_{cr}$, the traffic flow is in the congestion state.

Common theoretical fundamental diagrams include: the Greenshields fundamental diagram [26, 27]: the Greenberg fundamental diagram [28]: the Newell fundamental diagram [29]: the Underwood fundamental diagram [30], the Drake fundamental diagram [31], the Pipes-Munjal fundamental diagram [32].

Moreover, Newell [33] proposed a triangular fundamental diagram, where the speed-density model is denoted by:

$$v(\rho) = \begin{cases} v_f, & 0 \le \rho \le \rho_{cr}, \\ \dfrac{\rho_{cr}}{\rho_{max} - \rho_{cr}} v_f \left(\dfrac{\rho_{max}}{\rho} - 1\right), & \rho_{cr} \le \rho \le \rho_{max}. \end{cases} \tag{2.2}$$

and the flow-density model is described as:

$$f(\rho) = \begin{cases} v_f\rho, & 0 \le \rho \le \rho_{cr}, \\ \dfrac{\rho_{cr}}{\rho_{max}-\rho_{cr}} v_f(\rho_{max}-\rho), & \rho_{cr} \le \rho \le \rho_{max}. \end{cases} \tag{2.3}$$

Furthermore, Kerner and Konhäuse [34] and Herrmann and Kerner [35] proposed a nonconvex fundamental diagram

$$v(\rho) = v_f\{1-\exp[1-\exp(\frac{|c_j|}{v_f}(\frac{\rho_{max}}{\rho}-1))]\}. \tag{2.4}$$

Among these mentioned fundamental diagrams, the Greenshields fundamental diagram and the triangular fundamental diagram are commonly used in the research.

2.2.3 Macroscopic traffic flow models

The macroscopic traffic flow model describes the traffic flow as a compressible continuous fluid medium composed of a large number of vehicles, and studies the average behavior of vehicles rather than characteristics of individual vehicles.

A Continuous model

The macroscopic model of traffic flow was independently proposed by Lighhill and Whitham [8] and Richards [9], also known as LWR model. The LWR model is a first-order continuum model describing the traffic motion and demonstrates the existence and characteristics of traffic shock waves. According to the basic principle of conservation of the number of vehicles in the traffic system, the traffic flow of system without ramps satisfies the following equation:

$$\frac{\partial f}{\partial x} + \frac{\partial \rho}{\partial t} = 0, \tag{2.5}$$

where ρ denotes the density, f refers to the flow, and t and x are the time and the location, respectively.

In the LWR model, the traffic flow is assumed to satisfy the velocity-density relation of equilibrium state:

$$v = v_e(\rho), \tag{2.6}$$

where v is the average speed of traffic flow, subscript e denotes the equilibrium state of traffic flow. Combined with (2.5) and (2.6), the closed hyperbolic equation

can be obtained as follows:

$$(v_e + \rho\frac{\partial v_e}{\partial \rho})\frac{\partial \rho}{\partial x} + \frac{\partial \rho}{\partial t} = 0, \tag{2.7}$$

where $f_e(\rho) = \rho v_e(\rho)$. This model describes the propagation of nonlinear density wave and its solution can be obtained by using the method of characteristic equation and numerical simulation. However, this model usually assumes that the vehicle speed always satisfies the equilibrium relation, and thus it cannot describe the unbalanced traffic phenomena, for example, the instability of traffic flow and the formation of stop-and-go traffic phenomenon under certain conditions.

To address the shortcoming of LWR model, Payne [10] and Whitham [36] proposed a higher-order model (called PW model for short), which replaced the equilibrium speed-density relationship in the LWR model with the kinetic equation of velocity, allowing the velocity to deviate from the equilibrium speed. The kinetic equation is expressed as follows:

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} + \frac{1}{\rho}\frac{\partial p(\rho)}{\partial x} = \frac{v_e(\rho) - v}{\tau}, \tag{2.8}$$

where r is the relaxation time, which describes that the driver adjusts the speed v to reach the equilibrium speed v_e within the time; $p(\rho)$ refers to the pressure term, which satisfies $p^{'}(\rho) > 0$ and $2p^{'}(\rho) + \rho p^{''}(\rho) > 0$. After that, many higher-order continuous models were developed, such as the Papageorgiou model [37], the Kühne equation model [38], and the Berg model [39], etc. Because the dynamics equations of these models all have density gradient terms, these models are also called density gradient models. Daganzo [40] criticized the density gradient model, believing that there is a case that the characteristic velocity is greater than the macroscopic traffic flow speed. This indicates that vehicles will be affected by rear vehicles in the operation, which is contrary to the anisotropy of the traffic flow, thus causing the phenomenon of vehicle regression. To handle this issue, Aw-Rascle [41] designed a new high-order macroscopic model, where the kinetic equation is formulated as:

$$\frac{\partial(v + p(\rho))}{\partial t} + v\frac{\partial v}{\partial x} + \frac{\partial v p(\rho)}{\partial x} = 0, \tag{2.9}$$

where $p(\rho)$ has various expressions. Classically, $p(\rho) = \rho^{\gamma}$, $\gamma > 0$.

Zhang [42] proposed a similar model, where the kinetic equation is denoted

by:

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} + \rho v_e^{'}(\rho)\frac{\partial v}{\partial x} = 0. \tag{2.10}$$

By the connection between microscopic and macroscopic variables, Jiang et al. [11] deduced a higher-order model from the microscopic-following model, whose dynamics equation is:

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} - c_0\frac{\partial v}{\partial x} = \frac{v_e(\rho) - v}{\tau}, \tag{2.11}$$

where $c_0 = \Delta x\lambda$. Xue and Dai [43] further proposed a similar model with relaxation terms.

Furthermore, researchers have extended and developed all kinds of mixed traffic flow models. The hydrodynamic models of mixed traffic flow can be divided into two categories: the multi-lane model and the multi-class vehicle model. The former establishes a governing equation for each lane and considers the influence of vehicle lane change behavior in the equation. In the latter, the conservation equation is established for each type of vehicles, and the speed density function is set to obtain the model. Both multi-lane and multi-class vehicle models can be divided into LWR model and high-order model.

Michalopoulos et al. [44] proposed a multi-lane LWR model, which introduced variable lane change intensity parameters and lane change time delay when calculating the lane change flow. The model proposed by Holland and Woods [45] was mainly aimed at the situation with small density of two lanes. Lava and Daganzo [46] proposed a discrete multi-lane LWR model. [47] proposed a macroscopic model that could reveal the non-equilibrium characteristics of vehicle flow such as the small disturbance instability.

In reference [48], the multi-class vehicle LWR model was proposed, which can simulate the interaction between vehicles at different speeds on the road, and predict and explain many phenomena in the traffic flow. Zhang et al. [49] thoroughly discussed the hyperbolic properties of the model. This model was extended to non-uniform roads, in which the number of lanes and traffic capacity change with space [50]. The model proposed by Benzoni-Gavage and Colombo [51] took into account the effect of vehicle length. Chanut and Buisson [52] introduced the concept of critical density.

B Lattice hydrodynamic model

Nagatani [53] simplified the macroscopic model proposed in [54] and referred to the idea of Car-following model to obtain the lattice hydrodynamic model, whose dynamic equation is written as below:

$$\partial_t \rho + \partial_x(\rho v) = 0, \tag{2.12}$$

$$\partial_t(\rho v) = a\rho_0 V(\rho_{x+\delta}) - a\rho v, \tag{2.13}$$

where ρ and v denote the local average density and the local average speed, respectively; ρ_0 is the average density; $\delta = 1/\rho_0$ is the average headway; $\rho_{x+\delta}$ denotes the local density at the location $x + \delta$ and at the time t; and a refers to the driver's sensitivity coefficient. For the spatial variable x in (2.12) to be dimensionless, let:

$$\widehat{x} = \frac{x}{\delta}, \tag{2.14}$$

we can get:

$$\frac{\partial(\rho v)}{\partial x} = \frac{\partial(\rho v)}{\delta \partial x/\delta} = \rho_0 \frac{\partial(\rho v)}{\delta \partial \widehat{x}}. \tag{2.15}$$

For convenience, $\widehat{x}$ is still denoted as x, and the following model is obtained:

$$\partial_t \rho + \rho_0 \partial_x \rho v = 0, \tag{2.16}$$

$$\partial_t(\rho v) = a\rho_0 V(\rho_{x+1}) - a\rho v. \tag{2.17}$$

Discretization of variable x in (2.16) and (2.17), we can obtain the following lattice fluid dynamics model:

$$\partial_t \rho_i + \rho_0(\rho_i v_i - \rho_{i-1} v_{i-1}) = 0, \tag{2.18}$$

$$\partial_t(\rho_i v_i) = a\rho_0 V(\rho_{i+1,t}) - a\rho_i v_i. \tag{2.19}$$

where ρ_i and v_i represent the local density and velocity at the i grid point on the one-dimensional grid at the time t, respectively; $V(\rho_{i+1,t})$ is the optimization

velocity, which depends on the density at the $i+1$ grid point ahead; $\rho_0 V(\rho_{i+1,t})$ represents the expected flow. Formula (2.19) indicates that the driver obtains the expected flow based on the observed density ahead, and relaxes the traffic flow to the expected flow.

In recent years, many scholars have done a lot of studies and generalizations for the Nagatani model [53], and proposed the delayed lattice hydrodynamics model [55–57], and the multi-expectation lattice hydrodynamics model [58–61]. Nagatani [62] extended the model to two lanes by introducing the change lane effect. Furthermore, Nagatani [63] proposed a two-dimensional lattice model, where the evolution of traffic flow density waves is different from that of the one-dimensional model.

C Cell transmission model

Both the macroscopic LWR model and the higher-order macroscopic model have discrete schemes for finding numerical solutions, such as the Godunov scheme [64] and the Lax-Wendroff scheme [65]. Daganzo [66–68] further simplified the flow-density relationship and the solving process of the traffic flow in the model, and established the CTM through the mathematical processing method of finite differentiation and Godunov scheme. The model can not only describe the discontinuous phenomena, for example, the generation and dissipation of shock waves in traffic flow, but also easily model the real road network and calculate the solution.

CTM proposed by Daganzo [66,67] divides the road into several cells, and the time is discretized. The length of each cell refers to the distance traveled by free flow in a time step. The change of traffic flow on the whole road is described by the number of vehicles $N_{i,t}$ in each cell i and the change of the number of vehicles flowing into and out of each cell at different times.

CTM assumes that the flow-density follows the fundamental diagram, as shown in Fig. 2.1. We introduce the flow-density relation as follows:

$$f = \min\{v\rho, C, w(\rho_{max} - \rho)\}, \tag{2.20}$$

where ρ_{max} represents the jam density, C denotes the traffic capacity, v refers to the free flow speed, w is the speed of backward wave. Further, using the trapezoidal fundamental diagram to discretize LWR model of one-way freeway section, we can get the evolution of traffic density and flow as time changes:

$$n_{i,t+1} = n_{i,t} + f_{i,t} - f_{i+1,t}, \tag{2.21}$$

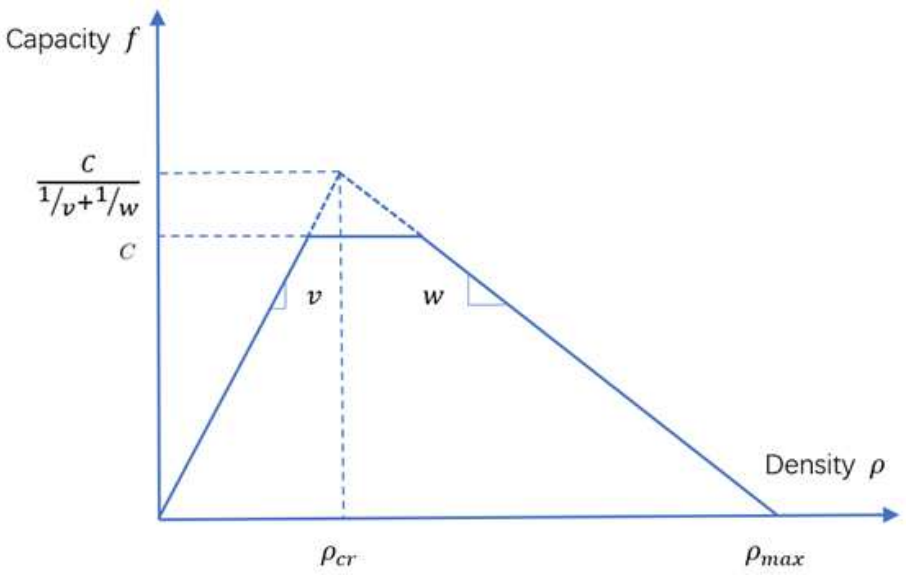


Figure 2.1: Trapezoidal fundamental diagram

$$f_{i,t} = \min\{n_{i-1,t}, C, \frac{w}{v}(N_{i,t} - n_{i,t})\}, \tag{2.22}$$

where $N_{i,t}$ denotes the maximum number of vehicles that can be carried at the cell i and the time t.

In the past two decades, scholars have made many improvements and promotions on CTM, such as Lagged CTM (LCTM) [69], Improved LCTM (ILCTM) [70], Modified CTM (MCTM [71, 72], State switching model (SSM), Asymmetric CTM (ACTM) [73, 74], Location specific CTM (LSCTM) [75].

Because of its advantages of describing the discontinuous phenomena such as the generation and dissipation of shock waves in traffic flow, CTM has been widely utilized in traffic signal optimization control [76–79], dynamic traffic prediction [80], dynamic traffic simulation [81], dynamic traffic allocation [82–84] and other fields [85, 86].

2.3 Koopman operator

2.3.1 Overview of the Koopman operator

The Koopman operator was proposed by B.O.Koopman [87]. The author showed that any observable evolution in dynamic systems can be expressed by infinite dimensional linear operators. The early work mainly focused on the spectrum properties of the Koopman operator in statistical mechanical systems and nonlinear dynamical systems [88–90]. Recently, the dynamic mode decomposition

algorithm (DMD) on the basis of Koopman operator and other improved algorithms have been applied to different fields, such as biological networks [91], power systems [92, 93], human motion analysis [94], building energy efficiency models [95, 96], transportation systems [21].

Theoretically, the Koopman operator can transform the nonlinear evolution of dynamic system into the linear evolution of infinite dimensional space. However, in practice, it is very difficult to use infinite dimensional operator, and so how to calculate the finite dimensional approximation of Koopman operators is a big challenge. One of the most commonly used algorithms to approximate the Koopman operator is DMD, which can obtain the spectrum of Koopman operators from the associated data of time or space, that is, eigenvalues and eigenfunctions. Murshed and Uddin [97] indicated that DMD can predict the future state information of dynamic system, such as the distribution of electric field [98] and the change of temperature [99].

Using the Krylov subspace to construct a unitary matrix that approximate the Koopman operator, Rowley et al. [100] proposed a Koopman spectrum analysis and pattern decomposition method by Arnoldi algorithm. By analyzing the advantages and disadvantages of Arnoldi algorithm, Schmid [101] pointed out that the Arnoldi algorithm relies too much on the last observation data, and the results obtained by the Arnoldi algorithm are unstable when the influence of system noise or measurement error large. Hence, the method of singular value decomposition (SVD) was developed to obtain the spectrum of Koopman operator [101].

Budisic [102] summarized the theory and application of Koopman spectrum analysis, and pointed out that in the given data, the tuple of Koopman operator can be obtained, that is, eigenvalues, eigenfunctions and Koopman modes. The eigenfunctions can be used to divide the ergodic states of the system, and the Koopman mode is used to simplify the model and study the correlation between different variables. After that, an exact DMD algorithm was proposed in [103]. William [104] proposed the extension of DMD algorithm, and named it EDMD. Both approaches depend on more standard methods to represent linear operators with respect to a particular basis, rather than DMD sampling methods based solely on adjoint matrices [105]. Korda and Mezić [106] further proved the relation between the spectrum of EDMD matrix and the eigenvalues of Koopman operator. Drmac et al. [105] proposed a data-driven mode decomposition (D-DMD) algorithm, which can scale the original data, so that the Koopman mode corresponding to small singular values can also be obtained. Moreover, DDMD adopted more accurate eigenvalues and modes, thus reducing the residual term.

Recently, by combining neural networks with the Koopman operator, Yeung et al. [107] has introduced a computational framework using deep learning to obtain the Koopman operator for nonlinear dynamical systems, and quantitatively predict the oscillation behavior of several steps in the future. Similarly, an optimal control method for dynamic systems using the same principle was also proposed in [108].

2.3.2 Definition of Koopman operator

The Koopman operator describes the evolution of observable in the phase space of dynamic system. Observable is a function of the state space, so the Koopman operator is acting on the infinite dimensional function space [102]. Consider a nonlinear dynamic system represented by discrete model in the phase space $\mathcal{M}$:

$$\mathbf{x}_{k+1} = f(\mathbf{x}_k), \mathbf{x}_k \in \mathcal{M}, \tag{2.23}$$

where $k \in \mathbb{N}$ represents the time variable, f is a nonlinear mapping between different states in the phase space $\mathcal{M}$. Define the observation function of the system $y_k = g(\mathbf{x}_k)$, $g : \mathcal{M} \to \mathbb{C}$, where $\mathbb{C}$ represents the complex field. The Koopman operator $\mathcal{K}$ is defined as follows:

$$\mathcal{K}g(\mathbf{x}_k) = g(f(\mathbf{x}_k)) = g(\mathbf{x}_{k+1}). \tag{2.24}$$

The Koopman operator defines a new dynamic system, where the evolution of g is described. If the initial state is $\mathbf{x}_0$, the observation function value after time step $t \in \mathbb{N}^+$ steps is written as $y_t = g(\mathbf{x}_t) = \mathcal{K}^t g(\mathbf{x}_0)$.

Assuming $g^{'}(\mathbf{x}) = g(f(\mathbf{x}))$, we can get by applying the Koopman operator

$$\mathcal{K}g(\mathbf{x}) = g^{'}(\mathbf{x}). \tag{2.25}$$

Therefore, the function of $\mathcal{K}$ can be regarded as the evolution of the value of function $g(\mathbf{x})$ under the nonlinear mapping f, as well as the evolution of function form of $(g(\mathbf{x}) \to g^{'}(\mathbf{x}))$. The Koopman operator $\mathcal{K}$ is a linear operator, even though the dynamical system f is nonlinear. Hence, it holds:

$$\mathcal{K}(\alpha g_1(\mathbf{x}) + \beta g_2(\mathbf{x})) = \alpha \mathcal{K} g_1(\mathbf{x}) + \beta \mathcal{K} g_2(\mathbf{x}), \tag{2.26}$$

where $g_1(\mathbf{x}), g_2(\mathbf{x})$ denote the scalar functions in the phase space $\mathcal{M}$, and $\alpha, \beta \in \mathbb{C}$. Note that the Koopman operator is infinite-dimensional unless the state space is

finite.

The Koopman operator can be used in various fields due to the linearity, especially for real-time control in traffic management. Thus, using Koopman operators to model the freeway traffic flow evolution to obtain optimal control is valuable.

CHAPTER 3

Distributionally Robust Ramp Metering Under Traffic Demand Uncertainty

3.1 Introduction

In this Chapter, we presents a distributionally robust optimization model to address the ramp metering problem with uncertain traffic demand flows. The aim of this model is to minimize the total travel delay of the system based on the macroscopic CTM of traffic flow. In our model, the only required data is the partial distributional information of stochastic demand flows. Using the Worst-Case Conditional Value-at-Risk (WCVaR) constraints to approximate the distributionally robust chance constraints, the proposed problem can be conservatively approximated as a semi-definite programming (SDP), which is computationally efficient. The performances of our proposed model are illustrated by practical applications. Experimental results show that the distributionally robust control strategy can achieve reliable performances over a range of uncertain scenarios.

The seminal research of optimization-based ramp metering can be traced back to the work [109], where a static model of traffic behavior was used to formulate the problem. This model was subsequently investigated and extended by [110–112]. One of the most widely adopted classes of models in the freeway control design is the macroscopic models, including the first order CTM model and the second order model (Metanet). The CTM model initially proposed by [66,67] can be regarded as a first-order Godunov approximation of the continuous Lighthill-Whitham-Richards-model (LWR) [8,9], and Metanet was proposed in [113]. Particularly, Papageorgiou et al. [114, 115] concluded that the freeway ramp metering is an useful and effective tool to improve traffic flows on congestion-prone freeways. Kotsialos and Papageorgiou [116] proposed a model-predictive framework for coordinated ramp metering rooted in the METANET model and

formulated the considered problem as a nonlinear optimization problem. Based on the asymmetric cell transmission model, Gomes and Horowitz [73] considered an on-ramp metering problem, where both free flow and congested conditions can be captured by the problem formulation. Gomes et al. [74] further provided a theoretical analysis to study the behavior of the CTM model of a freeway with steady demand. To analyse traffic flow density on freeway sections with random demand and supply, Sumalee et al. [117] developed a stochastic cell transmission model by characterizing the probability distributions of occurrence of each mode.

Recently, Chow and Li [118] proposed a robust optimization model of dynamic motorway traffic flow to optimize the total travel delay of the system with random traffic flow demands as well as set-valued fundamental diagrams [119]. The problem was reformulated as a minimization and maximization problem when an ellipsoidal likelihood set was considered. Roncoli et al. [120] developed a novel first-order multi-lane macroscopic traffic flow model for motorways to consider lane changing and capacity drop via appropriate procedures for computing lateral and longitudinal flows. Based on the work [120], Roncoli et al. [121] proposed a linearly constrained optimal control model by permitting the deployment of lane changing control, variable speed limits, and ramp metering. Han et al. [122] proposed a general first-order traffic flow model to simulate the capacity drop at the on-ramp bottleneck and lane drop bottleneck. On this basis, a linear quadratic model for predictive control strategy was proposed to realize the integration of dynamic path guidance and ramp metering. Furthermore, Han et al. [123] considered the propagation of shockwave on the freeway network, and modified the supply function that depends on the density difference between cell i and its upstream cell $i-1$, where both cells are congested. In addition, on the arterial links where shockwave is generated during each cycle, if the upstream is in a free-flow condition, as in the modification made by Han [123], the demand function of the target cell will have the same structure as the traditional CTM, and it will overestimate the actual discharge rate in the target cell. Under the assumption that equipped vehicles can bidirectionally communicate with the infrastructures, a novel feedback based integrated control strategy was proposed by [124] to implement ramp metering and lane-changing control. By adjusting the adaptive cruise control (ACC) settings of equipped and connected vehicles in real time on the basis of the current traffic conditions, a simple and effective ACC-based control strategy is proposed by [125], where this control strategy relies only on real-time information about the current traffic conditions (no network topology information is required). Kontorinaki et al. [126] proposed a local and coordinat-

ed ramp metering strategy based on the nonlinear adaptive control scheme, which consists of a nominal feedback law and a nonlinear observer aimed at estimating some unknown system variables.

However, solutions obtained from the deterministic optimization models (DOM) and robust optimization models (ROM) are overly conservative. An adjustable robust optimization approach has been developed to alleviate the conservatism [127]. Based on the work, Zymler et al. [128] proposed a novel method to approximate the distributionally robust individual and joint chance constraints with the first- and second-order moments and the support of the uncertainties of parameters. The approach is effective and outperforms the approximation proposed by Chen [129] and Bonferroni approximation. Although a number of researches have studied the robust solutions of ramp metering optimization, little has been done on problems where only partial information (such as mean and variance) of traffic parameters is available. Since the information provided by loop detectors may be incomplete in practice, using exact information of traffic parameters to study the traffic problem is often not possible in practice. It is an interesting point that we can conduct our research.

The rest of the chapter is organized as follows: Section 3.2 models the traffic flow dynamics. In section 3.3, we introduce a DRCCOM rooted in the deterministic cell transmission model. The approximation of distributionally robust chance constraints are presented in Section 3.4. In Section 3.5, the performances of various control strategies are illustrated and compared using practical examples. Finally, Section 3.6 gives some conclusions.

3.2 Modeling Traffic Flow Dynamics

The CTM proposed by Daganzo in [66] is one of the most widely utilized discrete models. Due to the popularity and credibility of CTM, we utilize CTM to model the traffic flow dynamics in this work.

In the formulation of CTM, a freeway is divided into I subsections or cells (see Fig. 3.1). Each cell has an external incoming flow $r_{i,t}$ from an on-ramp i to the freeway and an external outgoing flow $s_{i,t}$ from the freeway to an off-ramp i at time step t, and the flow $f_{i,t}$ and density $\rho_{i,t}$ in each cell i at each time step t can be used to characterize the traffic flow dynamics. Let $f_{i-1,t}$ denote the traffic inflow to downstream cell i at each simulation time step t and, hence $f_{i,t}$ (inflow to downstream cell $i+1$) denotes the traffic outflow from cell i at the same simulation time step t. Based on the conservation equation, the evolution

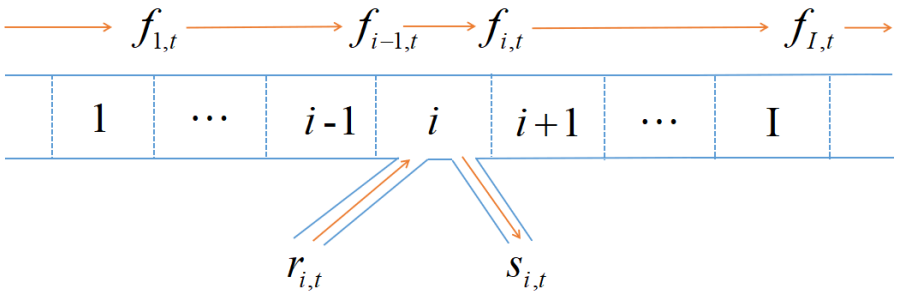


Figure 3.1: Schematic diagram

of density in cell i is described as follows:

$$\rho_{i,t+1} = \rho_{i,t} + \frac{\Delta t}{\Delta x_i}(f_{i-1,t} - f_{i,t} + r_{i,t} - s_{i,t}), \tag{3.1}$$

where Δx_i and Δt are the length of the cell i and the size of the simulation time step t, respectively. Depending on the network topology, some terms of Equation (3.1) may not be present. In particular, the inflow $f_{0,t}$ does not exist for the first cell of the network, the inflow $r_{i,t}$ does not exist for the cell without an on-ramp, while the outflow $s_{i,t}$ exists only for the cell with an off-ramp. It is noted that the time step Δt is defined such that $\Delta t \leq \min_i(\Delta x_i / v_i)$, which is the smallest ratio of the cell length Δx_i to the corresponding free flow velocity v_i on the freeway. The condition is used in traffic flow modeling to guarantee the numerical stability and nonnegativity of traffic quantities by limiting the distance traveled by vehicles in one simulation time step to no more than the length of the cell.

In the case of a given cell density, the outflow from the cell i during the time step t is controlled by the min(.) function as follows:

$$f_{i,t} = \min\{v_i \rho_{i,t}, C_i, C_{i+1}, w_{i+1}(\rho_{max,i+1} - \rho_{i+1,t})\}, \tag{3.2}$$

where C_i represents the capacity flow of cell i and C_{i+1} denotes the capacity flow of cell $i+1$. Due to the heterogeneous segments with different capacities at different locations, we consider both capacity flows in adjacent cells. Furthermore, v_i represents the free flow velocity of cell i, w_{i+1} denotes the backward wave speed of cell $i+1$ and can be obtained from the equality $w_{i+1} = C_{i+1}/(\rho_{max,i+1} - \rho_{cr,i+1})$, where $\rho_{max,i+1}$ is the jam density and $\rho_{cr,i+1}$ corresponds to the critical density and can be derived as $\rho_{cr,i+1} = C_{i+1}/v_{i+1}$. Define $y^d_{i,t} = \min\{v_i \rho_{i,t}, C_i\}$ and $y^s_{i+1,t} = \min\{C_{i+1}, w_{i+1}(\rho_{max,i+1} - \rho_{i+1,t})\}$, where $y^d_{i,t}$ denotes the demand function corresponding to the maximum outflow from cell i at the time step t, and $y^s_{i+1,t}$ is the supply function corresponding to the maximum flow received by cell $i+1$

at the same time step t. Note that variables v_i, C_i, w_i, $\rho_{max,i}$, $\rho_{cr,i}$ represent the model parameters of CTM, which can be calibrated using collected data by loop detectors [130]. Based on the results of [73, 76], equation (3.2) is reformulated as a linear programming problem, which is the key of the optimization model presented in the next section.

3.3 Optimization Model

In this section, we firstly review the CTM-based deterministic optimization model adopted by [73, 76, 118], and then rewrite it as a distributionally robust chance constrained problem with consideration of the uncertain demand flows.

3.3.1 Deterministic Optimization Model

We rewrite the CTM-based freeway optimization model as follows:

$$\text{(DOM)} \min_r D = \sum_{i=1}^{I}\sum_{t=1}^{T}(\rho_{i,t}\Delta x_i\Delta t - \frac{f_{i,t}\Delta x_i\Delta t}{v_i}) + \sum_{j=1}^{J}\sum_{t=1}^{T} q_{j,t}\Delta t, \tag{3.3}$$

$s.t.$

$$\rho_{i,t+1} = \rho_{i,t} + \frac{\Delta t}{\Delta x_i} \times (f_{i-1,t} - f_{i,t} + r_{i,t} - s_{i,t}), \forall i, t, \tag{3.4}$$

$$f_{i,t} \leq v_i\rho_{i,t}, \forall i, t, \tag{3.5}$$

$$f_{i,t} \leq C_i, \forall i, t, \tag{3.6}$$

$$f_{i,t} \leq C_{i+1}, \forall i, t, \tag{3.7}$$

$$f_{i,t} \leq w_{i+1}(\rho_{max,i+1} - \rho_{i+1,t}), \forall i, t, \tag{3.8}$$

$$q_{j,t+1} = q_{j,t} + (d_{j,t} - r_{j,t})\Delta t, \forall j, t, \tag{3.9}$$

$$q_{j,t} \leq q_{max,j}, \forall j, t. \tag{3.10}$$

$$0 \leq r_{j,t} \leq r_{max,j}, \forall j, t. \tag{3.11}$$

To seek the optimal ramp metering r, the optimization problem above is to minimize the total delay D of the system over cells $i = 1, 2, \ldots, I$ and time $t = 1, 2, \ldots, T$. The objective function D includes the total delay in both mainline and on-ramps, where $\rho_{i,t}\Delta x_i \Delta t - \frac{f_{i,t}\Delta x_i \Delta t}{v_i}$ represents the mainline delay in cell i at time t (see [118]), and $q_{j,t}$ denotes the queue length on the on-ramp j at time t. The constraint set (equations (3.4)-(3.8)) is equivalent to the CTM, as stated by [73, 76, 131]. Both constraints (3.5) and (3.6) specify the demand limitations when the flow is under the free flow condition, whereas the constraints (3.7) and (3.8) characterize the supply limitations when the flow is under congested condition. We assume that the exit flow $s_{i,t}$ is given throughout the paper. Constraint (3.9) characterizes the evolution of queues $q_{j,t}$ on the on-ramps $j = 1, 2, \ldots, J$ at time step t, where J denotes the total number of on-ramps. We use $d_{j,t}$ to denote the variable of the demand flow intending to enter the freeway from on-ramp j at time step t and $r_{j,t}$ to denote the actual incoming demand flow entering the freeway from on-ramp j at time step t. Constraint (3.10) is used to govern the maximum queue size on the on-ramps avoiding that the unacceptably long queue on the on-ramps will be adopted as an optimal solution. Finally, the constraint (3.11) gives the lower and upper bounds of the control variable r. The optimal ramp metering r is used to obtain the optimal control, which can also be realized through the hard shoulder running as well as the mainline speed control [132].

3.3.2 Distributionally Robust Chance Constrained Optimization Model

Before presenting the distributionally robust chance constrained optimization model, we summarize the state-of-the-art of work in solving chance constraints in Table 3.1. By comparison, we adopt the approach proposed by Zymler in [128] to approximate distributionally robust chance constraints because of its established theoretical analysis and computational efficiency.

The deterministic optimization model (DOM) can be extended to the distributionally robust chance constrained optimization model (DRCCOM) with consideration of uncertain demand flows. Since the optimization problem (3.3) is a minimization problem and the constraint (3.9) is the only constraint associated to the traffic demand flows on the source links (such as on-ramps), both constraints

Table 3.1: Methods for solving chance constraints

Type	Handling technique	Disadvantage
Scenario approximation [133]	Using the constraints for the scenario sample points to replace the chance constraints	Prohibitively time consuming
Generator-based approximation [134]	The chance constraints are approximated by CVaR inequalities	The approximated problem might be intractable
Chebyshevs Relaxation [135]	Relaxing the chance constraints by Chebyshevs inequality	Solution is too conservative
Robust chance constraints [128, 136]	Replacing the chance constraints by the Worst-case chance constraints under moment information, which can be transformed into a convex optimization problem with SDP or conic constraints	High computational cost

(3.9) and (3.10) can be rewritten as the following constraints:

$$q_{j,t+1} = q_{j,t} + (\widetilde{d}_{j,t} - r_{j,t})\Delta t, \forall j, t. \tag{3.12}$$

and

$$\mathbb{P}(q_{j,t} - q_{max,j} \leq 0) \geq \epsilon_d, \forall j, t, \tag{3.13}$$

where $\widetilde{d}_{j,t}$ denotes the random demand flow variable on the on-ramp j at time step t and $\epsilon_d \in (0, 1)$ is the confidence parameter. The violation of constraint (3.13) means that the waiting queue length is longer than the maximum queue on source links. Due to the fact that the mean and covariance of uncertain demand flows are given, the chance constraint (3.13) can be rewritten as follows:

$$\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}(q_{j,t} - q_{max,j} \leq 0) \geq \epsilon_d, \forall j, t, \tag{3.14}$$

where $\mathcal{P}$ is the set of all probability distributions.

By equation (3.4), the density $\rho_{i,t}$ can be rewritten as

$$\rho_{i,t} = \rho_{i,1} + \sum_{l=0}^{t-1} \frac{\Delta t}{\Delta x_i}(f_{i-1,l} - f_{i,l} + r_{i,l} - s_{i,l}), \forall i, t. \tag{3.15}$$

Similarly, the queue length $q_{j,t}$ can be reformulated as

$$q_{j,t} = q_{j,1} + \sum_{l=0}^{t-1}(\widetilde{d}_{j,l} - r_{j,l})\Delta t. \tag{3.16}$$

Let $\widetilde{\mathbf{d}}_j = \{0, \widetilde{d}_{j,1}, \widetilde{d}_{j,2}, \ldots, \widetilde{d}_{j,T-1}\}^\top \in \mathbb{R}^T$, and $B_{t-1} = \{\Delta t, \Delta t, \Delta t, \ldots, \Delta t, 0, \ldots, 0\}^\top \in \mathbb{R}^T$. By the above relationship (3.16), the relationships (3.12) and (3.13) can be simplified as follows:

$$\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\{q_{j,1} - \sum_{l=0}^{t-1} r_{j,l}\Delta t - q_{max,j} + B_{t-1}^\top \widetilde{\mathbf{d}}_j \leq 0\} \geq \epsilon_d, \forall j, t. \tag{3.17}$$

Then, the robust ramp metering problem with uncertain demand can be rewritten as:

$$\begin{aligned}(\text{DRCCP}) \min_r D \quad = \quad & \sum_{i=1}^{I}\sum_{t=1}^{T}\big([\rho_{i,1} + \sum_{l=0}^{t-1}\frac{\Delta t}{\Delta x_i}(f_{i-1,l} - f_{i,l} + r_{i,l} - s_{i,l})]\Delta x_i \Delta t \\ & - \frac{f_{i,t}\Delta x_i \Delta t}{v_i}\big) + \mathbb{E}\big[\sum_{j=1}^{J}\sum_{t=1}^{T}(q_{j,1} - \sum_{l=0}^{t-1} r_{j,l}\Delta t + B_{t-1}^\top \widetilde{\mathbf{d}}_j)\Delta t\big],\end{aligned} \tag{3.18}$$

subject to constraints (3.6), (3.7), (3.15), (3.17) and

$$f_{i,t} \leq v_i[\rho_{i,1} + \sum_{l=0}^{t-1}\frac{\Delta t}{\Delta x_i}(f_{i-1,l} - f_{i,l} + r_{i,l} - s_{i,l})], \forall i, t, \tag{3.19}$$

$$f_{i,t} \leq w_{i+1}[\rho_{max,i+1} - \rho_{i+1,1} - \sum_{l=0}^{t-1}\frac{\Delta t}{\Delta x_i}(f_{i,l} - f_{i+1,l} + r_{i+1,l} - s_{i+1,l})], \forall i, t. \tag{3.20}$$

Due to the distributionally robust chance constraint (3.17), we have difficulties to solve the Problem (DRCCP) directly. Thus, we need to transform the problem into a solvable problem, which is presented in the next section.

3.4 Approximation of Distributionally Robust Chance Constraint

An approximation approach proposed by [128] is utilized to approximate the constraint (3.17) in this section.

We let $\mu_j \in \mathbb{R}^T$ be the mean vector and $\Sigma_j \in \mathbb{S}^T$ be the covariance matrix of the random demand flow vector $\widetilde{\mathbf{d}}_j$ under true distribution $\mathbb{P}$ throughout this paper. Thus, we implicitly assume that $\mathbb{P}$ has finite second-order moments. Without loss of generality, we assume that $\Sigma_j \succ 0$. To simplify the notation, we let

$$\Omega_j = \begin{bmatrix} \Sigma_j + \mu_j^\top \mu_j & \mu_j^\top \\ \mu_j & 1 \end{bmatrix} \tag{3.21}$$

denote the second-order moment matrix of $\widetilde{\mathbf{d}}_j$. Chen et al. [129] has proved that the constraint (3.17) can be approximated by the Worst-Case CVaR constraint. Thus, we have

$$\begin{aligned} & R(\alpha_{j,t}) \\ = \; & \big\{(f,r) : \sup_{\mathbb{P}\in\mathcal{P}} \mathrm{CVaR}_{1-\epsilon_d}(q_{j,1} - \sum_{l=0}^{t-1} r_{j,l}\Delta t - q_{max,j} + B_{t-1}^\top \widetilde{\mathbf{d}}_j) \le 0\big\}, \forall j,t, \end{aligned} \tag{3.22}$$

where

$$\begin{aligned} & \mathrm{CVaR}_{1-\epsilon_d}(q_{j,1} - \sum_{l=0}^{t-1} r_{j,l}\Delta t - q_{max,j} + B_{t-1}^\top \widetilde{\mathbf{d}}_j) \\ = \; & \inf_{\alpha_{j,t}\in\mathbb{R}} \big\{\alpha_{j,t} + \frac{1}{1-\epsilon_d}\mathbb{E}_{\mathbb{P}}\big[(q_{j,1} - \sum_{l=0}^{t-1} r_{j,l}\Delta t - q_{max,j} + B_{t-1}^\top \widetilde{\mathbf{d}}_j - \alpha_{j,t})^+\big]\big\}, \forall j,t, \end{aligned} \tag{3.23}$$

where $\alpha_{j,t}$ is the decision variable in chance constraints, $\mathbb{E}_{\mathbb{P}}(\bullet)$ denotes the expectation of the distribution $\mathbb{P}$, and $(\bullet)^+ = \max\{\bullet, 0\}$ (see [137, 138] for more details).

An approximation approach proposed in [128] based on semidefinite programming (SDP) is utilized to approximate the constraint (3.22). By supposing that the mean and covariance matrix of stochastic variables are available, Zymler et

al. [128] firstly used the Worst-case Conditional Value-at-Risk (WCVaR) constraints to approximate distributionally robust chance constraints, and then the WCVaR constraints were reformulated into SDP constraints. The results indicate that the approximation is exact when the roust individual chance constraint is a concave or quadratic function. In this work, the approximation approach proposed in [128] is adopted to approximate the constraint (3.17) and the equivalent form of constraint (3.22) is presented as the following theorem.

Theorem 3.1. *If the demand flow $\widetilde{\mathbf{d}}_j$ follows an unknown probability distribution with given mean μ_j and covariance matrix Σ_j, then the constraint (3.17) can be approximated as follows:*

$$R(\alpha_{j,t}) = \left\{ (f,r): \begin{array}{l} \exists(\alpha_{j,t}, \mathbf{A}_{j,t}) \in \mathbb{R} \times \mathbb{S}^{T+1} \\ \alpha_{j,t} + \dfrac{1}{1-\epsilon_d}\langle \Omega_j, \mathbf{A}_{j,t}\rangle \leq 0, \mathbf{A}_{j,t} \succeq \mathbf{0} \\ \mathbf{A}_{j,t} - \begin{bmatrix} \mathbf{0} & \dfrac{B_{t-1}^{\top}}{2} \\ \dfrac{B_{t-1}}{2} & q_{j,1} - \sum_{l=0}^{t-1} r_{j,l}\Delta t - q_{max,j} - \alpha_{j,t} \end{bmatrix} \succeq \mathbf{0} \end{array}, \forall j,t \right\},$$

where $\mathbf{A}_{j,t} \in \mathbb{S}^{T+1}$ is the $T+1$-dimensional real symmetric matrices, $\langle \Omega_j, \mathbf{A}_{j,t}\rangle = trace(\Omega_j, \mathbf{A}_{j,t})$, which denotes a trace scalar product of matrices Ω_j and $\mathbf{A}_{j,t}$, and $\mathbf{A}_{j,t} \succeq \mathbf{0}$ implies that the matrix $\mathbf{A}_{j,t}$ is semidefinite.

Proof of Theorem 1: It is noted that the constraint (3.22) can be equivalently expressed as $J(f, r, \alpha_{j,t}) \leq 0$, where

$$\begin{aligned} J(f,r,\alpha_{j,t}) &= \sup_{\mathbb{P}\in\mathcal{P}} \mathrm{CVaR}_{1-\epsilon_d}(q_{j,1} - \sum_{l=0}^{t-1} r_{j,l}\Delta t - q_{max,j} + B_{t-1}^{\top}\widetilde{\mathbf{d}}_j) \\ &= \sup_{\mathbb{P}\in\mathcal{P}} \inf_{\alpha_{j,t}\in\mathbb{R}} \big\{ \alpha_{j,t} + \frac{1}{1-\epsilon_d}\mathbb{E}_{\mathbb{P}}\big[(q_{j,1} - \sum_{l=0}^{t-1} r_{j,l}\Delta t - q_{max,j} + B_{t-1}^{\top}\widetilde{\mathbf{d}}_j \\ &\quad -\alpha_{j,t})^{+}\big]\big\}. \end{aligned} \tag{3.24}$$

By the stochastic saddle point theorem [139], the maximization and minimization operations can be interchanged as follows:

$$\begin{aligned} J(f,r,\alpha_{j,t}) &= \inf_{\alpha_{j,t}\in\mathbb{R}} \big\{ \alpha_{j,t} + \frac{1}{1-\epsilon_d} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\big[(q_{j,1} - \sum_{l=0}^{t-1} r_{j,l}\Delta t - q_{max,j} + B_{t-1}^{\top}\widetilde{\mathbf{d}}_j \\ &\quad -\alpha_{j,t})^{+}\big]\big\}. \end{aligned} \tag{3.25}$$

Next, a SDP reformulation of the following worst-case expectation problem can be derived:

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[(q_{j,1} - \sum_{l=0}^{t-1} r_{j,l}\Delta t - q_{max,j} + B_{t-1}^{\top}\widetilde{\mathbf{d}}_j - \alpha_{j,t})^{+}\right], \tag{3.26}$$

which can be regarded as the subordinate maximization problem in (3.25). Based on the Lemma 1 in [128], we obtain

$$\begin{aligned} &\inf_{\mathbf{A}_{j,t}\in\mathbb{S}^{T+1}} \langle \Omega_j, \mathbf{A}_{j,t}\rangle \\ \text{s.t.} \quad &\mathbf{A}_{j,t} \succeq \mathbf{0}, \\ &[\widetilde{\mathbf{d}}_j^{\top} \quad 1]\mathbf{A}_{j,t}[\widetilde{\mathbf{d}}_j^{\top} \quad 1]^{\top} \geq q_{j,1} - \sum_{l=0}^{t-1} r_{j,l}\Delta t - q_{max,j} + B_{t-1}^{\top}\widetilde{\mathbf{d}}_j - \alpha_{j,t}, \\ &\qquad\qquad \forall \widetilde{\mathbf{d}}_j \in \mathbb{R}^T, j \in J, t \in T. \end{aligned} \tag{3.27}$$

The constraint (3.27) can be written as follows:

$$\begin{aligned} &[\widetilde{\mathbf{d}}_j^{\top} \quad 1]\mathbf{A}_{j,t}[\widetilde{\mathbf{d}}_j^{\top} \quad 1]^{\top} \\ \geq \quad & q_{j,1} - \sum_{l=0}^{t-1} r_{j,l}\Delta t - q_{max,j} + B_{t-1}^{\top}\widetilde{\mathbf{d}}_j - \alpha_{j,t}, \forall \widetilde{\mathbf{d}}_j \in \mathbb{R}^T, j \in J, t \in T. \end{aligned} \tag{3.28}$$

Furthermore, constraint (3.28) can be equivalently expressed as

$$\mathbf{A}_{j,t} - \begin{bmatrix} \mathbf{0} & \frac{B_{t-1}}{2} \\ \frac{B_{t-1}^{\top}}{2} & q_{j,1} - \sum_{l=0}^{t-1} r_{j,l}\Delta t - q_{max,j} - \alpha_{j,t} \end{bmatrix} \succeq \mathbf{0}, \forall j, t. \tag{3.29}$$

Therefore, the worst-case expectation problem (3.26) can be reformulated into

$$\begin{aligned} &\inf_{\mathbf{A}_{j,t}\in\mathbb{S}^{T+1}} \langle \Omega_j, \mathbf{A}_{j,t}\rangle \\ \text{s.t.} \quad &\mathbf{A}_{j,t} \succeq \mathbf{0}, \\ &\mathbf{A}_{j,t} - \begin{bmatrix} \mathbf{0} & \frac{B_{t-1}}{2} \\ \frac{B_{t-1}^{\top}}{2} & q_{j,1} - \sum_{l=0}^{t-1} r_{j,l}\Delta t - q_{max,j} - \alpha_{j,t} \end{bmatrix} \succeq \mathbf{0}, \forall j, t. \end{aligned} \tag{3.30}$$

Substituting (3.30) into (3.25) yields

$$\begin{aligned} J(f,r,\alpha_{j,t}) \quad = \quad & \inf_{\alpha_{j,t}\in\mathbb{R}} \ \alpha_{j,t} + \frac{1}{1-\epsilon_d}\langle \Omega_j, \mathbf{A}_{j,t}\rangle \\ \text{s.t.} \quad & \mathbf{A}_{j,t} \in \mathbb{S}^{T+1}, \mathbf{A}_{j,t} \succeq \mathbf{0}, \\ & \mathbf{A}_{j,t} - \begin{bmatrix} \mathbf{0} & \frac{B_{t-1}}{2} \\ \frac{B_{t-1}^\top}{2} & q_{j,1} - \sum_{l=0}^{t-1} r_{j,l}\Delta t - q_{max,j} - \alpha_{j,t} \end{bmatrix} \succeq \mathbf{0}, \forall j,t, \end{aligned} \tag{3.31}$$

and thus the proof of Theorem 3.1 is completed.

Based on Theorem 3.1, the optimization problem (DRCCP) can be reformulated into a problem with SDP constraints as follows:

$$\begin{aligned} (\text{DRCCOM}) \min_{f,r,\alpha,\mathbf{A}} D \quad = \quad & \sum_{i=1}^{I}\sum_{t=1}^{T}\big([\rho_{i,1} + \sum_{l=0}^{t-1}\frac{\Delta t}{\Delta x_i}(f_{i-1,l} - f_{i,l} + r_{i,l} - s_{i,l})]\Delta x_i \Delta t \\ & - \frac{f_{i,t}\Delta x_i \Delta t}{v_i}\big) + \mathbb{E}\big[\sum_{j=1}^{J}\sum_{t=1}^{T}(q_{j,1} - \sum_{l=0}^{t-1} r_{j,l}\Delta t \\ & + B_{t-1}^\top \widetilde{\mathbf{d}}_j)\Delta t\big], \end{aligned} \tag{3.32}$$

subject to constraints (3.6), (3.7), (3.15), (3.19), (3.20) and

$$\alpha_{j,t} + \frac{1}{1-\epsilon_d}\langle \Omega_j, \mathbf{A}_{j,t}\rangle \leq 0, \forall j,t, \tag{3.33}$$

$$\mathbf{A}_{j,t} - \begin{bmatrix} \mathbf{0} & \frac{B_{t-1}}{2} \\ \frac{B_{t-1}^\top}{2} & q_{j,1} - \sum_{l=0}^{t-1} r_{j,l}\Delta t - q_{max,j} - \alpha_{j,t} \end{bmatrix} \succeq \mathbf{0}, \forall j,t, \tag{3.34}$$

$$\mathbf{A}_{j,t} \succeq \mathbf{0}, \mathbf{A}_{j,t} \in \mathbb{S}^{T+1}, \forall j,t. \tag{3.35}$$

Based on the result in [128], we get that DRCCOM can be solved efficiently.

3.5 Case Study

We select a 13 km road of the Kwinana Freeway in the vicinity of Perth in Australia (refer to Fig. 3.2). We discretize the section into 26 cells with 500 m for each cell. The road section covers eight on-ramps and four off-ramps, and is one of the busiest sections in Perth. We select the duration from 6 : 00am to 10 : 00am, which represents the peak hours. The on-ramps are located at cells $2, 5, 8, 9, 10, 16, 17, 25$ and the off-ramps are located at cells $3, 7, 15, 26$. The data of flow and density collected from loop detectors are used to obtain the piecewise linear fundamental diagram for each cell. We use the least square method to compute model parameters and summarize them in Table 3.2.

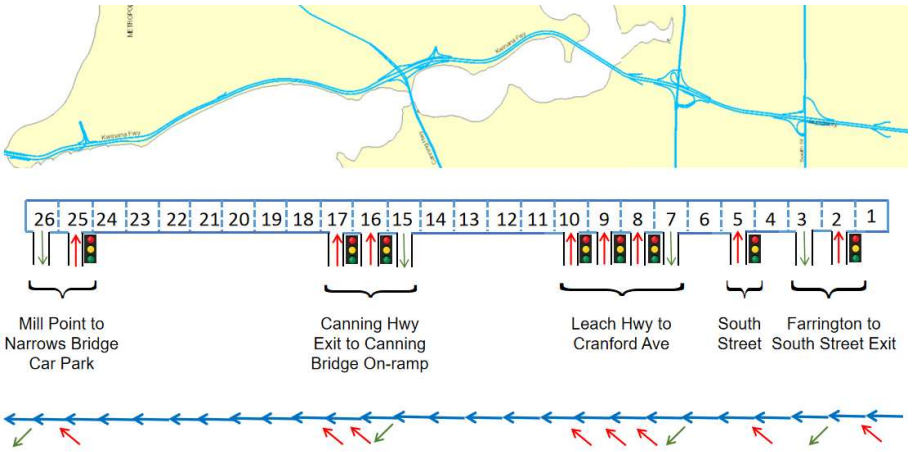


Figure 3.2: Road map

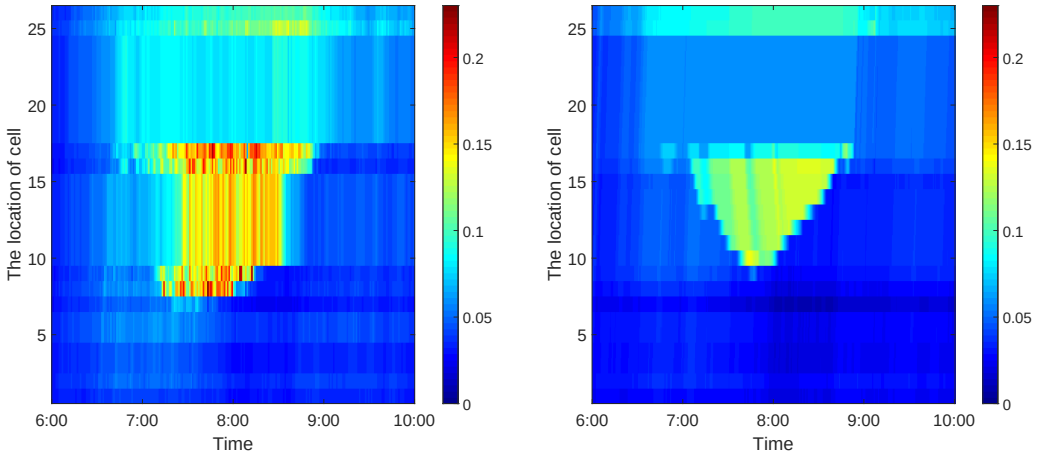
3.5.1 Results from the Deterministic Model

The objective of deterministic optimization model (DOM) is to find an optimal ramp metering by minimizing the total travel delay with consideration of deterministic demand flows under triangular fundamental diagrams. We use MATLAB R2019a with SeDuMi [140] solver and YALMIP interface [141] to solve the proposed models in numerical application. For the no control model and DOM, we show the corresponding results on the June 11, and June 13, 2018, in Figs. 3.3-3.4, respectively. The bar on the right side shows the size of mainline density that increases from bottom to top. The lighter color in the figures implies smaller value of density and better traffic condition and vis versa. For the no control model on June 11, 2018, the total system delay with zero ramp delay is 1074.6552veh-hr and the total system delay for DOM is 488.8045veh-hr and the associated ramp delay is 174.8159veh-hr. On June 13, 2018, the total system delay for no

Table 3.2: Model parameters

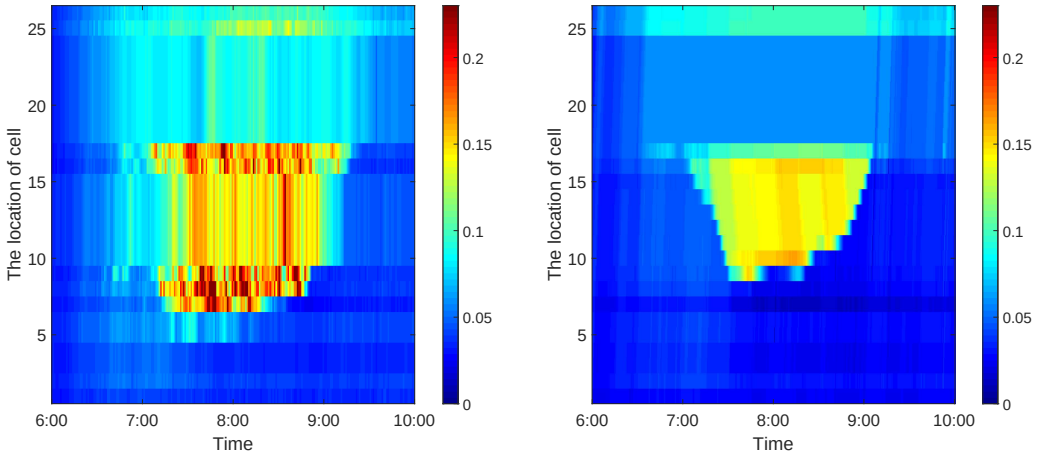
Name of parameter	Value
$v_i, i = 1, \ldots, 24$	27.7778m/s
$v_i, i = 25, 26$	22.2222m/s
$w_i, i = 1, 3, 4, 6, \ldots, 24$	9.8029m/s
$w_i, i = 2, 5$	9.7895m/s
$w_i, i = 25, 26$	10.7830m/s
$C_i, i = 1, 3, 4, 6, \ldots, 24$	1.6667veh/s
$C_i, i = 2, 5, 25, 26$	2.2222veh/s
$q_{max,j}, j = 7$	120veh
$q_{max,j}, j = 1, \ldots, 6, 8$	60veh
$r_{max,j}, j = 1, \ldots, 8$	0.5500veh/s
$\rho_{max,i}, i = 1, 3, 4, 6, \ldots, 24$	0.2300veh/m
$\rho_{max,i}, i = 2, 5, 25, 26$	0.3067veh/m
Δt	15s
$\Delta x_i, i = 1, \ldots, 26$	500m

control with zero ramp delay is 1583.9869veh-hr and the total system delay for DOM is 1105.5529veh-hr and the associated ramp delay is 474.3417veh-hr. By comparison, we can see that there are 54.5152% and 30.2045% improvement for total delay, respectively. Therefore, ramp metering can reduce the congestion on freeway and is an effective strategy to improve freeways operations.



(a) Actual density (veh/m) on June 11, 2018 (b) Density (veh/m) for DOM on June 11, 2018

Figure 3.3: Traffic density (veh/m) for no control and DOM on June 11, 2018



(a) Actual density (veh/m) on June 13, 2018 (b) Density (veh/m) for DOM on June 13, 2018

Figure 3.4: Traffic density (veh/m) for no control and DOM on June 13, 2018

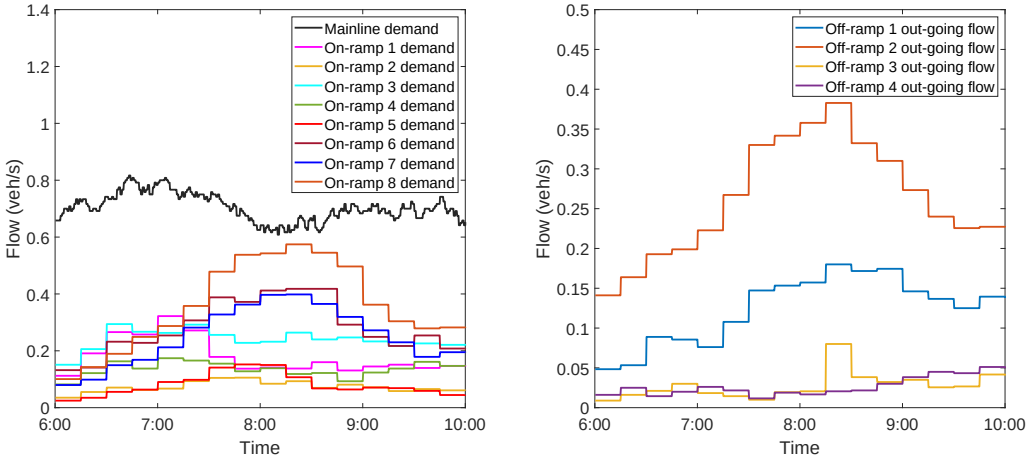
3.5.2 Results from the Disributionally Robust Chance Constrained Optimization Model

Now, we show the corresponding results of the disributionally robust chance constrained optimization model (DRCCOM), where we take into account ten scenarios (refer to Table 3.3 for more details) based on the measured data of the demand flows. On the second column in Table 3.3, "1.00" denotes the situation where the demand flow $\widetilde{d}_{j,t}$ is set as the mean demand flow $d_{j,t}^{mean}$; "0.96" denotes the situation where the demand flow $\widetilde{d}_{j,t}$ is reduced to $0.96d_{j,t}^{mean}$; "1.05" refers to the situation in which the demand flow $\widetilde{d}_{j,t}$ is increased to $1.05d_{j,t}^{mean}$. A demand multiplier more than 1 refers to the situation where the actual demand flow is being underestimated and vis versa. Fig. 3.5 shows the mean mainline demand, mean on-ramp demands and mean out-going flows.

For comparison, the robust optimization method (ROM) proposed by Chow and Li [118] is adopted in this paper, where we suppose that there is an uncertainty of ± 0.05 associated with the demand flows based on the collected data from detectors and utilize triangular fundamental diagrams. This gives $\widetilde{d}_{j,t}^{min} = 0.95 * \widetilde{d}_{j,t}$ and $\widetilde{d}_{j,t}^{max} = 1.05 * \widetilde{d}_{j,t}$. First, we take into account the scenario 5, i.e., letting $\widetilde{d}_{j,t} = d_{j,t}^{mean}$. Table 3.4 shows the corresponding results. By analysing, we can see that the performance of DRCCOM is the best when $\epsilon_d = 0.95$. For this case, the total delay and ramp delay for the DRCCOM are 731.6660veh-hr and 215.0415veh-hr, respectively. Compared to ROM, there are, respectively, 0.2442% and 0.0670% improvement. Compared with DOM, the total delay of DRCCOM

Table 3.3: Ten scenarios based on the measured data of the demand flows

Case	Demand flows	control strategy
1	$0.96*\widetilde{d}_{j,t}$	deterministic, robust, distributionally robust
2	$0.97*\widetilde{d}_{j,t}$	deterministic, robust, distributionally robust
3	$0.98*\widetilde{d}_{j,t}$	deterministic, robust, distributionally robust
4	$0.99*\widetilde{d}_{j,t}$	deterministic, robust, distributionally robust
5	$1.00*\widetilde{d}_{j,t}$	deterministic, robust, distributionally robust
6	$1.01*\widetilde{d}_{j,t}$	deterministic, robust, distributionally robust
7	$1.02*\widetilde{d}_{j,t}$	deterministic, robust, distributionally robust
8	$1.03*\widetilde{d}_{j,t}$	deterministic, robust, distributionally robust
9	$1.04*\widetilde{d}_{j,t}$	deterministic, robust, distributionally robust
10	$1.05*\widetilde{d}_{j,t}$	deterministic, robust, distributionally robust



(a) Mean mainline demand and mean on-ramp demands

(b) Mean out-going flows

Figure 3.5: Mean mainline demand, mean on-ramp demands and mean out-going flows

is less, but the ramp delay is more than that of DOM. The performances of total delay of DRCCOM when $\epsilon_d = 0.90$ or $\epsilon_d = 0.97$ do not outperform those of DOM and ROM. Furthermore, we can seen that the ramp delay is increasing with the value of ϵ_d decreasing, which is consistent with our theoretical analyses because smaller ϵ_d implies more probability of violation of constraint in terms of that the waiting queue length is longer than the maximum queue on an on-ramp.

Table 3.4 also shows the total delay and ramp delay over 10 scenarios for the three different control strategies. We can see that the total delay and ramp delay increase with the demand increasing for the three control strategies. In particular, the total delay of DRCCOM when $\epsilon = 0.95$ is less than those of ROM for the

Table 3.4: Total delay and ramp delay for ten scenarios

Demand	DOM Total delay(veh-hr) Ramp delay(veh-hr)	ROM Total delay(veh-hr) Ramp delay(veh-hr)	DRCCOM with $\epsilon_d = 0.90$ Total delay(veh-hr) Ramp delay(veh-hr)	DRCCOM with $\epsilon_d = 0.95$ Total delay(veh-hr) Ramp delay(veh-hr)	DRCCOM with $\epsilon_d = 0.97$ Total delay(veh-hr) Ramp delay(veh-hr)
$0.96*d_{j,t}^{mean}$	332.6767	338.9177	334.7076	328.2172	329.8285
	40.0145	48.1446	48.5633	45.3767	42.9587
$0.97*d_{j,t}^{mean}$	411.3381	420.2148	417.3771	410.2358	413.6383
	82.0282	94.3712	89.4254	85.7465	83.1289
$0.98*d_{j,t}^{mean}$	502.4933	500.5120	503.7418	498.2489	500.9168
	102.9199	102.9337	110.8222	105.0563	102.9788
$0.99*d_{j,t}^{mean}$	599.4085	602.6364	605.5120	598.6296	603.5045
	185.6659	208.0696	202.9336	196.4971	186.2164
$1.00*d_{j,t}^{mean}$	734.4483	733.4570	736.4828	731.6660	733.4828
	213.7042	216.4920	218.5041	215.0415	214.7027
$1.01*d_{j,t}^{mean}$	909.7404	906.5402	915.7383	896.2559	907.6203
	346.3851	357.6635	366.2991	356.7829	350.5417
$1.02*d_{j,t}^{mean}$	1174.9983	1081.2249	1168.0312	1084.1379	1164.1543
	397.6578	446.8358	429.1136	418.0353	408.7932
$1.03*d_{j,t}^{mean}$	1338.1348	1336.3029	1344.1208	1304.6056	1321.0637
	503.1153	564.3454	592.6797	551.3854	529.0421
$1.04*d_{j,t}^{mean}$	1569.6749	1521.8225	1581.2552	1555.3102	1571.8093
	629.6253	659.5434	663.7006	658.4050	646.6241
$1.05*d_{j,t}^{mean}$	1845.0960	1805.6598	1838.0398	1798.5056	1824.0285
	727.3990	725.9099	787.2781	765.9304	743.8254

cases that $\widetilde{d}_{j,t} \neq 1.02 * d_{j,t}^{mean}, 1.04 * d_{j,t}^{mean}$. Compared to DOM, the total delay of DRCCOM when $\epsilon = 0.95$ is less than those of DOM for all the 10 different scenarios, while the ramp delay of DRCCOM shows a opposite trend. Moreover, the performances of total delay and ramp delay of DOM are better than those of DRCCOM when $\epsilon = 0.90$ for most cases. When $\epsilon = 0.97$, the total delay of DRCCOM is less than those of DOM for most cases, but the ramp delay is longer than DOM. This is reasonable because the objective function is to minimize total delay rather than ramp delay. When more vehicles are holding on on-ramps, total delay may reduce but ramp delay will increase.

3.6 Conclusion

A distributionally robust chance constrained optimization model is presented in this chapter to address the ramp metering problem with uncertain demand flows. The model is formulated as a semidefinite programming using the Worst-Case Conditional Value-at-Risk (WCVaR) constraints to approximate distributionally robust chance constraints. Given partial information (such as mean and covariance matrix) of stochastic demand flows and the triangular fundamental diagrams, an optimal ramp metering strategy can be obtained by minimizing the

total delay of mainline and on-ramps. The associated results of three different ramp metering strategies, i.e., deterministic optimization model, robust optimization model and distributionally robust chance constrained optimization model, are presented and compared. The results show that, considering the uncertainty encountered by the system over a range of scenarios, distributionally robust optimization is an effective and useful method to control the total delay of the system and mitigate traffic congestions. The results also show that the distributionally robust optimization is more effective in managing the total system delay.

CHAPTER 4

A Smoothing Method for Ramp Metering

4.1 Introduction

In this Chapter, we propose an optimization program for freeway dynamic ramp metering based on CTM. This problem has been formulated as a discrete time optimal control problem with smooth state equations and constraints to meter traffic inflow from on-ramps. In the proposed model, the 'min' operators in the primal CTM are non-differentiable and thus, the corresponding optimal control problem cannot be solved directly using conventional gradient based methods. In this work, we introduce a smooth approximation to approximate the 'min' operators and then a unified computational approach is developed to solve the problem. Theoretical analysis is carried out, showing that the optimal solution obtained from the approximated problem converges to the optimal solution of the primal CTM. Compared to the classical inequality relaxation method, our method can resolve the flow holding-back problem and reduce under fundamental diagram phenomenon. Compared with the Big-M method, our method has better efficiency. To achieve the desired traffic response control in real application, a series of online optimal control problems are solved using MPC. Simulation studies show that our method can significantly improve freeway traffic management efficiency.

Queue-spill overs decrease the capacity of freeway in handling vehicles, thus wasting commuting time in congested traffic conditions, especially during peak hours. Ramp metering is an efficient way to protect congested areas of the freeway from oversaturated flow conditions. ALINEA, which is one of the first ramp metering controllers, is a feedback control of integral type [142]. Based on the ALINEA, a proportional integral regulator known as Pi-ALINEA [143, 144] was then proposed. Tracking a reference point for the traffic density (or occupancy) is the main purpose of these regulators. There are also some other rule-based control systems, for example coordinated control HERO [145], and others (see [146,147]).

In addition to these simple regulators, many complex controllers have been designed based on optimization approaches. To alleviate the freeway traffic congestion, different cost functions, for maximising the system throughput or minimizing the total travel time spent by drivers in a traffic network system, are proposed. One of the most widely adopted optimization-based models in freeway control with ramp metering design is the macroscopic model, including the first order models [66, 67]) and the second order models [113]. The second order models have obvious advantages because they can show the phenomenon of capacity drop. However, compared with first-order models, they contain more parameters that are required to be customized. To date, the constructed optimization problems based on these models have only been solved in the sense of local optimality [148, 149] due to the complexity of these optimization problems.

Various formulations and solution techniques have been proposed for the optimal ramp metering problems. The seminal research of optimization-based ramp metering can be traced back to [109], where a static model of traffic behavior was utilized for the formulation of the problem. This model has been extended in various forms in [112]. In particular, a deterministic ramp metering optimization program was proposed in [73] based on the Asymmetric Cell transmission model (ACTM), where it was demonstrated that the solution to the linear relaxation problem is feasible for freeway segments with only on-ramp and off-ramp junctions.

In [150], the Link-Node Cell Transmission Model (LN-CTM) was used to reformulate this problem as a linear program under relaxed piecewise-affine fundamental diagrams. Similarly, the 'min' operators were relaxed in [151] to obtain a feasible solution by employing traffic demand control in each cell. Particulary, if the fundamental diagrams are assumed to be symmetric triangular, meaning that the backward wave speed equals to the free flow speed of each cell, the solution of the relaxed problem is feasible. In [120], a novel first-order multi-lane macroscopic traffic flow model was first proposed for motorways, which considers the changes of lateral flow and longitudinal flow. In [121], the model proposed in [120] was formulated as a discrete time optimal control problem with linear relaxation through the use of ramp metering and variable speed limits. Recently, a traffic network finite-horizon optimal control model with exact linear relation for ramp metering controls and variable speed limits is proposed in [152], for which the distributed alternating direction method of multipliers (ADMM) is used to optimize the proposed model. A decentralized MPC approach is proposed in [153] for the freeway system on lossy communication networks under the mainline de-

mand control. The priority parameter is used to ensure the flow of vehicles from both cells.

In [154], a Modified Cell Transmission Model (MCTM) is formulated as a Linear Complementarity System (lCS) which can be efficiently solved, thus successfully avoiding the hard nonlinearity caused by the 'min' operators. In [155], it is shown that if the objective is to minimize the total time spent and the turning rates are invariant, then through the use of ramp metering and partial mainline demand control, convex relaxation can be used to accurately obtain the optimal solution of the original problem by introducing an alternative representation system.

Literature review shows that many ramp metering studies used CTM for freeway network loading. According to their approaches of reducing the complexity of the problem, the papers mentioned above can be divided into two categories: (i) The 'min' operators are directly replaced by a set of linear inequality constraints, where the effect of the flow holding-back problem is ignored; and (ii) The 'min' operators are relaxed through the use of inequality constraints under specific assumptions, such as symmetric triangular fundamental diagrams, or combined with traffic demand control; and (iii) The original 'min' operators are transformed equivalently to a series of linear inequality constraints using the Big-M method [78] through the introduction of binary variables. Although this transformation is equivalent, the transformed problem is computationally demanding because many auxiliary binary variables are being introduced, and hence it is not possible to rely on online computation. However, even for methods based on the equivalent transformations, they are specific methods and hard to calculate, and are hard to be extended to solve general problems. Based on the above discussion, these methods either cannot ensure the optimality of the solution obtained or lack the generality.

The rest of the chapter is organised as follows. Section 4.2 formulates and models the optimization problem of freeway ramp metering. In Section 4.3, our approximate model is proposed and the main convergent results are established. Solution techniques are developed in Section 4.4. Sections 4.5 shows the numerical studies and Section 4.6 gives the conclusion.

4.2 Optimization problem

In this section, we will formulate the traffic flow dynamic as an optimal control problem to optimize the inflow from on-ramps to the freeway. Dynamic inflows

from on-ramps are the main control variables over the study period. The aim is to improve freeway network performance by regulating the number of vehicles to enter the freeway. The problem is formulated and modeled based on the proposed CTM [66, 67]. The CTM is a numerical method developed based on the space-time discretization of the hydrodynamic traffic flow model [8]. See, for example, [156, 157]. In the formulation of CTM, each freeway is discretized into several homogenous segments, called subsections or cells, and the traffic flow is analysed in each cell through discretize time steps.

We summary in Table 4.1 the definitions of all the parameters, sets and decision variables used in this study.

Table 4.1: Models' variables and parameters

Parameters	
Δx_i	length of cell i
Δt	size of the simulation time step t
C_i	capacity of cell i
v_i	free flow speed at cell i
$\rho_{max,i}$	jam density at cell i
w_i	backward wave speed at cell i
$q_{max,j}$	maximum queue length at on-ramp j
$r_{max,j}$	maximum ramp metering at on-ramp j
ϵ	parameter for adjusting approximate error
Sets	
T	set of all time steps
I	set of all cells
J	set of all on-ramps
F	feasible set of Problem PCTM
F^ϵ	feasible set of Problem APCTM
V	convex and compact subset of $\mathbb{R}^J$
$\mathcal{U}$	set of admission controls
Control variable	
$r_{j,t}$	ramp metering at on-ramp j and time step t
Variables	
$\rho_{i,t}$	density at cell i and time step t
$f_{i,t}$	flow at cell i and time step t
$d_{j,t}$	demand flow to enter freeway from on-ramp j at time step t
$q_{j,t}$	queue length waiting for on-ramp j at time step t

4.2.1 Primal cell transmission model

The proposed cost function of the optimization problem is to optimize the total delay, including the mainline delay and the ramp delay. As shown in [118], this cost function is suitable for the oversaturation of the freeway system, and it tries to allow as many vehicles as possible to reach the destinations. The problem can be formally stated as follows:

$$\begin{aligned}
(\text{PCTM}) \min_{r} D &= \sum_{t=1}^{T}\sum_{i=1}^{I}(\rho_{i,t}\Delta x_i \Delta t - \frac{f_{i,t}\Delta x_i \Delta t}{v_i}) + \sum_{t=1}^{T}\sum_{j=1}^{J} q_{j,t}\Delta t, && (4.1)\\
\text{s.t.}& \\
\rho_{i,t+1} &= \rho_{i,t} + \frac{\Delta t}{\Delta x_i} \times (f_{i-1,t} - f_{i,t} + r_{i,t} - s_{i,t}), \forall i = 1, \ldots, I, \forall t, && (4.2)\\
f_{i,t} &= \min\{v_i\rho_{i,t}, C_i, C_{i+1}, w_{i+1}(\rho_{max,i+1} - \rho_{i+1,t})\}, \forall i, t, && (4.3)\\
q_{j,t+1} &= q_{j,t} + \Delta t(d_{j,t} - r_{j,t}), \forall j, t, && (4.4)\\
0 &\leq \rho_{i,t} \leq \rho_{max,i}, \forall i, t, && (4.5)\\
0 &\leq q_{j,t} \leq q_{max,j}, \forall j, t, && (4.6)\\
0 &\leq r_{j,t} \leq r_{max,j}, \forall j, t. && (4.7)
\end{aligned}$$

The dynamics of the density in each cell i at time step t is governed by (4.2). The outflow from the cell i during the time step t is controlled by a piecewise linear fundamental diagram described in (4.3). The dynamics of the queue length for each on-ramp j at time step t is given by (4.4). The upper and lower bounds for the density $\rho_{i,t}$, queue length $q_{j,t}$ and ramp-metering $r_{j,t}$ are given by (4.5)-(4.7), respectively.

Let $V = \{r = [r_1, r_2, \ldots, r_J] \in \mathbb{R}^J : 0 \leq r_j \leq r_{max,j}, \forall j\}$, where $r_{max,j}$ is a given constant. It is noted that V is a compact and convex subset of $\mathbb{R}^J$. Let r be a control sequence $\{r_t : t = 1, \ldots, T-1\}$ in V. Then, r is called an admissible control. We use $\mathcal{U}$ to denote the class of all such admissible controls. For a control r in $\mathcal{U}$, if it satisfies the constraints (4.5)-(4.7), then it is called a feasible control sequence. Let F be the class of all such feasible controls.

Note that the constraint (4.3) is equivalent to the following constraint:

$$f_{i,t} = \min\{\min\{v_i\rho_{i,t}, C_i\}, \min\{C_{i+1}, w_{i+1}(\rho_{max,i+1} - \rho_{i+1,t})\}\}, \forall i, t, \quad (4.8)$$

where $f^D_{i,t} = \min\{v_i\rho_{i,t}, C_i\}$ and $f^S_{i+1,t} = \min\{C_{i+1}, w_{i+1}(\rho_{max,i+1} - \rho_{i+1,t})\}$ represent the demand function and supply function, respectively.

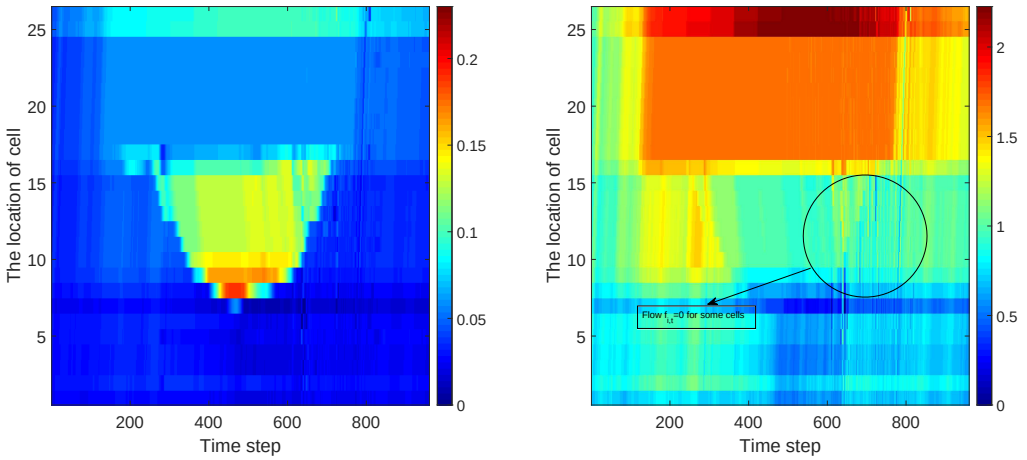
4.2.2 Existing Solution Methods

A Linear inequality relaxation based method

Due to the 'min' operators in constraint (4.3), the optimization problem PCTM is difficult to solve directly. To overcome this difficulty, the 'min' operators are relaxed to the following inequality constraints in [73]:

$$f_{i,t} \leq v_i \rho_{i,t}, \forall i, t, \tag{4.9}$$
$$f_{i,t} \leq C_i, \forall i, t, \tag{4.10}$$
$$f_{i,t} \leq C_{i+1}, \forall i, t, \tag{4.11}$$
$$f_{i,t} \leq w_{i+1}(\rho_{max,i+1} - \rho_{i+1,t}), \forall i, t. \tag{4.12}$$

In general, the optimal value of $f_{i,t}$ obtained for solving the Optimal Control Problem PCTM with the constraint (4.3) being relaxed to constraints (4.9)-(4.12) is strictly less than the right-hand sides of all the constraints (4.9)-(4.12). This problem is known as the flow holding-back problem. In [73], it was shown that for ramp metering control problems, the traffic flow states caused by the inequality relaxation constraints are lower than the fundamental diagram of the CTM. In fact, it is observed that for some cases in our experiments, the traffic flow values of some cells are even zero due to the flow holding-back problem as shown in Fig. 4.1. In this chapter, we will introduce a smoothing approximation of the 'min' operator to resolve the flow holding-back problem (or under the fundamental diagram problem) approximately.



(a) Density (veh/m) for inequality relaxation CTM (b) Flow (veh/s) for inequality relaxation CTM

Figure 4.1: Density and flow for inequality relaxation

Under some strong assumptions, such as symmetric triangular fundamental diagrams, or using the traffic demand control in each cell, the 'min' operators can be relaxed equivalently to the inequality constraints (4.9)-(4.12), see [150–153]. However, it is doubtful whether the assumption of demand control in every cell is realistic. Thus, this method is not a general method.

B Big-M method

Another approach is to convert the 'min' operators in constraint (4.3) into equivalent standard inequality constraints through the introduction of binary variables. This method is known as Big-M method [78]. In fact, constraint (4.3) can be written equivalently as the following constraints:

$$
\begin{aligned}
f_{i,t} &\leq v_i \rho_{i,t}, \forall i, t, && (4.13)\\
f_{i,t} &\leq C_i, \forall i, t, && (4.14)\\
f_{i,t} &\leq C_{i+1}, \forall i, t, && (4.15)\\
f_{i,t} &\leq w_{i+1}(\rho_{max,i+1} - \rho_{i+1,t}), \forall i, t, && (4.16)
\end{aligned}
$$

$$
\begin{aligned}
f_{i,t} &\geq v_i \rho_{i,t} - M(1 - \alpha_{i,t}), \forall i, t, && (4.17)\\
f_{i,t} &\geq C_i - M(1 - \beta_{i,t}), \forall i, t, && (4.18)\\
f_{i,t} &\geq C_{i+1} - M(1 - \gamma_{i,t}), \forall i, t, && (4.19)\\
f_{i,t} &\geq w_{i+1}(\rho_{max,i+1} - \rho_{i+1,t}) - M(1 - \chi_{i,t}), \forall i, t, && (4.20)\\
\alpha_{i,t} &+ \beta_{i,t} + \gamma_{i,t} + \chi_{i,t} = 1, \forall i, t, && (4.21)\\
\alpha_{i,t} &\in \{0,1\}, \beta_{i,t} \in \{0,1\}, \gamma_{i,t} \in \{0,1\}, \chi_{i,t} \in \{0,1\}. && (4.22)
\end{aligned}
$$

Note that after the reformation of the constraints, the resulting optimization problem can be solved only for small-size networks using traditional optimal algorithms. Due to the introduction of a large number of binary variables, it is not feasible to be solved even for medium-sized networks. In this work, using the structure of the model, an effective solution approach is designed to avoid the flow holding-back problem and it is applicable for large-size networks. Details are given in the next section.

4.3 Approximate Model

4.3.1 Min Approximation

It is difficult to solve Problem PCTM directly because the flow $f_{i,t}$ between cells is required to equal to the minimum value in constraint (4.3). However, as discussed before, constraint (4.3) is the key to eliminating internal metering and removing inflow holding-back. To resolve this problem, we propose a novel smoothing approach to approximate the 'min' operators in constraint (4.3) and show that the solution obtained from the approximate problem converges to the solution of Problem PCTM.

The main technical challenge of constraint (4.3) is how to deal with its non-smoothness. Based on (4.8), we can equivalently represent the demand function as $f^D_{i,t} = \min\{v_i\rho_{i,t}, C_i\} = \frac{1}{2}(v_i\rho_{i,t} + C_i - |v_i\rho_{i,t} - C_i|)$, and the supply function as $f^S_{i+1,t} = \min\{C_{i+1}, w_{i+1}(\rho_{max,i+1} - \rho_{i+1,t})\} = \frac{1}{2}(w_{i+1}\ (\rho_{max,i+1} - \rho_{i+1,t}) + C_{i+1} - |w_{i+1}(\rho_{max,i+1} - \rho_{i+1,t}) - C_{i+1}|)$. Let $\epsilon > 0$ be a small number, we use $f^{D\epsilon}_{i,t}$ and $f^{S\epsilon}_{i+1,t}$ to approximate $f^D_{i,t}$ and $f^S_{i+1,t}$, respectively, i.e.,

$$f^{D\epsilon}_{i,t} = \tfrac{1}{2}[v_i\rho_{i,t} + C_i - \sqrt{(v_i\rho_{i,t} - C_i)^2 + \tfrac{\epsilon^2}{4}}], \tag{4.23}$$

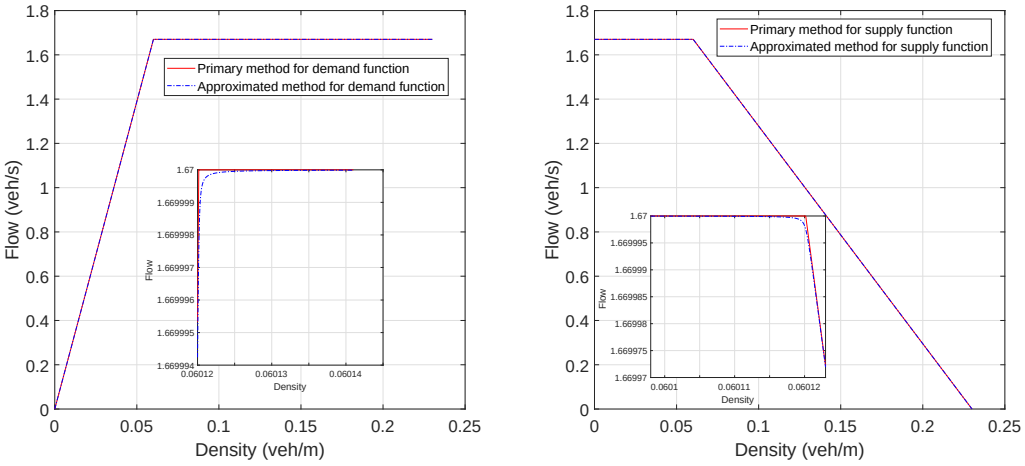
and

$$\begin{aligned} f^{S\epsilon}_{i+1,t} \ = \ & \frac{1}{2}\big[w_{i+1}(\rho_{max,i+1} - \rho_{i+1,t}) + C_{i+1} \\ & -\sqrt{(w_{i+1}(\rho_{max,i+1} - \rho_{i+1,t}) - C_{i+1})^2 + \frac{\epsilon^2}{4}}\big]. \end{aligned} \tag{4.24}$$

It is obvious that the functions $f^{D\epsilon}_{i,t}$ and $f^{S\epsilon}_{i+1,t}$ are smooth and differentiable. Fig. 4.2 shows the demand function and supply function for the 'min' and the approximate operators with $\epsilon = 10^{-5}$. From Fig. 4.2, we can observe that the approximate functions not only retain the main characteristics of the 'min' operators, such as the trend, but also ensure differentiability at the inflection point.

Before carrying out further analysis, we estimate the bounds of $f^D_{i,t} - f^{D\epsilon}_{i,t}$ and $f^S_{i+1,t} - f^{S\epsilon}_{i+1,t}$. By the definitions of $f^{D\epsilon}_{i,t}$ and $f^{S\epsilon}_{i+1,t}$, it holds that

$$0 \le f^D_{i,t} - f^{D\epsilon}_{i,t} \le \frac{\epsilon}{4}, \tag{4.25}$$



(a) Demand function (b) Supply function

Figure 4.2: Demand and supply function for 'min' and approximate operators

Similarly, we can show that

$$0 \leq f^{S}_{i+1,t} - f^{S\epsilon}_{i+1,t} \leq \frac{\epsilon}{4}. \tag{4.26}$$

By virtue of relation (4.8), we have

$$f_{i,t} \quad = \quad \frac{1}{2}(f^{D}_{i,t} + f^{S}_{i+1,t} - |f^{D}_{i,t} - f^{S}_{i+1,t}|). \tag{4.27}$$

Similarly, let

$$f^{\epsilon\epsilon}_{i,t} \quad = \quad \frac{1}{2}(f^{D\epsilon}_{i,t} + f^{S\epsilon}_{i+1,t} - |f^{D\epsilon}_{i,t} - f^{S\epsilon}_{i+1,t}|). \tag{4.28}$$

Using the same approximation method, we obtain

$$f^{\epsilon}_{i,t} = \frac{1}{2}(f^{D\epsilon}_{i,t} + f^{S\epsilon}_{i+1,t} - \sqrt{(f^{D\epsilon}_{i,t} - f^{S\epsilon}_{i+1,t})^2 + \frac{\epsilon^2}{4}}). \tag{4.29}$$

Based on inequalities (4.28) and (4.29), we have

$$0 \leq f^{\epsilon\epsilon}_{i,t} - f^{\epsilon}_{i,t} \leq \frac{\epsilon}{4}. \tag{4.30}$$

4.3.2 Approximate model

Now, the approximate problem of Problem PCTM may be stated as follows:

$$\text{(APCTM)}\min_{r} D^{\epsilon} = \sum_{t=1}^{T}\sum_{i=1}^{I}(\rho_{i,t}^{\epsilon}\Delta x_i\Delta t - \frac{f_{i,t}^{\epsilon}(\rho_{i,t}^{\epsilon})\Delta x_i\Delta t}{v_i}) + \sum_{t=1}^{T}\sum_{j=1}^{J} q_{j,t}\Delta t, \tag{4.31}$$

s.t.

$$\rho_{i,t+1}^{\epsilon} = \rho_{i,t}^{\epsilon} + \frac{\Delta t}{\Delta x_i} \times (f_{i-1,t}^{\epsilon}(\rho_{i-1,t}^{\epsilon}) - f_{i,t}^{\epsilon}(\rho_{i,t}^{\epsilon}) + r_{i,t} - s_{i,t}), \forall i, t, \tag{4.32}$$

$$0 \le \rho_{i,t}^{\epsilon} \le \rho_{max,i}, \forall i, t, \tag{4.33}$$

and constraints (4.4), (4.6), (4.7), (4.23), (4.24) and (4.29).

For a control r in $\mathcal{U}$, if it satisfies the constraints (4.6), (4.7) and (4.33), then it is called a feasible control sequence. Let F^{ϵ} be the class of all such feasible controls. Clearly, Problem APCTM is a smooth discrete time optimal control problem which can be efficiently solved using traditional optimization approaches, such as sequential quadratic programming algorithm (SQP) [158].

Next, we will give some auxiliary lemmas, which are used in the convergence analysis. Lemma 4.1 is to estimate upper and lower bounds of $f_{i,t} - f_{i,t}^{\epsilon}$, which play an essential role in the proofs of Lemma 4.2 and Theorem 4.1.

Lemma 4.1. *Consider (4.23), (4.24) and (4.29). Then, for all $\epsilon > 0$, it holds that $-\frac{\epsilon}{4} \le f_{i,t} - f_{i,t}^{\epsilon} \le \frac{3\epsilon}{4}$, and $\lim_{\epsilon\to 0} f_{i,t}^{\epsilon} = f_{i,t}$.*

Proof. First, we estimate an upper bound of $f_{i,t} - f_{i,t}^{\epsilon}$ for any i and t. By (4.27) and (4.29), it gives

$$\begin{aligned} f_{i,t} - f_{i,t}^{\epsilon} &= f_{i,t} - f_{i,t}^{\epsilon\epsilon} + f_{i,t}^{\epsilon\epsilon} - f_{i,t}^{\epsilon} \\ &= \frac{1}{2}[f_{i,t}^{D} + f_{i+1,t}^{S} - |f_{i,t}^{D} - f_{i+1,t}^{S}| - (f_{i,t}^{D\epsilon} + f_{i+1,t}^{S\epsilon} - |f_{i,t}^{D\epsilon} - f_{i+1,t}^{S\epsilon}|)] \\ &\quad + f_{i,t}^{\epsilon\epsilon} - f_{i,t}^{\epsilon} \\ &\le \frac{\epsilon}{2} + \frac{1}{2}(|f_{i,t}^{D\epsilon} - f_{i+1,t}^{S\epsilon}| - |f_{i,t}^{D} - f_{i+1,t}^{S}|), \end{aligned} \tag{4.34}$$

where the last inequality is due to (4.25), (4.26) and (4.30). We next discuss the positivity and negativity of $f_{i,t}^{D\epsilon} - f_{i+1,t}^{S\epsilon}$ and $f_{i,t}^{D} - f_{i+1,t}^{S}$. Let

$$a = f_{i,t}^{D\epsilon} - f_{i+1,t}^{S\epsilon},\ b = f_{i,t}^{D} - f_{i+1,t}^{S}. \tag{4.35}$$

1): When $a \geq 0,\ b \geq 0$, we obtain

$$|a| - |b| = f_{i,t}^{D\epsilon} - f_{i+1,t}^{S\epsilon} - f_{i,t}^{D} + f_{i+1,t}^{S} \leq \frac{\epsilon}{4},$$

and thus

$$f_{i,t} - f_{i,t}^{\epsilon} \leq \frac{5\epsilon}{8}. \tag{4.36}$$

2): When $a < 0,\ b < 0$, we have

$$|a| - |b| = f_{i+1,t}^{S\epsilon} - f_{i,t}^{D\epsilon} + f_{i,t}^{D} - f_{i+1,t}^{S} \leq \frac{\epsilon}{4},$$

and thus

$$f_{i,t} - f_{i,t}^{\epsilon} \leq \frac{5\epsilon}{8}. \tag{4.37}$$

3): When $a \geq 0,\ b < 0$, we have

$$|a| - |b| = f_{i,t}^{D\epsilon} - f_{i+1,t}^{S\epsilon} + f_{i,t}^{D} - f_{i+1,t}^{S} \leq \frac{\epsilon}{2},$$

and thus

$$f_{i,t} - f_{i,t}^{\epsilon} \leq \frac{3\epsilon}{4}. \tag{4.38}$$

4): When $a < 0,\ b \geq 0$, we obtain

$$|a| - |b| = f_{i+1,t}^{S\epsilon} - f_{i,t}^{D\epsilon} - f_{i,t}^{D} + f_{i+1,t}^{S} \leq \frac{\epsilon}{2},$$

and thus

$$f_{i,t} - f_{i,t}^{\epsilon} \leq \frac{3\epsilon}{4}. \tag{4.39}$$

By (4.36), (4.37), (4.38) and (4.39), it follows that

$$f_{i,t} - f_{i,t}^{\epsilon} \leq \frac{3\epsilon}{4}. \tag{4.40}$$

Next, we estimate a lower bound of $f_{i,t} - f_{i,t}^{\epsilon}$ for any i and t. By (4.27) and (4.29), it gives

$$f_{i,t} - f_{i,t}^{\epsilon} \ \geq \ \frac{1}{2}(|f_{i,t}^{D\epsilon} - f_{i+1,t}^{S\epsilon}| - |f_{i,t}^{D} - f_{i+1,t}^{S}|). \tag{4.41}$$

Similarly, we study the positivity and negativity of $f_{i,t}^{D\epsilon} - f_{i+1,t}^{S\epsilon}$ and $f_{i,t}^{D} - f_{i+1,t}^{S}$.

1): When $a \geq 0,\ b \geq 0$,

$$|a| - |b| = f_{i,t}^{D\epsilon} - f_{i+1,t}^{S\epsilon} - f_{i,t}^{D} + f_{i+1,t}^{S} \geq -\frac{\epsilon}{4},$$

and thus

$$f_{i,t} - f_{i,t}^{\epsilon} \geq -\frac{\epsilon}{8}. \tag{4.42}$$

2): When $a < 0,\ b < 0$,

$$|a| - |b| = f_{i+1,t}^{S\epsilon} - f_{i,t}^{D\epsilon} + f_{i,t}^{D} - f_{i+1,t}^{S} \geq -\frac{\epsilon}{4},$$

and thus

$$f_{i,t} - f_{i,t}^{\epsilon} \geq -\frac{\epsilon}{8}, \tag{4.43}$$

3): When $a \geq 0,\ b < 0$,

$$|a| - |b| = f_{i,t}^{D\epsilon} - f_{i+1,t}^{S\epsilon} + f_{i,t}^{D} - f_{i+1,t}^{S} \geq -\frac{\epsilon}{2},$$

and thus

$$f_{i,t} - f_{i,t}^{\epsilon} \geq -\frac{\epsilon}{4}. \tag{4.44}$$

4): When $a < 0,\ b \geq 0$,

$$|a| - |b| = f_{i+1,t}^{S\epsilon} - f_{i,t}^{D\epsilon} - f_{i,t}^{D} + f_{i+1,t}^{S} \geq -\frac{\epsilon}{2},$$

and thus

$$f_{i,t} - f_{i,t}^{\epsilon} \geq -\frac{\epsilon}{4}. \tag{4.45}$$

By (4.42), (4.43), (4.44) and (4.45), we have

$$f_{i,t} - f_{i,t}^{\epsilon} \geq -\frac{\epsilon}{4}. \tag{4.46}$$

Due to (4.40) and (4.46), we obtain the first part of the result. The second part of the result is obvious when $\epsilon \to 0$, and thus the proof is completed. □

Lemma 4.2. *Let $r^{*\epsilon}$ be an optimal solution to Problem* APCTM. *Then, there*

exists a subsequence of $\{r^{*\epsilon}\}$*, which is again denoted by the original sequence, and a control vector* $\overline{r} \in F$ *such that*

$$\lim_{\epsilon \to 0} ||r^{*\epsilon} - \overline{r}|| = 0, \tag{4.47}$$

$$\lim_{\epsilon \to 0} \rho_{i,t}^{*\epsilon} = \overline{\rho}_{i,t}, \ \forall i, t. \tag{4.48}$$

Similarly, let r^* *be an optimal solution to Problem* PCTM*. Then, there exists a sequence* $\{\overline{r}^{\epsilon}\}$*, which is again denoted by the original sequence, such that*

$$\lim_{\epsilon \to 0} ||\overline{r}^{\epsilon} - r^*|| = 0, \tag{4.49}$$

$$\lim_{\epsilon \to 0} \overline{\rho}_{i,t}^{\epsilon} = \rho_{i,t}^*, \ \forall i, t. \tag{4.50}$$

Proof. Note that V is a compact subset of $\mathbb{R}^J$. Since $\{r^{*\epsilon}\}$ as a sequence in ϵ is in V, it is clear that there exists a subsequence, which is again denoted by the original sequence, and a control vector $\overline{r} \in F$ such that

$$\lim_{\epsilon \to 0} ||r^{*\epsilon} - \overline{r}|| = 0. \tag{4.51}$$

We prove $\lim_{\epsilon \to 0} \rho_{i,t}^{*\epsilon} = \overline{\rho}_{i,t}$ by exploiting the mathematical induction. The result is true when $t = 1$ for each i, because $\lim_{\epsilon \to 0} \rho_{i,1}^{*\epsilon} = \rho_{i,1}$. Next, we prove that the result is true when $t = 2$ for any $i = 1, \ldots, I$. Based on Equations (4.2) and (4.32), we have, for each i,

$$\overline{\rho}_{i,2} = \overline{\rho}_{i,1} + \frac{\Delta t}{\Delta x_i} \times (\overline{f}_{i-1,1} - \overline{f}_{i,1} + \overline{r}_{i,1} - s_{i,1}),$$

and

$$\rho_{i,2}^{*\epsilon} = \rho_{i,1}^{*\epsilon} + \frac{\Delta t}{\Delta x_i} \times (f_{i-1,1}^{\epsilon}(\rho_{i-1,1}^{*\epsilon}) - f_{i,1}^{\epsilon}(\rho_{i,1}^{*\epsilon}) + r_{i,1}^{*\epsilon} - s_{i,1}),$$

where $\overline{\rho}_{i,1} = \rho_{i,1}^{*\epsilon} = \rho_{i,1}$ for each i. Using Lemma 4.1, we obtain for any i

$$\lim_{\epsilon \to 0} f_{i,1}^{\epsilon}(\rho_{i,1}^{*\epsilon}) = \overline{f}_{i,1}. \tag{4.52}$$

Since $0 < \frac{\Delta t}{\Delta x_i} < 1$, by (4.51) and (4.52), we have

$$\lim_{\epsilon \to 0} \rho_{i,2}^{*\epsilon} = \overline{\rho}_{i,2}, \forall i. \tag{4.53}$$

Now, we assume that the result is true when $t = T - 1$ for any $i = 1, \ldots, I$, i.e.,

$$\lim_{\epsilon \to 0} \rho^{*\epsilon}_{i,T-1} = \overline{\rho}_{i,T-1}. \tag{4.54}$$

Next, we shall prove that it is also true when $t = T$. Based on equations (4.2) and (4.32), we have, for any $i = 1, \ldots, I$,

$$\overline{\rho}_{i,T} = \overline{\rho}_{i,T-1} + \frac{\Delta t}{\Delta x_i} \times (\overline{f}_{i-1,T-1} - \overline{f}_{i,T-1} + \overline{r}_{i,T-1} - s_{i,T-1}), \tag{4.55}$$

and

$$\rho^{*\epsilon}_{i,T} = \rho^{*\epsilon}_{i,T-1} + \frac{\Delta t}{\Delta x_i} \times (f^{\epsilon}_{i-1,T-1}(\rho^{*\epsilon}_{i-1,T-1}) - f^{\epsilon}_{i,T-1}(\rho^{*\epsilon}_{i,T-1}) + r^{*\epsilon}_{i,T-1} - s_{i,T-1}). \tag{4.56}$$

Using Lemma 4.1, we obtain, for any $i = 1, \ldots, I$,

$$\lim_{\epsilon \to 0} f^{\epsilon}_{i,T-1}(\overline{\rho}_{i,T-1}) = \overline{f}_{i,T-1}. \tag{4.57}$$

Furthermore, $f^{\epsilon}_{i,t}$ is continuously differentiable with respect to each of the components of ρ and r. Thus, it follows from (4.51) and (4.54) that

$$\lim_{\epsilon \to 0} f^{\epsilon}_{i,T-1}(\rho^{*\epsilon}_{i,T-1}) = \lim_{\epsilon \to 0} f^{\epsilon}_{i,T-1}(\overline{\rho}_{i,T-1}). \tag{4.58}$$

The following result follows readily from (4.57) and (4.58)

$$\lim_{\epsilon \to 0} f^{\epsilon}_{i,T-1}(\rho^{*\epsilon}_{i,T-1}) = \overline{f}_{i,T-1}. \tag{4.59}$$

Then, $\lim_{\epsilon \to 0} \rho^{*\epsilon}_{i,T} = \overline{\rho}_{i,T}$ is obvious by virtue of (4.54), (4.55), (4.56) and (4.59). The second part of the result follows readily by using a similar approach, and thus the details are omitted. The proof is completed. □

Theorem 4.1 below shows that the solution of our approximate model APCT-M will converge to the solution of the Problem PCTM through appropriately controlling the parameter ϵ.

Theorem 4.1. *Under the conditions of Lemma 4.2, $\overline{r}$ is an optimal control vector of Problem* PCTM.

Proof. By induction, it follows from Lemma 4.2 and the differentiability of $q_{j,t}$

that, for each $j = 1, \ldots, J, t = 1, \ldots, T$,

$$\lim_{\epsilon \to 0} q_{j,t}^{*\epsilon} = \overline{q}_{j,t}, \tag{4.60}$$

and

$$\lim_{\epsilon \to 0} \overline{q}_{j,t}^{\epsilon} = q_{j,t}^{*}. \tag{4.61}$$

Based on Lemma 4.1, we have, for each i, t,

$$\lim_{\epsilon \to 0} f_{i,t}^{\epsilon}(\overline{\rho}_{i,t}) = \overline{f}_{i,t}, \tag{4.62}$$

and

$$\lim_{\epsilon \to 0} f_{i,t}^{\epsilon}(\rho_{i,t}^{*}) = f_{i,t}^{*}. \tag{4.63}$$

By Lemma 4.2, $\lim_{\epsilon \to 0} \rho_{i,t}^{*\epsilon} = \overline{\rho}_{i,t}$ and $\lim_{\epsilon \to 0} \overline{\rho}_{i,t}^{\epsilon} = \rho_{i,t}^{*}$ for each i and t. Since $f_{i,t}^{\epsilon}$ is continuously differentiable with respect to each of the components of ρ and r, we have

$$\lim_{\epsilon \to 0} f_{i,t}^{\epsilon}(\rho_{i,t}^{*\epsilon}) = \lim_{\epsilon \to 0} f_{i,t}^{\epsilon}(\overline{\rho}_{i,t}). \tag{4.64}$$

and

$$\lim_{\epsilon \to 0} f_{i,t}^{\epsilon}(\overline{\rho}_{i,t}^{\epsilon}) = \lim_{\epsilon \to 0} f_{i,t}^{\epsilon}(\rho_{i,t}^{*}). \tag{4.65}$$

Combining (4.62) and (4.64) yields

$$\lim_{\epsilon \to 0} f_{i,t}^{\epsilon}(\rho_{i,t}^{*\epsilon}) = \overline{f}_{i,t}. \tag{4.66}$$

By (4.63) and (4.65), we have

$$\lim_{\epsilon \to 0} f_{i,t}^{\epsilon}(\overline{\rho}_{i,t}^{\epsilon}) = f_{i,t}^{*}. \tag{4.67}$$

By virtue of (4.48) of Lemma 4.2, (4.60) and (4.66), we obtain

$$\lim_{\epsilon \to 0} D^{*\epsilon}(\rho^{*\epsilon}, r^{*\epsilon}, q^{*\epsilon}) = D(\overline{\rho}, \overline{r}, \overline{q}), \tag{4.68}$$

where $D^{*\epsilon}(\rho^{*\epsilon}, r^{*\epsilon}, q^{*\epsilon})$ is the optimal function value of Problem APCTM and $D(\overline{\rho}, \overline{r}, \overline{q})$ is the associated objective function value of Problem PCTM. Similarly,

by (4.50) of Lemma 4.2, (4.61) and (4.67), we obtain

$$\lim_{\epsilon \to 0} D^{\epsilon}(\overline{\rho}^{\epsilon}, \overline{r}^{\epsilon}, \overline{q}^{\epsilon}) = D^{*}(\rho^{*}, r^{*}, q^{*}), \tag{4.69}$$

where $D^{\epsilon}(\overline{\rho}^{\epsilon}, \overline{r}^{\epsilon}, \overline{q}^{\epsilon})$ is the associated objective function value of Problem APCTM and $D^{*}(\rho^{*}, r^{*}, q^{*})$ is the optimal function value of Problem PCTM.

For any $\overline{r} \in F$ and $\overline{r}^{\epsilon} \in F^{\epsilon}$, we have

$$D^{*}(\rho^{*}, r^{*}, q^{*}) \leq D(\overline{\rho}, \overline{r}, \overline{q}), \tag{4.70}$$

and

$$\lim_{\epsilon \to 0} D^{*\epsilon}(\rho^{*\epsilon}, r^{*\epsilon}, q^{*\epsilon}) \leq \lim_{\epsilon \to 0} D^{\epsilon}(\overline{\rho}^{\epsilon}, \overline{r}^{\epsilon}, \overline{q}^{\epsilon}). \tag{4.71}$$

Combining (4.68), (4.69), (4.70) and (4.71), we obtain

$$D^{*}(\rho^{*}, r^{*}, q^{*}) \leq D(\overline{\rho}, \overline{r}, \overline{q}) \leq \lim_{\epsilon \to 0} D^{\epsilon}(\overline{\rho}^{\epsilon}, \overline{r}^{\epsilon}, \overline{q}^{\epsilon}) = D^{*}(\rho^{*}, r^{*}, q^{*}).$$

Thus,

$$D(\overline{\rho}, \overline{r}, \overline{q}) = D^{*}(\rho^{*}, r^{*}, q^{*}).$$

This completes the proof. □

Problem APCTM is a smooth discrete time optimal control problem which can be solved efficiently using gradient-based optimization methods. The required gradient formulas for the objective and constraint functions will be derived in the next section.

4.4 Gradients of the Objective and Constraint Functions

To solve Problem APCTM, we need the gradients of the objective and the constraint functions with respect to the variable r.

4.4.1 Gradient formulas

Let $y_t = \{\rho^{\epsilon}_{1,t}, \rho^{\epsilon}_{2,t}, \ldots, \rho^{\epsilon}_{I,t}, q_{1,t}, q_{2,t}, \ldots, q_{J,t}\}^{\top} \in \mathbb{R}^{I+J}$ and $r_t = \{r_{1,t}, r_{2,t} \ldots, r_{J,t}\}^{\top} \in \mathbb{R}^{J}$ be the state and control vectors, respectively. Then, for any $t = 1, 2, \ldots, T-$

1, the state equations of (4.4) and (4.32) are written in compact form as follows:

$$
\begin{aligned}
y_{t+1} \;=\; & \begin{bmatrix} \rho^{\epsilon}_{1,t} \\ \rho^{\epsilon}_{2,t} \\ \vdots \\ \rho^{\epsilon}_{I,t} \\ q_{1,t} \\ q_{2,t} \\ \vdots \\ q_{J,t} \end{bmatrix} + \begin{bmatrix} \frac{\Delta t}{\Delta x_1} \cdot r_{0,t} \\ \frac{\Delta t}{\Delta x_2} \cdot f^{\epsilon}_{1,t} \\ \vdots \\ \frac{\Delta t}{\Delta x_I} \cdot f^{\epsilon}_{I-1,t} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} - \begin{bmatrix} \frac{\Delta t}{\Delta x_1} \cdot f^{\epsilon}_{1,t} \\ \frac{\Delta t}{\Delta x_2} \cdot f^{\epsilon}_{2,t} \\ \vdots \\ \frac{\Delta t}{\Delta x_I} \cdot f^{\epsilon}_{I,t} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} \frac{\Delta t}{\Delta x_1} \cdot r_{1,t} \\ \frac{\Delta t}{\Delta x_2} \cdot r_{2,t} \\ \vdots \\ \frac{\Delta t}{\Delta x_I} \cdot r_{I,t} \\ -\Delta t \cdot r_{1,t} \\ -\Delta t \cdot r_{2,t} \\ \vdots \\ -\Delta t \cdot r_{J,t} \end{bmatrix} \\
& + \begin{bmatrix} -\frac{\Delta t}{\Delta x_1} \cdot s_{1,t} \\ -\frac{\Delta t}{\Delta x_2} \cdot s_{2,t} \\ \vdots \\ -\frac{\Delta t}{\Delta x_I} \cdot s_{I,t} \\ \Delta t \cdot d_{1,t} \\ \Delta t \cdot d_{2,t} \\ \vdots \\ \Delta t \cdot d_{J,t} \end{bmatrix}. \qquad (4.72)
\end{aligned}
$$

Let $F(t, y_t, r_t)$ denote the right hand side of the difference equation (4.72). It contains state variables, control variables and time. The initial condition for the system of difference equations is

$$y_1 = [\rho_{1,1}, \ldots, \rho_{I,1}, q_{1,1}, \ldots, q_{J,1}]^{\top} \in \mathbb{R}^{I+J}. \qquad (4.73)$$

We now consider the following class of discrete time optimal control problems in canonical formulation. Let

$$\Phi_0(y_T(r)) = \sum_{i=1}^{I} (\rho^{\epsilon}_{i,T} \Delta t \Delta x_i - \frac{f^{\epsilon}_{i,T} \Delta t \Delta x_i}{v_i}) + \sum_{j=1}^{J} q_{j,T} \Delta t,$$

$$L_0(y_t(r)) \;=\; \sum_{i=1}^{I} (\rho^{\epsilon}_{i,t} \Delta t \Delta x_i - \frac{f^{\epsilon}_{i,t} \Delta t \Delta x_i}{v_i} + \sum_{j=1}^{J} q_{j,t} \Delta t.$$

Then,

$$g_0(r) = \Phi_0(y_T(r)) + \sum_{t=1}^{T-1} L_0(t, y_t(r)),$$

where $g_0(r)$ is the objective function which is to be optimized subject to $r \in \mathcal{U}$. Similarly, we rewrite the constraints as follows:

$$\begin{aligned}
g_l(r) &= \Phi_l(y_T(r)) + \sum_{t=1}^{T-1} L_l(t, y_t(r), r_t) \\
&= -\rho^{\epsilon}_{i,t}, \;\; 1 \le l \le IT, i = 1 \ldots, I, \; t = 1, \ldots, T, \\
g_l(r) &= \Phi_l(y_T(r)) + \sum_{t=1}^{T-1} L_l(t, y_t(r), r_t) \\
&= \rho^{\epsilon}_{i,t} - \rho_{max,i}, \; IT + 1 \le l \le 2IT, i = 1 \ldots, I, \; t = 1, \ldots, T, \\
g_l(r) &= \Phi_l(y_T(r)) + \sum_{t=1}^{T-1} L_l(t, y_t(r), r_t) \\
&= -q_{j,t}, \; 2IT + 1 \le l \le (2I + J)T, j = 1 \ldots, J, \; t = 1, \ldots, T, \\
g_l(r) &= \Phi_l(y_T(r)) + \sum_{t=1}^{T-1} L_l(t, y_t(r), r_t) \\
&= q_{j,t} - q_{max,i}, \; (2I + J)T + 1 \le l \le 2(I + J)T, \\
&\quad j = 1 \ldots, J, \; t = 1, \ldots, T.
\end{aligned}$$

These constraint functions are said to be in canonical form, because they are in the same form as the objective function. Now, we can derive the gradient formulas of the objective and constraint functions in a unified way. Define

$$r = [(r_1)^\top, (r_2)^\top, \ldots, (r_{T-1})^\top]^\top.$$

Let the control vector r be perturbed by $\xi\widehat{r}$, where $\xi > 0$ is a small constant and $\widehat{r}$ is an arbitrary but fixed perturbation of r given by

$$\widehat{r} = [(\widehat{r}_1)^\top, (\widehat{r}_2)^\top, \ldots, (\widehat{r}_{T-1})^\top]^\top.$$

Then, we have

$$r(\xi) = r + \xi\widehat{r} = [(r_1(\xi))^\top, (r_2(\xi))^\top, \ldots, (r_{T-1}(\xi))^\top]^\top,$$

where

$$r_t(\xi) = r_t + \xi\widehat{r}(\xi), t = 1, \ldots, T-1.$$

Consequently, the state of the system will be perturbed, and so are the objective and constraint functions.

Define

$$y_t(\xi) = y_t(r(\xi)), t = 2, \ldots, T. \tag{4.74}$$

Then,

$$y_{t+1}(\xi) = F(t, y_t(\xi), r_t(\xi)). \tag{4.75}$$

The variation of the state for $t = 1, 2, \ldots, T-1$ is:

$$\Delta y_{t+1} = \frac{\partial F(t, y_t, r_t)}{\partial y_t}\Delta y_t + \frac{\partial F(t, y_t, r_t)}{\partial r_t}\widehat{r}_t \tag{4.76}$$

with

$$\Delta y_1 = 0. \tag{4.77}$$

For the $l-$th function ($l = 0$ is the objective function), we have

$$\frac{\partial g_l(r)}{\partial r}\widehat{r} = \frac{\partial \Phi_l(y_T)}{\partial y_T}\Delta y_T + \sum_{t=1}^{T-1}[\frac{\partial L_l(t, y_t, r_t)}{\partial y_t}\Delta y_t + \frac{\partial L_l(t, y_t, r_t)}{\partial r_t}\widehat{r}_t]. \tag{4.78}$$

For each $l = 0, 1, \ldots, 2(I+J)T$, define the Hamiltonian

$$H_l(t, y_t, r_t, \lambda^l_{t+1}) = L_l(t, y_t, r_t) + (\lambda^l_{t+1})^\top F(t, y_t, r_t),$$

where $\lambda^l_t \in \mathbb{R}^{I+J}, t = T, T-1, \ldots, 2$, denotes the co-state sequence for the $l-$th canonical constraint. Then, it follows from (4.78) that

$$\begin{aligned}\frac{\partial g_l(r)}{\partial r}\widehat{r} = {} & \frac{\partial \Phi_l(y_T)}{\partial y_T}\Delta y_T + \sum_{t=1}^{T-1}\{\frac{\partial H_l(t, y_t, r_t, \lambda^l_{t+1})}{\partial y_t}\Delta y_t - (\lambda^l_{t+1})^\top \frac{\partial F(t, y_t, r_t)}{\partial y_t}\Delta y_t \\ & + \frac{\partial H_l(t, y_t, r_t, \lambda^l_{t+1})}{\partial r_t}\widehat{r}_t - (\lambda^l_{t+1})^\top \frac{\partial F(t, y_t, r_t)}{\partial r_t}\widehat{r}_t\}.\end{aligned} \tag{4.79}$$

Based on (4.76) and (4.77), we have

$$\Delta y_{t+1} = \frac{\partial F(t, y_t, r_t)}{\partial y_t}\Delta y_t + \frac{\partial F(t, y_t, r_t)}{\partial r_t}\widehat{r}_t. \tag{4.80}$$

Let the co-state λ_t^l be determined by the following system of difference equations:

$$(\lambda_t^l)^\top = \frac{\partial H_l(t, y_t, r_t, \lambda_{t+1}^l)}{\partial y_t}, t = T-1, T-2, \ldots, 2, \tag{4.81}$$

and

$$(\lambda_T^l)^\top = \frac{\partial \Phi_l(y_T)}{\partial y_T}. \tag{4.82}$$

By virtue of (4.79), (4.80), (4.81), (4.82), (4.74) and (4.75), we obtain

$$\frac{\partial g_l(r)}{\partial r}\widehat{r} = [\frac{\partial H_l(1, y_1, r_1, \lambda_2^l)}{\partial r_1}, \ldots, \frac{\partial H_l(T-1, y_{T-1}, r_{T-1}, \lambda_T^l)}{\partial r_{T-1}}]\widehat{r}.$$

Because $\widehat{r}$ is arbitrary, we have the following gradient formula:

$$\frac{\partial g_l(r)}{\partial r} = [\frac{\partial H_l(1, y_1, r_1, \lambda_2^l)}{\partial r_1}, \ldots, \frac{\partial H_l(T-1, y_{T-1}, r_{T-1}, \lambda_T^l)}{\partial r_{T-1}}]. \tag{4.83}$$

Now we summarize the gradient computation in the following theorem.

Theorem 4.2. *Consider Problem* APCTM. *Then, for each* $l = 0, 1, \ldots, 2(I + J)T$, *the gradient of* $g_l(r)$ *with respect to control vector* r *is given by (4.83), where* $r = [(r_1)^\top, (r_2)^\top, \ldots, (r_{T-1})^\top]^\top$.

4.4.2 Algorithm

Problem APCTM is essentially a nonlinear mathematical programming problem where the decision vector is the control vector r. Many gradient-based optimization methods, such as SQP, can be used to solve it. To apply gradient-based optimization methods, for each $r \in V$, we need the values of the objective function $g_0(r)$ and the constraint functions $g_l(r)$, $l = 1, 2, \ldots, 2(I + J)T$, together with their corresponding gradients. Detailed computation is given in Algorithm 1.

Step 2 in Algorithm 1 is to compute the values of $y_t(r)$ corresponding to each given r. Then, $g_l(r)$ is computed in Step 9 based on Step 4 and Step 8. After that, the co-state system (4.81) and (4.82) is solved backward in time from $t = T$ to $t = 2$ to acquire $\lambda_t^l(r)$. Finally, the gradients of the objective function and the

Algorithm 1 : Algorithm to Compute Gradients of Objective function and Constraints

Initialization: For a given $r \in V$ and initial condition (4.73).
Output: Compute the solution $y_{t+1}(r)$, $t = 1, 2, \ldots, T-1$ of system (4.72) forward in time from $t = 1$ to $t = T-1$.
for $l = 0$ to $2(I+J)T$ **do**
    Compute $\Phi_l(y_T(r))$.
    for $t = 1$ to $T-1$ **do**
        Compute $L_l(t, y_t(r), r_t)$.
    end for
    Compute $\sum_{t=1}^{T-1} L_l(t, y_t(r), r_t)$.
    Compute $g_l(r) = \Phi_l(y_T(r)) + \sum_{t=1}^{T-1} L_l(t, y_t(r), r_t)$.
    for $t = T$ to 2 **do**
        Solve the system of the co-state system (4.81) and (4.82) backward in time. Let $\lambda_t^l(r)$ be the solution obtained.
    end for
    Calculate the gradients of g_l using (4.83).
end for

constraint functions are calculated using the gradient formulas given in Theorem 4.2.

4.5 Numerical Studies

In this section, the numerical performance of a ramp metering method based on the proposed model is given. A MPC approach is utilized to achieve traffic response controls.

The MPC approach has been widely utilized in freeway traffic control problems, see [123,159,160]. Based on the current state of the traffic system, the MPC approach utilizes a traffic model to predict dynamic of the state, and finds an optimal control signal which gives the optimal value of the objective function. This property guarantees that the controller can take advantage of potentially larger future gains at a current (smaller) cost, thereby avoiding short-sighted control action.

After optimization, the values for the control variables of the first sample of the optimal control action are applied to the process. The remaining control signals are recalculated in a finite rolling horizon scheme. Readers can refer to [161] for a detailed description of the MPC method.

4.5.1 MPC Design

In this section, we first redefine the MPC objective function, which minimizes the total delay of the freeway system, including mainline delay and on-ramp delay for $l = t, \ldots, t + \mathrm{Np} - 1$. It is similar to the cost function defined in (4.31). The optimization problem is reformulated as follows:

$$D_t = \sum_{l=t}^{t+\mathrm{Np}-1} [\sum_{i=1}^{I} (\rho_{i,l} \Delta x_i \Delta t - \frac{f_{i,l} \Delta x_i \Delta t}{v_i}) + \sum_{j=1}^{J} q_{j,l} \Delta t]. \tag{4.84}$$

Note that the tuning rules used to select the appropriate value of the prediction horizon Np and control sample time Nc are very important to the performance of MPC [162]. Normally, the value of Np should be larger than the typical travel time from the controlled segment to the exit of the network. The reason is that if the prediction horizon Np is shorter than the typical travel time, the vehicles affected by the current control action have no effect on the network operation before exiting. On the other hand, Np should not be too large due to the computational complexity of the MPC optimization problem. Hence, we choose Np as the typical time in the network based on this reasoning. For the control sample time Nc, we will choose a value that represents the trade-off between performance and computational effort.

Performances of the following models are compared:

- Primal cell transmission model (PCTM)
- Linear inequality relaxation based method (LICTM)
- Big-M method (BMM)
- Approximate model (APCTM)
- Generalized non-holding back linear programming formulation (GNHBLP) proposed by Zhu [163] to address holding back problems.

All the experiments were run on a computer with Intel(R) Core(TM) i7-8565U CPU-1.80GHz 1.99 GHz and RAM 16GB. We use the MATLAB implementation of the SQP algorithm (fmincon) to solve the models PCTM, LICTM, APCTM and GNHBLP, and use the GUROBI [164] to solve the model BMM.

4.5.2 Example 1

A Scenario

The example is selected from the Kwinana Freeway in the vicinity of Perth in Australia. This section of freeway is divided into $I = 26$ cells with $J = 8$ on-ramps and 4 off-ramps. Each segment has a longitude of $\Delta x_i = 500$m for any $i = 1, \ldots, I$. Fig. 4.3 shows the details of this example.

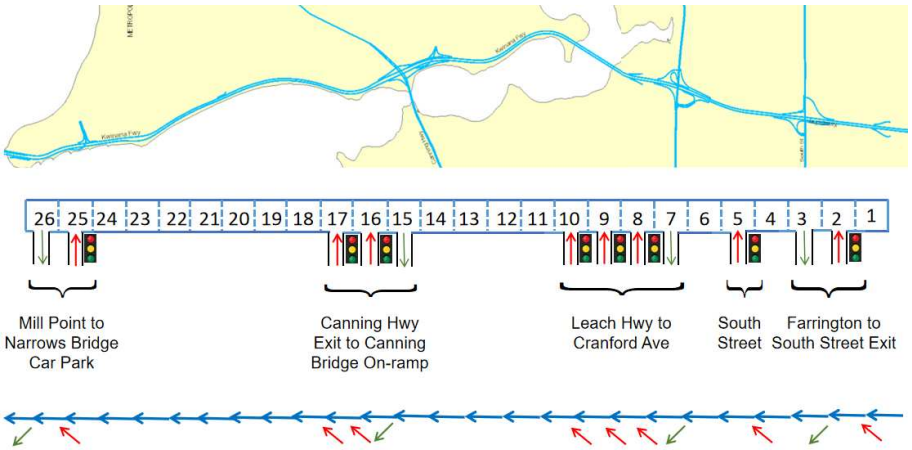


Figure 4.3: Road graph

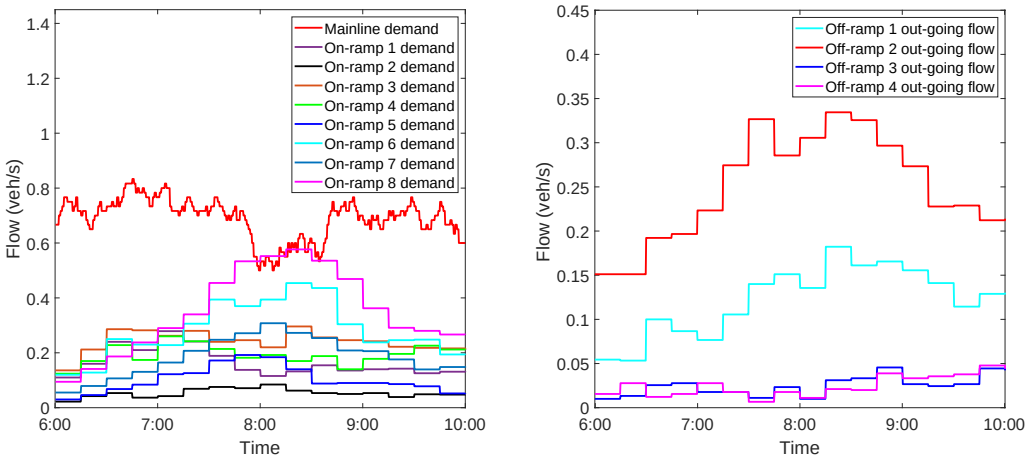
There are eight control signals: ramp metering in cells 2, 5, 8, 9, 10, 16, 17, 25, respectively. We measured density, flow, speed, on-ramp demands and out-going flows at each sample time t. We can obtain the flow-density fundamental diagrams from the data measured by detectors installed in the Kwinana Freeway. These model parameters are calculated using the least square method and are given in Table 4.2. The mainline demand, ramp demands and out-going flows are measured from the detectors, see Fig. 4.4. Traffic control can be used to improve the performance of the freeway system. The time chosen is 4h from 6 : 00 am to 10 : 00 am, which corresponds to 960 steps. In this example, we select Np = 33, i.e., 8 minutes as predictive horizon, and Nc = 8, i.e., 2 minutes for control sample time, which meet the above requirements.

B Experiment Results

The density results for the six models (actually being measured, PCTM, LICTM, BMM, APCTM and GNHBLP) are shown in Fig. 4.5. The bars on the right side of each of Figs. 4.5(a)-4.5(f) show the changes in the values of the corresponding state variables that increase from bottom to top. From Figs. 4.5(a)-4.5(f), we can observe the evolution for density for each of the models (PCTM, LICTM, BMM, APCTM and GNHBLP). As the demands for the mainline and on-ramps increase,

Table 4.2: Model parameters

Name of parameter	Value
$v_i, i = 1, \ldots, 24$	27.7778m/s
$v_i, i = 25, 26$	22.2222m/s
$w_i, i = 1, 3, 4, 6, \ldots, 24$	9.8029m/s
$w_i, i = 2, 5$	9.7895m/s
$w_i, i = 25, 26$	10.783m/s
$C_i, i = 1, 3, 4, 6, \ldots, 24$	1.6667veh/s
$C_i, i = 2, 5, 25, 26$	2.2222veh/s
$q_{max,j}, j = 7$	120veh
$q_{max,j}, j = 1, \ldots, 6, 8$	60veh
$r_{max,j}, j = 1, \ldots, 8$	0.55veh/s
$\rho_{max,i}, i = 1, 3, 4, 6, \ldots, 24$	0.23veh/m
$\rho_{max,i}, i = 2, 5, 25, 26$	0.3067veh/m
Δt	15s
ϵ	0.00000001

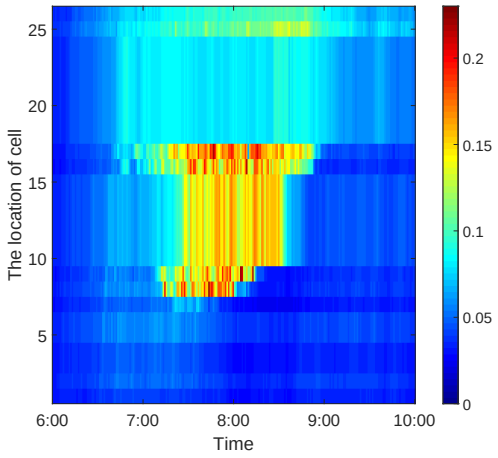


(a) Mainline demand and on-ramp demands (b) Out-going flows

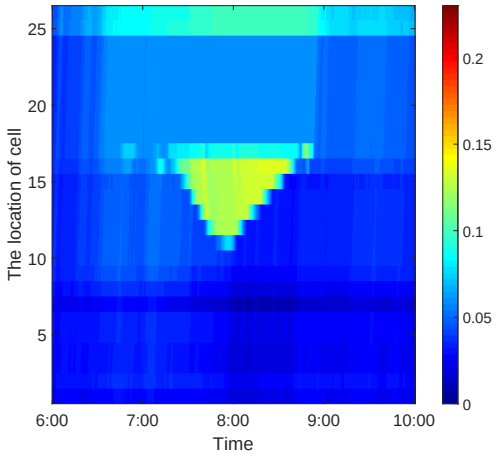
Figure 4.4: Mainline demand, on-ramp demands and out-going flows

the density at the junctions of the mainline and on-ramps increases. Consequently, the speed decreases, congestion gradually occurs, and the congestion wave propagates upstream from the junction. This situation lasts approximately two hours, which is consistent with the traffic flow during the morning rush hour, i.e., from 7:00 am to 9:00 am. From Figs. 4.5(b), 4.5(d) and 4.5(e), we observe that the changing trend of our model APCTM is almost the same as that of PCTM and BMM. This is consistent with our theoretical analysis.

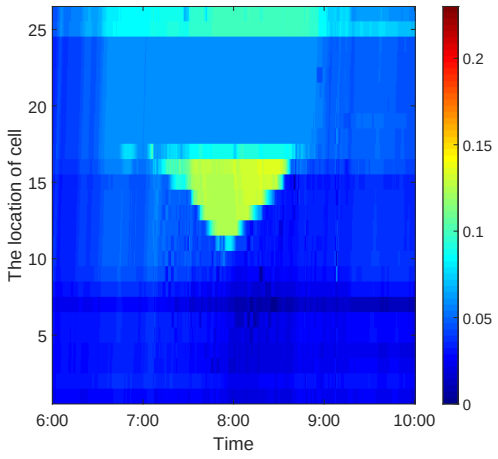
The traffic flow results for the six models (actually being measured, PCTM,



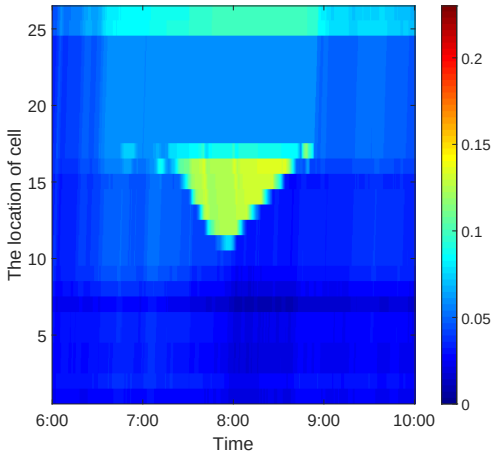
(a) Actual density (veh/m)



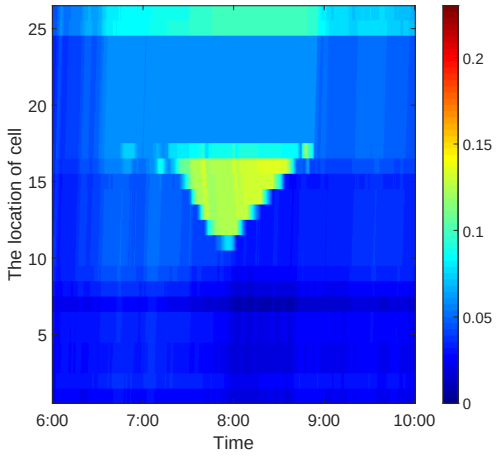
(b) Density (veh/m) for (PCTM) with Np = 33, Nc = 8



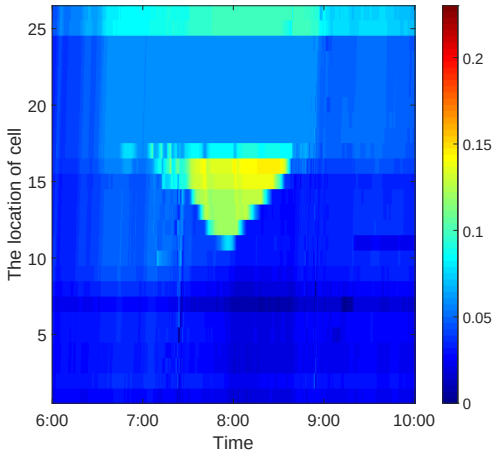
(c) Density (veh/m) for (LICTM) with Np = 33, Nc = 8



(d) Density (veh/m) for (BMM) with Np = 33, Nc = 8



(e) Density (veh/m) for (APCTM) with Np = 33, Nc = 8



(f) Density (veh/m) for (GNHBLP) with Np = 33, Nc = 8

Figure 4.5: Traffic density for actually being measured, PCTM, LICTM, BMM, APCTM and GNHBLP

Table 4.3: The Sizes of Variables and Constraints of The Five Models with Different Predictive Horizon Np

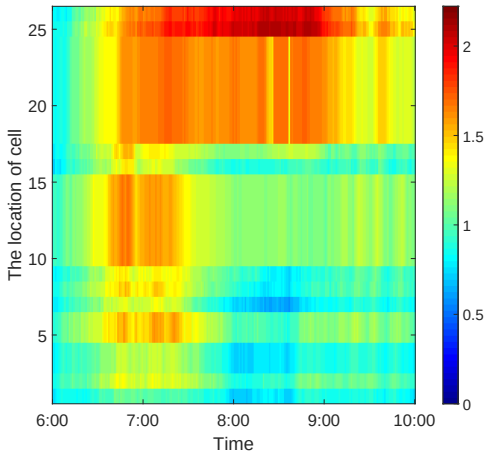
Model	Np = 33		Np = 121		Np = 481	
	N_v	N_c	N_v	N_c	N_v	N_c
PCTM	256	2244	960	8160	3840	30784
LICTM	1088	5610	4080	20570	16320	81770
BMM	4352	9834	16320	36058	65280	143338
APCTM	256	2244	960	8160	3840	30784
GNHBLP	1089	5611	4081	20571	16321	81771

LICTM, BMM, APCTM and GNHBLP) are shown in Fig. 4.6. From Figs. 4.6(a)- 4.6(f), we can observe the evolution of traffic flow of the models (PCTM, LICTM, BMM, APCTM and GNHBLP). Compared the result of LICTM with those of other models, we see that PCTM, BMM, APCTM and GNHBLP can solve the flow holding-back problem. However, the performance of PCTM, BMM and APCTM is better than that of GNHBLP. The experimental results agree with the theoretical analysis.

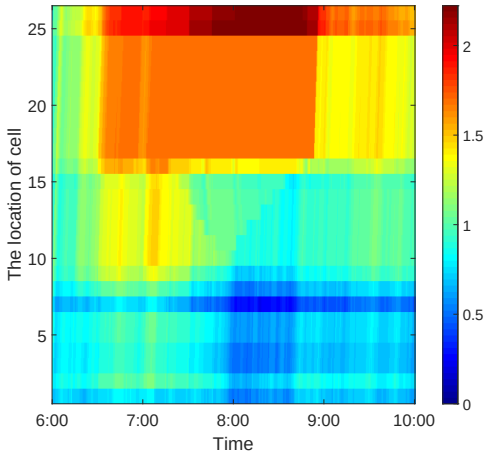
We show the ramp metering and queue length of the five models (PCTM, LICTM, BMM, APCTM and GNHBLP) in Fig. 4.7 and Fig. 4.8. From Figs. 4.7(a), 4.7(c) and 4.7(d), we see that the solution of our model APCTM converges to those of PCTM and BMM when the parameter ϵ is appropriately chosen. Due to the flow holding-back problem arising from linear relaxation, the solution of LICTM shows a different trend. GNHBLP only solves the flow holding-back problem to a certain extent.

We compare the sizes of the fives models (PCTM, LICTM, BMM, APCTM, GNHBLP) with different prediction horizon Np in Table 4.3, where N_v and N_c denote, respectively, the numbers of variables and constraints. We find that PCTM and APCTM have considerably fewer numbers of variables and constraints than LICTM, BMM and GNHBLP. Even though PCTM and APCTM have the same numbers of variables and constraints, we find that the time cost of APCTM is far less than that required by PCTM model (Table 4.4). In particular, the trend is more evident when the number of the variables increases. BMM equivalently represents the 'min' operators by introducing a large number of auxiliary variables, so the CPU time is more than that required by APCTM. Since LICTM and GNHBLP are linear program problems, the computational time will obviously be less than that of APCTM. However, they are relaxation problems of the original problem, so the total delay and ramp delay obtained tend to be longer.

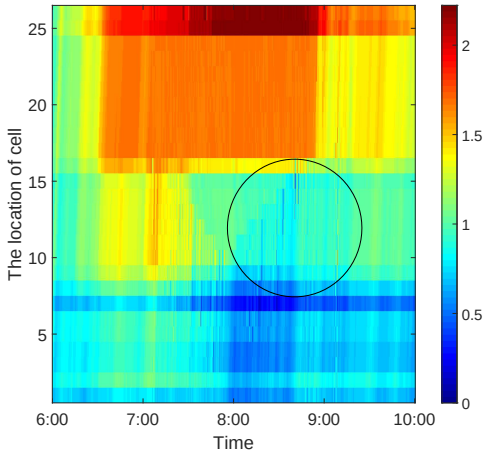
For comparison, we summarize the total delay, ramp delay and total CPU time (in seconds) of the above five models in Table 4.5. Clearly, we can see that



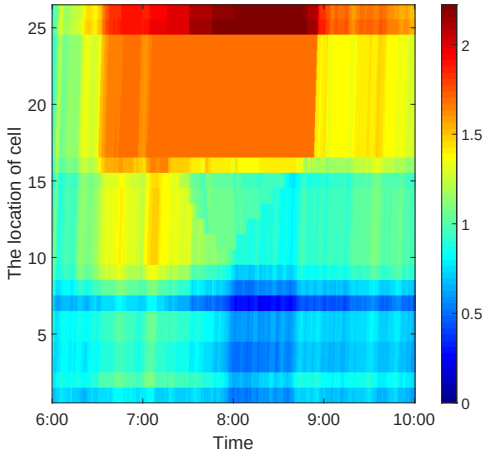
(a) Actual flow (veh/s)



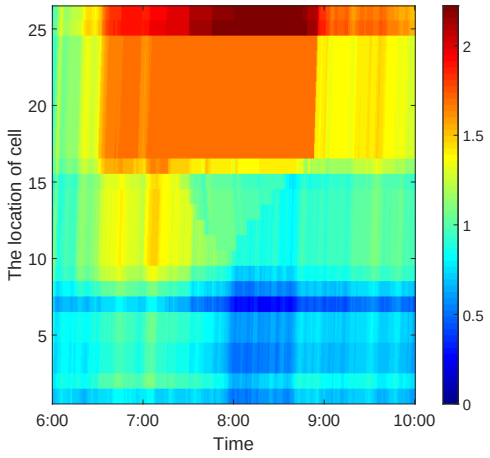
(b) Flow (veh/s) for (PCTM) with Np = 33, Nc = 8



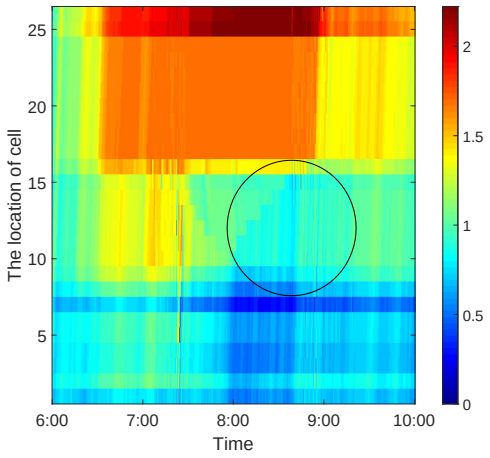
(c) Flow (veh/s) for (LICTM) with Np = 33, Nc = 8



(d) Flow (veh/s) for (BMM) with Np = 33, Nc = 8

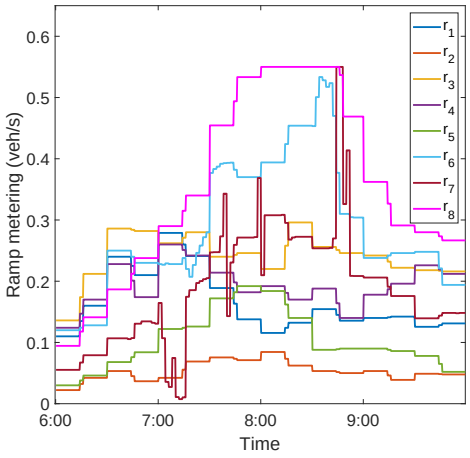


(e) Flow (veh/s) for (APCTM) with Np = 33, Nc = 8

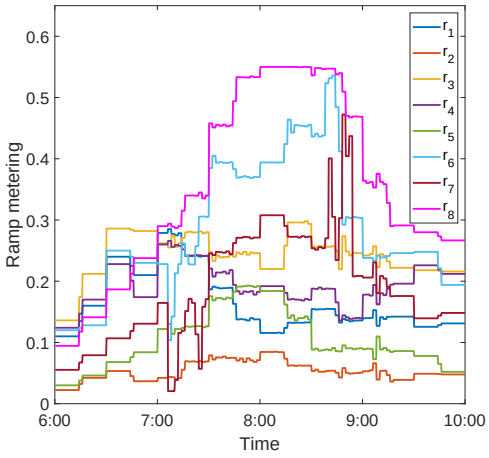


(f) Flow (veh/s) for (GNHBLP) with Np = 33, Nc = 8

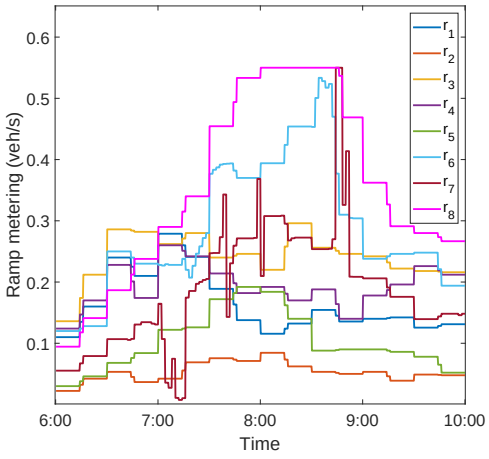
Figure 4.6: Flow for actually being measured, PCTM, LICTM, BMM, APCTM and GNHBLP



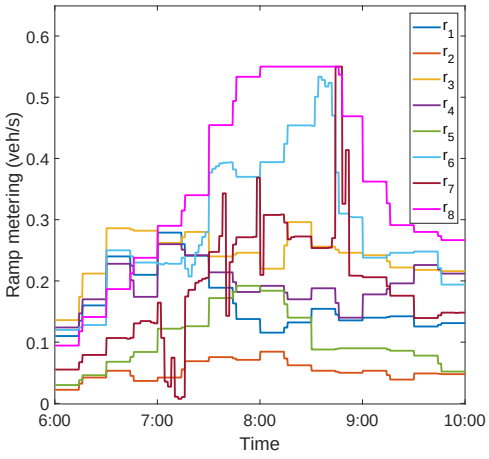
(a) Ramp metering (veh/s) for (PCTM) with Np = 33, Nc = 8



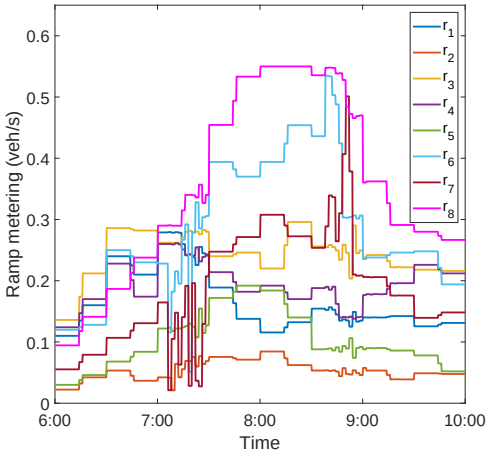
(b) Ramp metering (veh/s) for (LICTM) with Np = 33, Nc = 8



(c) Ramp metering (veh/s) for (BMM) with Np = 33, Nc = 8

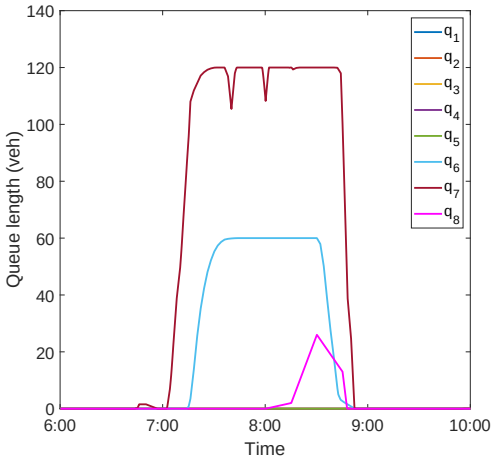


(d) Ramp metering (veh/s) for (APCTM) with Np = 33, Nc = 8

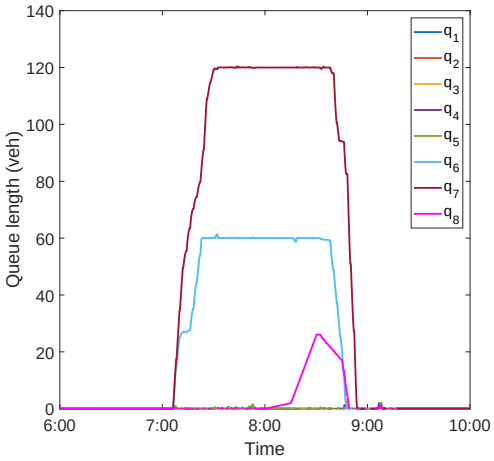


(e) Ramp metering (veh/s) for (GNHBLP) with Np = 33, Nc = 8

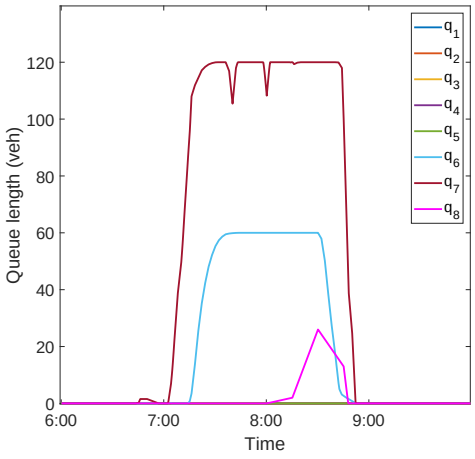
Figure 4.7: Ramp metering for PCTM, LICTM, BMM, APCTM and GNHBLP



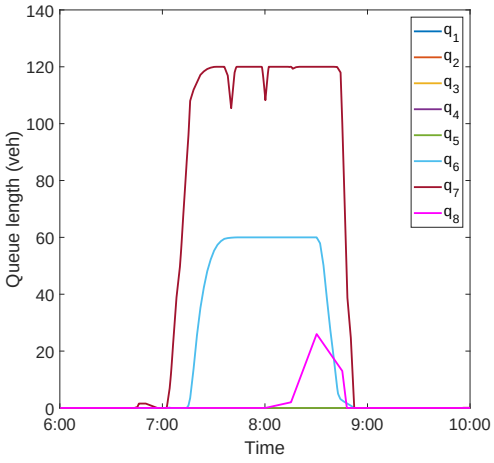
(a) Queue length (veh) for (PCTM) with Np = 33, Nc = 8



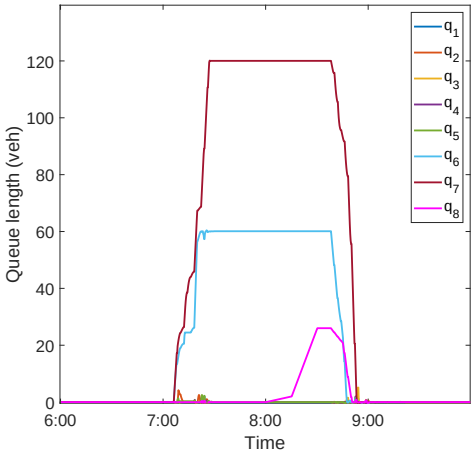
(b) Queue length (veh) for (LICTM) with Np = 33, Nc = 8



(c) Queue length (veh) for (BMM) with Np = 33, Nc = 8



(d) Queue length (veh) for (APCTM) with Np = 33, Nc = 8



(e) Queue length (veh) for (GNHBLP) with Np = 33, Nc = 8

Figure 4.8: Queue length for PCTM, LICTM, BMM, APCTM and GNHBLP

Table 4.4: The CPU time (in Seconds) Required for Application (4.84) with Different Predictive Horizon Np

Model	Np = 33	Np = 121	Np = 481
	CPU time	CPU time	CPU time
PCTM	2.0598	35.8489	10938.5695
LICTM	0.2289	2.0629	22.2301
BMM	1.3830	20.7851	2686.0885
APCTM	0.8402	8.3765	767.0149
GNHBLP	0.2521	2.3057	22.4705

the total delay of APCTM is better than LICTM and GNHBLP because our model APCTM does not have a flow holding-back problem. In Table 4.5, we can also see that the total delay and ramp delay of APCTM are, respectively, 99.9965% and 99.8162% of those of PCTM. Compared to the case with no control, our model can reduce the total delay by 55.6279%. Compared with the LICTM, our approximate model reduces the total delay and ramp delay by 2.0455% and 2.0192%, respectively. Furthermore, the total delay and ramp delay of APCTM are also better than those of GNHBLP. The total CPU time taken by APCTM is 119.1033s. Since in the real application, the control sample time is 2 minutes, it means that the CPU time taken by APCTM can meet the requirements of online control applications. By comparison, we can find the total CPU time of PCTM and BMM are, respectively, 2.6856 and 1.8263 times of that of APCTM. Compared with LICTM and GNHBLP, APCTM requires more computational time due to the involvement of nonlinearity. On the other hand, the total delay and ramp delay are shorter when compared with those of LICTM and GNHBLP.

Table 4.5: Total delay, ramp delay and Total CPU time (in Seconds) Required for Application (4.84) with Np = 33, Nc = 8

Model	Total delay	Ramp delay	Total CPU time
No control	1074.6552h	0h	0
PCTM	476.8305h	276.6920h	345.6726
LICTM	486.8045h	282.4111h	51.3336
BMM	476.8338h	276.6953h	217.5161
APCTM	476.8471h	276.7086h	119.1033
GNHBLP	480.2179h	279.5597h	53.6784

4.5.3 Example 2

A Scenario

The example is built based on the actual situation of the Freeway in the vicinity of Shapingba, Chongqing in China. The section of freeway under consideration is divided into $I = 15$ cells with $J = 2$ on-ramps. Cells $i = 1, 2, \ldots, 9$ have 4 lanes and cells $i = 10, 11, \ldots, 15$ have 3 lanes. Fig. 4.9 shows the information of road map. For this example, we use SUMO to simulate the traffic flow dynamic from 6:00 am to 10:00 am to obtain the required data. Mainline demand and on-ramp demands are presented in Fig. 4.10. The values of parameters are listed in Table 4.6. In this example, we let Np = 17, i.e., 2 minutes and 40 seconds as predictive horizon, and Nc = 6, i.e., 1 minute for control sample time, which can meet the above requirements. Note that this example aims to validate that our proposed method is also efficient for the simulated case.

Figure 4.9: Road graph

Table 4.6: Model parameters

Name of parameter	Value	Name of parameter	Value
$v_i, i = 1, \ldots, 15$	22.2222m/s	$q_{max,j}, j = 1$	60veh
$w_i, i = 1, 2, \ldots, 9$	6.0206m/s	$q_{max,j}, j = 2$	70veh
$w_i, i = 10, 11, \ldots, 15$	11.6436m/s	$r_{max,j}, j = 1$	1veh/s
$C_i, i = 1, 2, \ldots, 9$	2.3685veh/s	$r_{max,j}, j = 2$	0.55veh/s
$C_i, i = 10, 11, \ldots, 15$	2.1916veh/s	Δt	10s
$\rho_{max,i}, i = 1, 2, \ldots, 9$	0.5veh/m	ϵ	0.00000001
$\rho_{max,i}, i = 10, 11, \ldots, 15$	0.2868veh/m		

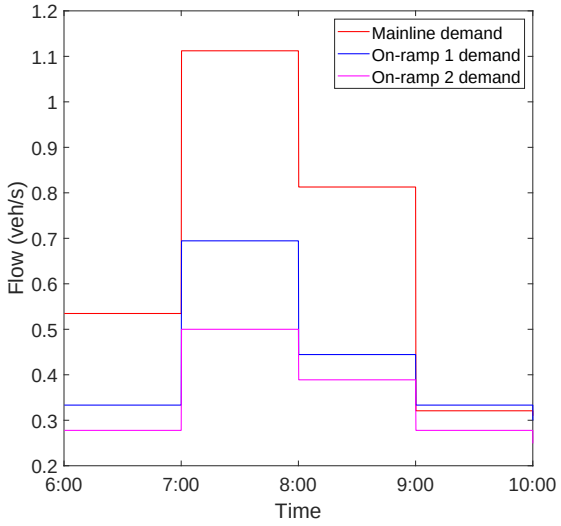


Figure 4.10: Road graph

B Experiment Results

We compare the sizes and CPU time (in seconds) of the five models (PCTM, LICTM, BMM, APCTM, GNHBLP) for each iteration with $\mathrm{Np} = 17, \mathrm{Nc} = 6$ in Table 4.7. Similar to Example 1, PCTM and APCTM have fewer numbers of variables and constraints than those of LICTM, BMM and GNHBLP. APCTM can be solved more efficiently than PCTM and BMM, but not better than LICTM and GNHBLP.

Table 4.7: The Sizes and CPU Time (in Seconds) Required for Application (4.84) with Np = 17, Nc = 6

Model	N_v	N_c	CPU time
PCTM	32	578	0.3502
LICTM	272	1564	0.0609
BMM	1200	2550	0.2686
APCTM	32	578	0.1736
GNHBLP	273	1565	0.0641

The total delay, ramp delay and total CPU time (in seconds) of the five models are summarized in Table 4.8. From Table 4.8, we clearly observe that the total delay of APCTM is better than those of LICTM and GNHBLP, because APCTM can solve the flow holding-back problem if the parameter ϵ is appropriately chosen. In addition, we can also observe that the total delay and ramp

delay of our APCTM are, respectively, 99.9775% and 99.678% of those obtained using PCTM. Compared with the case of no control, APCTM can reduce the total delay by 52.4996%, but the ramp delay increases by 18.6882%. Compared with LICTM and GNHBLP, APCTM reduces the total delay by 1.7654% and 0.872%, respectively. As for the total CPU time, PCTM and BMM are, respectively, 2.0173 and 1.6762 times of that of APCTM, but LICTM and GNHBLP take shorter times than that of APCTM. In this example, the total CPU time taken by APCTM is 38.8343s. Since in the real application, the control sample time is 1 minute, it means that the CPU time taken by APCTM can also meet the requirements of online control applications.

Table 4.8: Total delay, ramp delay and Total CPU time (in Seconds) Required for Application (4.84) with Np = 17, Nc = 6

Name of model	Total delay	Ramp delay	Total CPU time
No control	520.7225h	64.0275h	0
PCTM	247.2868h	75.9776h	78.3409
LICTM	251.7902h	81.0918h	12.8565
BMM	247.2896h	75.9931h	65.0922
APCTM	247.3452h	76.2231h	38.8343
GNHBLP	249.5209h	77.4397h	13.8184

Based on the results obtained for Example 1 and Example 2, we can conclude that the total delay and ramp delay of APCTM are better than those of LICTM and GNHBLP. Furthermore, the CPU time of APCTM is at least 2 and 1.5 times of those of PCTM and BMM. Therefore, APCTM is efficient in terms of computational time and the reduction of total delay and ramp delay.

4.6 Conclusion

This study proposed a novel approximate optimization model based on the CTM to overcome the flow holding-back problem caused by unrealistic fundamental diagrams if inequality convex relaxation is used directly. Theoretical analysis shows that the solution obtained from our model converges to the solution of the original CTM as ϵ approaches to zero. To obtain a feedback control law, a customized MPC approach is designed under the framework of our proposed

optimization problem. Experimental results show that our approach is better than the existing methods.

CHAPTER 5

Freeway Ramp Metering and Variable Speed Limit Control Based on Cell Transmission Model

5.1 Introduction

In this Chapter, a novel RM and VSL control model is developed, in which the influences of RM and VSL are incorporated into a macroscopic CTM as control components. The problem of traffic control in highway network via RM and VSL is modeled as a discrete-time optimal control problem with constraints. Due to non-smoothness of the 'min' operators, a smoothing method is introduced to approximate this non-smooth model, and an efficient MPC solution technique based on co-state system is used to solve this smooth approximate control problem. This approach is applicable to large-scale networks. Two practical examples of freeway stretching under different control scenarios are utilized to validate the performance of the developed model. The results show that the use of VSL control measures, especially in combination with the coordinated ramp metering control, can significantly improve the traffic flow efficiency and mitigate traffic congestion.

The purpose of RM is mainly focused on alleviating the traffic congestion to improve traffic operations by adjusting or restricting the inflows from on-ramps to the freeway main road. Due to the limited storage space and equity considerations, RM has been applied to multiple ramps (coordinated ramp metering) [114] to achieve maximum efficiency. The coordinated RM strategies take advantage of measurements across the entire network to control all metered ramps. Although RM leads to ramp queue delays, the total traffic throughput increases because the traffic congestion on mainstream of freeways is reduced, thereby reducing the

total time spent and delays for most drivers. In the past, extensively studies have been conducted on coordinated ramp metering, such as multivariable control strategies [165] and optimal control strategies [120,159,166]. Previous studies have shown that the influence of VSL on aggregate traffic flow behaviours on freeway networks is similar to that of RM, especially when addressing potential active bottlenecks. This strategy for controlling freeway traffic flow is to implement VSL on appropriate variable message signs (VMS) based on prevailing traffic conditions. Nowadays, there are mainly three kinds of VSL control methods in literatures, and for a review of using VSL for the freeway traffic control for readers is referred to [167].

- VSL for traffic safety: The aim of this kind of control methods is mainly to improve the traffic safety by decreasing the velocity limit when vehicles are approaching jammed roads or accidents. Relevant studies rooted in the traffic safety [168, 169] have found that VSL control system can help to improve the traffic safety in certain restricted scenarios. For instance, the finding in [168] showed that the number of crashes was reduced by 30%.

- VSL for pollution reduction: These VSL control methods aim to design a controller for improving the traffic efficiency while offering a tradeoff between the travel time, emissions, and fuel consumptions. However, the results obtained in [170] indicated that the emission reduction achieved is usually relatively low if the traffic efficiency did not decrease (2.5% for the case study considered in [170]).

- VSL for traffic operation improvement: The purpose of these VSL control methods is mainly to decrease the fluctuation of traffic behaviors by regulating or limiting the flow of arriving vehicles, so as to reduce, or avoid the traffic congestion on the freeway, thereby mitigating or avoiding the capacity drop and maximizing the out-going flow from activation bottlenecks [162, 166, 171–173]. These studies of these control methods are mainly through simulation. Thus, their potentially significant improvement in traffic operation efficiency depends on how the model is constructed. In addition, the application of model-based optimal control approaches (such as MPC [161, 162]) can significantly reduce the total time spent (TTS), fuel consumptions, and other performance indicators. However, the success is highly dependent on the precision of the proposed model [162].

From the literature review given above, we see that many VSL and RM collaborative studies use the second order models [113] for the freeway network loading.

The second order models have evident strengths due to the fact that they can explicitly characterize speeds and show the very common phenomena, such as capacity drop. However, in comparison with the first order model (CTM) [66,67]), the accuracy of these models relies highly on model parameters that need to be customized. Up to now, the proposed optimal control models on the basis of second order models have only produced local optimality [148] on account of the nonconvexity and the complexity. Even though few models proposed based on the first-order model can cooperatively implement VSL and RM, they have some limitations [171]. For instance, a mode dependent VSL control strategy is developed in [171] based on CTM for freeway networks, but this model can only resolve such problem without on-ramps. In this work, we propose a novel optimal control model based on CTM, which considers the capacity drop. In this model, VSL and RM are control variables that are simultaneously implemented.

The layout of this chapter is govern as follows: Section 5.2 formulates the traffic dynamics based on CTM by incorporating ramp metering as well as variable speed limits. In Section 5.3, a novel smooth model is proposed by approximating the 'min' operators of the model designed in Section 5.2. A customized solution technique is designed in Section 5.4. Numerical Application is conducted in Section 5.5. Finally, Section 5.6 draws some conclusions.

5.2 Modeling Traffic Flow Dynamics

5.2.1 The application framework

We take into account freeway traffic flows with a sufficient proportion of Vehicle Automation and Communication System (VACS)-equipped vehicles. Sufficient proportion implies that the considered decision variables in the optimization model can be actually carried out by obtaining information from real-time traffic conditions and installing appropriate actuators. This assumption is reasonable and can be currently satisfied due to the fact that VACS-equipped vehicles can obtain traffic information from different proprioceptive sensing technologies integrated within the vehicle such as cameras, sonar, navigation, radar, and lidar.

Ramp metering (RM): Controls are designed to regulate inflows from on-ramps to the freeway main road and have been used on a lot of freeways (see e.g. [174]). Because they can be implemented directly on on-ramps, any specific in-vehicle equipment does not necessarily be required to perform. The calculated inflows can be applied immediately by traffic signals.

Variable Speed Limits (VSL): Controls implement VSL on the appropriate variable message signs (VMS) on the basis of current traffic states. In the past decades, many countries have implemented these control methods (see e.g. [175]). The work in [166] has been studied and extended in many existing literatures. The work assumes that the exiting flows are specified by a central Decision Maker (DM) by computing the solution of the optimization problem for each segment-lane. Hence, all equipped vehicles traveling on the segment-lane will accept and apply their respective speeds or speed limits. This is also sufficient to impose a speed limit on unequipped vehicles based on the assumption that the proportion of equipped vehicles is sufficient.

Since the main purpose is to reduce and even eliminate traffic jams, these control strategies are considered as the decision variables in the optimal control model. To solve the considered problem for large freeway networks, it is approximated as a smooth discrete-time optimal control problem with constraints, described by a global cost function and constraints.

5.2.2 The CTM Model for the Problem

We write the conservation equation of flow and density relationship for each cell as follows:

$$\rho_{i,t+1} = \rho_{i,t} + \frac{\Delta t}{\Delta x_i}(f_{i-1,t} - f_{i,t} + r_{i,t} - s_{i,t}). \tag{5.1}$$

where $\rho_{i,t}$ and $f_{i,t}$ denote the density and flow at cell i and time step t, respectively; $r_{i,t}$ denotes the ramp metering at on-ramps i and time step t; and $s_{i,t}$ representees the out-going flow from off-ramps i at time step t; Δt refers to the size of simulation time step and Δx_i is the length of cell i. Eq. (5.1) describes the evolution of density $\rho_{i,t}$ and is directly derived from the CTM model proposed in [66, 67].

From the triangular fundamental diagram (FD) displayed in Fig. 5.1, the outflow from the cell i with a given density during the time step t can be controlled. In particular, Fig. 5.1 shows the demand function, which controls the outgoing flow according to the current density, and the supply function, which controls the inflow that relies also on the current density. It is noted that, as elaborated in [80], the displayed FD demand function offers the possibility to describe the capacity drop phenomenon when the upstream density exceeds the critical density ρ_{cr}, i.e., $\rho > \rho_{cr}$. Moreover, if the downstream cell has an on-ramp, the supply function needs to provide partial space to process the vehicles entering from the

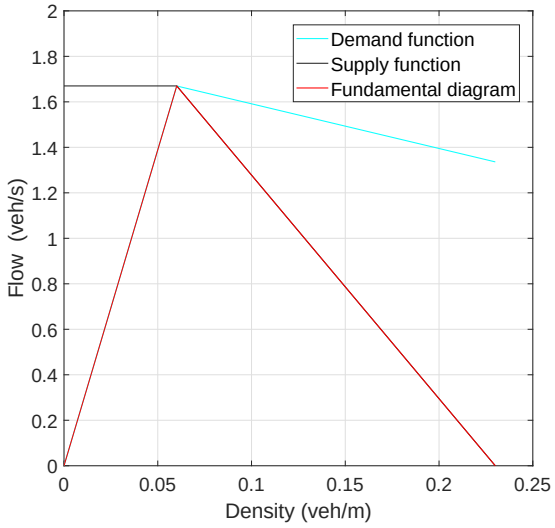


Figure 5.1: Fundamental diagram with capacity drop

on-ramp. Without control, the actual outflow is equal to the minimum between the demand part and supply part. In case of adopting the optimal control strategy, the outflows can be controlled by implementing VSL for travel vehicles or employing RM on metered on-ramps. Hence, by the model proposed in [80], the outflow at cell i during time step $(t, t+1]\Delta t$ is formulated as follows:

$$f_{i,t} = \min\{f^D_{i,t}, f^S_{i+1,t} - \theta r_{i+1,t}\}, \tag{5.2}$$

where the demand part and the supply part are:

$$f^D_{i,t} = \min\left\{v_{f,i}\rho_{i,t}, \left[C_i + \alpha C_i(\frac{\rho_{i,t} - \rho_{cr,i}}{\rho_{cr,i} - \rho_{max,i}})\right]\right\}, \tag{5.3}$$

and

$$f^S_{i+1,t} = \min\{C_{i+1}, w_{i+1}(\rho_{max,i+1} - \rho_{i+1,t})\}, \tag{5.4}$$

where $\theta > 0$ describes the size of the space that the downstream cell $i+1$ provides to hold vehicles going from the on-ramp $i+1$; $\alpha > 0$ shows the level of capacity drop; $v_{f,i}$ representees the free flow velocity at cell i; C_i denotes the capacity of cell i; $\rho_{max,i}$ refers to the jam density of cell i; w_{i+1} refers to the speed of backward wave of cell $i+1$. Here, $w_i = \frac{C_i}{\rho_{max,i} - \rho_{cr,i}} = \frac{v_{f,i}\rho_{cr,i}}{\rho_{max,i} - \rho_{cr,i}}$, and α and θ are pre-defined parameters, which can be determined according to obtained data. Equation (5.3) is equivalent to that in the CTM in [66, 67], without taking capacity drop into consideration. The min(.) operators in the constraints (5.2)-(5.4) ensure that the

traffic flow follows the fundamental diagram of the CTM by controlling the flow of vehicles between adjacent cells to be exactly equal to one of the terms contained in the min(.) operators [79].

5.2.3 Incorporating the RM impact

In this section, the impact of RM is incorporated into the developed model. Due to the uncertainty of external demand flows to each on-ramp, each on-ramp i accepts an uncontrollable demand flow $d_{i,t}$. Moreover, due to the limited space of the mainstreams of the freeway, each on-ramp outflow $r_{i,t}$ is controlled by the corresponding ramp metering measures, which may give rise to the production of ramp queue $q_{i,t}$. Therefore, the evolution on the on-ramps is expressed as follows:

$$q_{i,t+1} = q_{i,t} + \Delta t(d_{i,t} - r_{i,t}). \tag{5.5}$$

Furthermore, due to the limited storage capacities on freeway networks and on-ramps and non-negativity of the decision variables, they need to be imposed as hard constraints in the optimization problem. More specifically, the hard constraints on the on-ramp queues $q_{i,t}$ and the ramp metering actions $r_{i,t}$ are in the form of lower and upper bounds as given below:

$$0 \leq q_{i,t} \leq q_{max,i}, \tag{5.6}$$

and

$$0 \leq r_{i,t} \leq r_{max,i}. \tag{5.7}$$

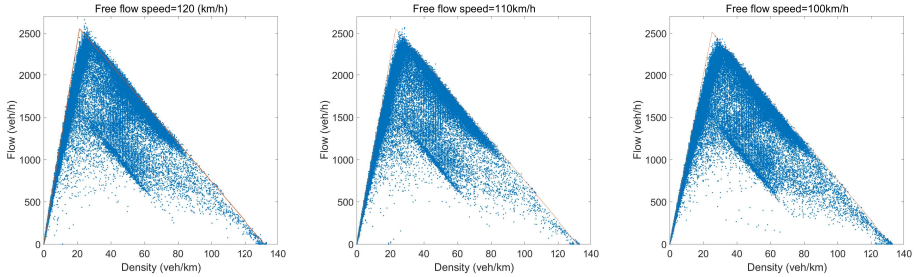
Note that Constraint (5.6) specifies the limited capacities on the on-ramps. It is important in practical situations where the on-ramp queues over-spill onto the freeway main road causing substantial congestions and delays. Under this constraint, the model can capture the potential excessive overflow, and the obtained optimal control strategies can resolve this problem.

5.2.4 Incorporating the VSL Impact

In this section, we derive a simple linear model to reflect the influence of VSL on the traffic flow.

To begin with, special VSL values are reflected in the cell-specific VSL rates b_t during $(t, t+1]\Delta t$. The VSL rates are ideal decision variables taking values from

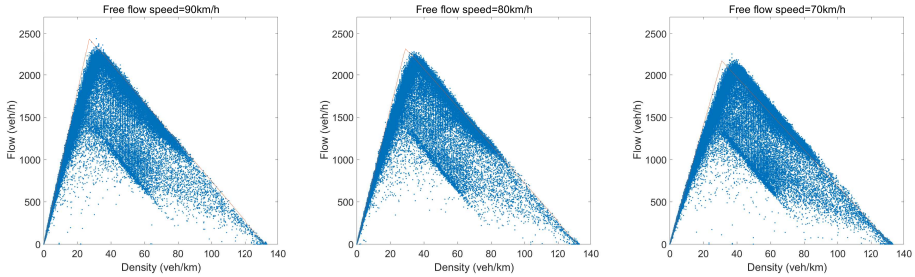
the admissible value range $b_t \in [b_{min}, 1]$, where $b_{min} \in (0, 1)$ is a lower feasible bound for VSL rates. Based on the simulation data obtained from simulation as detailed in section 5.5, we can obtain the relationship between the fundamental diagrams and different VSL rates, see Fig. 5.2. From Fig. 5.2, we can observe that implementing VSL (for any VSL values) does not increase capacity, which is consistent with the result obtained in [166]. Furthermore, using a lower VSL value may lead to capacity drop, resulting in a decrease in traffic efficiency.



(a) Fundamental diagram for free flow speed=120 (km/h)

(b) Fundamental diagram for free flow speed=110 (km/h)

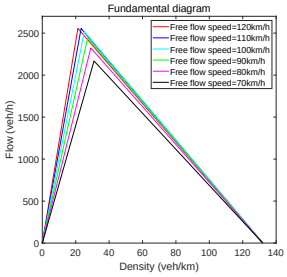
(c) Fundamental diagram for free flow speed=100 (km/h)



(d) Fundamental diagram for free flow speed=90 (km/h)

(e) Fundamental diagram for free flow speed=80 (km/h)

(f) Fundamental diagram for free flow speed=70 (km/h)



(g) Fundamental diagram

Figure 5.2: Fundamental diagrams for different VSL rates

By the assumption of VACS-equipped vehicles, drivers can control vehicle's speeds based on the obtained information of current traffic states, for instance, the density and mean speed, thereby reducing time spent to downstream. Due to the limited data available, we can only obtain partial fundamental diagrams

that are characterized for special VSL values, such as $b_t = 1, 0.9, 0.8, \ldots, b_{min}$. This is too limited to find any VSL value to improve traffic operation efficiency. In order to obtain the corresponding free flow velocity and critical density of any VSL rate b_t, we want to propose a model to express the relationship.

Now, we apply the characterized VSL rates to the CTM model (5.1)-(5.4) as the control variables. Based on the work of [166, 175, 176], we use the following two linear functions to denote the relation of b_t in (5.3)-(5.4) by introducing one auxiliary variable A as given below:

$$v_{f,i}[b_{i,t}] = v_{f,i} b_{i,t}, \tag{5.8}$$

and

$$\rho_{cr,i}[b_{i,t}] = \rho_{cr,i}(1 + A_i(1 - b_{i,t})), \tag{5.9}$$

where $v_{f,i}$ and $\rho_{cr,i}$ are the free flow velocity and critical density for the specific non-VSL values, and A_i is a constant parameter estimated based on obtained data. In this study, the least squares method is utilize to obtain the A_i.

As shown in (5.8), $b_{i,t} = 1$ if VSL is not employed or is equal to VSL-induced $v_{f,i}[b_{i,t}]$ divided by the free flow speed without VSL. Therefore, $b_{i,t} = 1$, no VSL is applied; otherwise $b_{i,t} < 1$. From Equation (5.9), we see that $A_i > 0$ and that it is a linear increasing function with reference to the decrease of $b_{i,t}$ starting from the usual non-VSL value for $b_{i,t} = 1$.

Since the VSL rate b_t should not be too large nor too small when taking into consideration of the traffic safety and practical requirements, constraints in the form of upper and lower bounds are imposed on the VSL rate b_t, i.e.,

$$b_{min,i} \leq b_{i,t} \leq 1. \tag{5.10}$$

Note that in our model, we can implement any VSL values to control the outflow of each cell to reduce traffic congestion. This model is different from that in [171], where a mode dependent VSL control action based on CTM is proposed for freeway networks without on-ramps. Compared with the model in [172], our proposed model does not require to customize many parameters, and due to its simplicity, it is applicable to much larger networks.

5.2.5 The Integrated Optimal Control Problem

With the preparations given above, we introduce a freeway optimization model on the basis of Model Predictive Control (MPC), which takes into account of VSL and RM, as shown below:

$$\begin{aligned}(\text{DMP}) \min_{r,b} Z \;=\; & \sum_{t=0}^{\text{Np}} \sum_{i=1}^{I} (\rho_{i,t} \Delta x_i \Delta t + q_{i,t} \Delta t) + \lambda_r \sum_{t=1}^{\text{Np}} \sum_{i=1}^{I} (r_{i,t} - r_{i,t-1})^2 \\ & + \lambda_{bt} \sum_{t=1}^{\text{Np}} \sum_{i=1}^{I} (b_{i,t} - b_{i,t-1})^2 + \lambda_{bs} \sum_{t=0}^{\text{Np}} \sum_{i=2}^{I} (b_{i,t} - b_{i-1,t})^2, \end{aligned} \tag{5.11}$$

subject to (5.1), (5.2), (5.3), (5.4), (5.5), (5.6), (5.7), (5.8), (5.9) and (5.10). Np is the predictive horizon of MPC and I is the number of cells. The objective function given by (5.11) consists of the weighted sum of four different terms, where the first term is linear and the last three terms are quadratic.

The first linear part denotes the Total Time Spent (TTS) by taking into account both the mainstream travel time of freeways as well as the on-ramps queuing time. It is the most important indicator for the evaluation of the traffic flow efficiency. The quadratic terms are penalty terms and are designed to punish the variation of decision variables within two consecutive time steps or adjacent cells. They are used to reduce, or even restrain, the temporal and spatial fluctuations of the control variables, which contribute little to the result. The first penalty part (weighted by λ_r) is associated to the time-variations of the on-ramp flows; and the last two terms (weighted by λ_{bt} and λ_{bs}) have purposes of penalising, respectively, the time and space variations of VSL rates $b_{i,t}$. Parameters $\lambda_r, \lambda_{bt}, \lambda_{bs}$ are given constants, and they are set as $\lambda_r = \lambda_{bt} = \lambda_{bs} = 0.00001$.

It is challengeable to resolve Problem DMP directly due to the appearing of the min(.) operators in Constraints (5.2), (5.3) and (5.4). Traditionally, there are two methods (Big-M method and inequality relaxation) to solve it. The min(.) operators can be equivalently denoted as linear constraints using the big-M technique [78]. This approach requires a lot of auxiliary binary variables to be introduced into the problem, thereby increasing the complexity of the problem. Alternatively, one constraint involving the min(.) operator can be relaxed into several linear inequality constraints without inserting any auxiliary binary variables. These relaxed constraints can guarantee the conservation of flow. However, they do not ensure that the minimum values of the right-hand-side of these constraints can be taken by the vehicle flow between neighbouring cells. As a re-

sult, the flow may be less than the minimum and the flow holding-back problem may occur. Even though this method decreases the complexity of the problem, the vehicle flow does not satisfy the fundamental diagram of CTM, and thus the vehicle flow may be impractical [78]. To efficiently handle Problem DMP for large scale networks, we will develop a novel smooth model and a more efficient solution technique later.

5.3 Approximation Approach and Model Construction

5.3.1 Smooth approximation

In Section 5.2, we model traffic flow dynamics with VSL and RM being taken into consideration based on CTM. Due to computational difficulties, we propose an approximate smoothing model in this section.

Let $V = \{c = [r_1, r_2, \ldots, r_I, b_1, b_2, \ldots, b_I] \in \mathbb{R}^{2I} : 0 \leq r_i \leq r_{max,i}, b_{min,i} \leq b_i \leq 1, \forall i = 1, \ldots, I\}$, where $r_{max,i}$ and $b_{min,i}$ are given constants. Let c be a control sequence $\{c_t : t = 1, \ldots, \mathrm{Np} - 1\}$ in V . Then, c is called an admissible control and $\mathcal{U}$ represents the class of all such admissible controls. For a control c in $\mathcal{U}$, if it satisfies constraints (5.6), (5.7) and (5.10), then c is called a feasible control sequence. Based on Equations (5.3), (5.4), (5.8) and (5.9), for any i and t, we can denote $f^D_{i,t}$ and $f^S_{i+1,t}$ equivalently as follows:

$$\begin{aligned} & f^D_{i,t} \\ & = \frac{1}{2}[v_{f,i}[b_{i,t}]\rho_{i,t} + C_i[b_{i,t}] + \alpha C_i[b_{i,t}](\frac{\rho_{i,t} - \rho_{cr,i}[b_{i,t}]}{\rho_{cr,i}[b_{i,t}] - \rho_{max,i}}) \\ & - |v_{f,i}[b_{i,t}]\rho_{i,t} - C_i[b_{i,t}] - \alpha C_i[b_{i,t}](\frac{\rho_{i,t} - \rho_{cr,i}[b_{i,t}]}{\rho_{cr,i}[b_{i,t}] - \rho_{max,i}})|], \end{aligned}$$

and

$$\begin{aligned} f^S_{i+1,t} = \frac{1}{2}[& w_{i+1}[b_{i+1,t}](\rho_{max,i+1} - \rho_{i+1,t}) + C_{i+1}[b_{i+1,t}] \\ & - |w_{i+1}[b_{i+1,t}](\rho_{max,i+1} - \rho_{i+1,t}) - C_{i+1}[b_{i+1,t}]|]. \end{aligned}$$

Let $\epsilon > 0$ be a given small number. $f^{D\epsilon}_{i,t}$ and $f^{S\epsilon}_{i+1,t}$ are used to approximate $f^D_{i,t}$

and $f^S_{i+1,t}$ respectively as follows:

$$f^{D\epsilon}_{i,t} = \frac{1}{2}[v_{f,i}[b_{i,t}]\rho_{i,t} + C_i[b_{i,t}] + \alpha C_i[b_{i,t}](\frac{\rho_{i,t} - \rho_{cr,i}[b_{i,t}]}{\rho_{cr,i}[b_{i,t}] - \rho_{max,i}}) - \sqrt{(v_{f,i}[b_{i,t}]\rho_{i,t} - C_i[b_{i,t}] - \alpha C_i[b_{i,t}](\frac{\rho_{i,t} - \rho_{cr,i}[b_{i,t}]}{\rho_{cr,i}[b_{i,t}] - \rho_{max,i}}))^2 + \frac{\epsilon^2}{4}}], \quad (5.12)$$

and

$$f^{S\epsilon}_{i+1,t} = \frac{1}{2}[w_{i+1}[b_{i+1,t}](\rho_{max,i+1} - \rho_{i+1,t}) + C_{i+1}[b_{i+1,t}] - \sqrt{(w_{i+1}[b_{i+1,t}](\rho_{max,i+1} - \rho_{i+1,t}) - C_{i+1}[b_{i+1,t}])^2 + \frac{\epsilon^2}{4}}]. \quad (5.13)$$

Note that the functions $f^{D\epsilon}_{i,t}$ and $f^{S\epsilon}_{i+1,t}$ are smooth and differentiable, and keep the properties of $f^D_{i,t}$ and $f^S_{i+1,t}$ (including the changing trends). To characterize the precision of the approximative model, we need to estimate the bounds of $f^D_{i,t} - f^{D\epsilon}_{i,t}$ and $f^S_{i+1,t} - f^{S\epsilon}_{i+1,t}$. By the definitions of $f^{D\epsilon}_{i,t}$ and $f^{S\epsilon}_{i+1,t}$, it is easy to obtain that

$$0 \leq f^D_{i,t} - f^{D\epsilon}_{i,t} \leq \frac{\epsilon}{4}. \quad (5.14)$$

Inequality (5.14) gives an estimate on the bound between the primal demand functions $f^D_{i,t}$ and the approximate demand function $f^{D\epsilon}_{i,t}$. Clearly, the difference can be controlled arbitrarily small by reducing the value of the parameter ϵ. Similarly, we have

$$0 \leq f^S_{i+1,t} - f^{S\epsilon}_{i+1,t} \leq \frac{\epsilon}{4}. \quad (5.15)$$

Next, we introduce the following auxiliary variable $f^{\epsilon\epsilon}_{i,t}$:

$$f^{\epsilon\epsilon}_{i,t} = \frac{1}{2}(f^{D\epsilon}_{i,t} + f^{S\epsilon}_{i+1,t} - \theta r_{i+1,t} - |f^{D\epsilon}_{i,t} - f^{S\epsilon}_{i+1,t} + \theta r_{i+1,t}|). \quad (5.16)$$

Using the same approximating method, we obtain

$$f^{\epsilon}_{i,t} = \frac{1}{2}[f^{D\epsilon}_{i,t} + f^{S\epsilon}_{i+1,t} - \theta r_{i+1,t} - \sqrt{(f^{D\epsilon}_{i,t} - f^{S\epsilon}_{i+1,t} + \theta r_{i+1,t})^2 + \frac{\epsilon^2}{4}}]. \quad (5.17)$$

Based on equality (5.16) and equality (5.17), we have

$$0 \leq f^{\epsilon\epsilon}_{i,t} - f^{\epsilon}_{i,t} \leq \frac{\epsilon}{4}. \quad (5.18)$$

Note that we can control the parameter ϵ to adjust the approximation accuracy of inequalities (5.14), (5.15) and (5.18). As $\epsilon \to 0$, the difference between the primal variables and the approximating variables will converge to zero. Based on the above preparatory work, we can formulate our approximation model in the next section.

5.3.2 Proposed approximation model

After approximating the 'min' operators, we can obtain the corresponding optimal control model. Hence, the new optimization model can be described as follows:

$$\begin{aligned}(\mathrm{DMP}^{\epsilon}) \min_{r,b} Z &= \sum_{t=0}^{\mathrm{Np}} \sum_{i=1}^{I} (\rho_{i,t}^{\epsilon} \Delta x_i \Delta t + q_{i,t} \Delta t) + \lambda_r \sum_{t=1}^{\mathrm{Np}} \sum_{i=1}^{I} (r_{i,t} - r_{i,t-1})^2 \\ &\quad + \lambda_{bt} \sum_{t=1}^{\mathrm{Np}} \sum_{i=1}^{I} (b_{i,t} - b_{i,t-1})^2 + \lambda_{bs} \sum_{t=0}^{\mathrm{Np}} \sum_{i=2}^{I} (b_{i,t} - b_{i-1,t})^2,\end{aligned}$$

subject to

$$\rho_{i,t+1}^{\epsilon} = \rho_{i,t}^{\epsilon} + \frac{\Delta t}{\Delta x_i} \times (f_{i-1,t}^{\epsilon} - f_{i,t}^{\epsilon} + r_{i,t} - s_{i,t}), \forall i, t, \tag{5.19}$$

$$0 \leq \rho_{i,t}^{\epsilon} \leq \rho_{max,i}, \forall i, t, \tag{5.20}$$

and (5.5), (5.6), (5.7), (5.8), (5.9), (5.10) and (5.17). Note that Problem DMP^{ϵ} is a discrete and smooth optimal control problem, and it can be resolved effectively utilizing gradient decent algorithms, for example, SQP [158].

Remark 1. Solutions of Problem DMP^{ϵ} will converge to that of Problem DMP as $\epsilon \to 0$. The proof can be divided into three parts. The first part is to compute the lower and upper bounds of $f_{i,t} - f_{i,t}^{\epsilon}$ followed by the proof $\lim_{\epsilon \to 0} f_{i,t} - f_{i,t}^{\epsilon} = 0$. The second part is to prove that $\lim_{\epsilon \to 0} \rho_{i,t}^{*\epsilon} = \overline{\rho}_{i,t}$ and $\lim_{\epsilon \to 0} \overline{\rho}_{i,t}^{\epsilon} = \rho_{i,t}^{*}$ by the mathematical induction, where $\{\rho_{i,t}^{*\epsilon}\}$ and $\{\rho_{i,t}^{\epsilon}\}$ denotes the state variables of Problem DMP^{ϵ}, and $\overline{\rho}_{i,t}$ and $\rho_{i,t}^{*}$ denotes state variables of Problem DMP. The superscript $*$ indicates that the variable concerned is optimal. Finally, we use the squeeze theorem to prove the solution convergence. The detail of proof is similar to that given in [177,178] and hence is omitted here.

5.4 Customized Solution Technique

In order to resolve Problem DMP^{ϵ}, we need to calculate the gradients of the objective and the constraint functions with respect to the control variables. In

this section, the required gradient formulas are derived.

5.4.1 Gradients of objective and constraint functions

Let $z_t = \{\rho^{\epsilon}_{1,t}, \rho^{\epsilon}_{2,t}, \ldots, \rho^{\epsilon}_{I,t}, q_{1,t}, q_{2,t}, \ldots, q_{I,t}\}^{\top} \in \mathbb{R}^{2I}$ denote the state variable and $c_t = \{r_{1,t}, r_{2,t} \ldots, r_{I,t}, b_{1,t}, b_{2,t}, \ldots, b_{I,t}\}^{\top} \in \mathbb{R}^{2I}$ be the decision variable. Then, we reformulate the state equations of (5.19) and (5.5) in the form given below:

$$
\begin{aligned}
z_{t+1} \quad = \quad & \begin{bmatrix} \rho^{\epsilon}_{1,t} \\ \rho^{\epsilon}_{2,t} \\ \vdots \\ \rho^{\epsilon}_{I,t} \\ q_{1,t} \\ q_{2,t} \\ \vdots \\ q_{I,t} \end{bmatrix} + \begin{bmatrix} \frac{\Delta t}{\Delta x_1} \cdot d_{0,t} \\ \frac{\Delta t}{\Delta x_2} \cdot f^{\epsilon}_{1,t} \\ \vdots \\ \frac{\Delta t}{\Delta x_I} \cdot f^{\epsilon}_{I-1,t} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} - \begin{bmatrix} \frac{\Delta t}{\Delta x_1} \cdot f^{\epsilon}_{1,t} \\ \frac{\Delta t}{\Delta x_2} \cdot f^{\epsilon}_{2,t} \\ \vdots \\ \frac{\Delta t}{\Delta x_I} \cdot f^{\epsilon}_{I,t} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} \frac{\Delta t}{\Delta x_1} \cdot c_{1,t} \\ \frac{\Delta t}{\Delta x_2} \cdot c_{2,t} \\ \vdots \\ \frac{\Delta t}{\Delta x_I} \cdot c_{I,t} \\ -\Delta t \cdot c_{1,t} \\ -\Delta t \cdot c_{2,t} \\ \vdots \\ -\Delta t \cdot c_{I,t} \end{bmatrix} \\
& + \begin{bmatrix} -\frac{\Delta t}{\Delta x_1} \cdot s_{1,t} \\ -\frac{\Delta t}{\Delta x_2} \cdot s_{2,t} \\ \vdots \\ -\frac{\Delta t}{\Delta x_I} \cdot s_{I,t} \\ \Delta t \cdot d_{1,t} \\ \Delta t \cdot d_{2,t} \\ \vdots \\ \Delta t \cdot d_{I,t} \end{bmatrix}, \qquad (5.21)
\end{aligned}
$$

where $d_{0,t}$ refers to the demand from the upstream of cell 1 at the time step t and $d_{i,t}$ denotes the demand from the on-ramp i and the same time step t and they are given. Let $\mathcal{F}(t, z_t, c_t)$ be the right hand side of the difference equation (5.21). It includes the state variable, the decision variable and the time. Let the initial

condition be

$$z_0 = [\rho_{1,0}, \ldots, \rho_{I,0}, q_{1,0}, \ldots, q_{I,0}]^\top \in \mathbb{R}^{2I}. \tag{5.22}$$

Let

$$\begin{aligned}\Phi_0(\mathrm{Np}, z_{\mathrm{Np}}, c_{\mathrm{Np}}) &= \sum_{i=1}^{I}(\rho^{\epsilon}_{i,\mathrm{Np}}\Delta t\Delta x_i + q_{i,\mathrm{Np}}\Delta t) + \lambda_r \sum_{i=1}^{I}(c_{i,\mathrm{Np}} - c_{i,\mathrm{Np}-1})^2 \\ &\quad +\lambda_{bt}\sum_{i=1+I}^{2I}(c_{i,\mathrm{Np}} - c_{i,\mathrm{Np}-1})^2 + \lambda_{bs}\sum_{i=2+I}^{2I}(c_{i,\mathrm{Np}} - c_{i-1,\mathrm{Np}})^2,\end{aligned}$$

$$\begin{aligned}L_0(t, z_t, c_t) &= \sum_{i=1}^{I}(\rho^{\epsilon}_{i,t}\Delta t\Delta x_i + q_{i,t}\Delta t) + \lambda_r \sum_{i=1}^{I}(c_{i,t} - c_{i,t-1})^2 \\ &\quad +\lambda_{bt}\sum_{i=1+I}^{2I}(c_{i,t} - c_{i,t-1})^2 + \lambda_{bs}\sum_{i=2+I}^{2I}(c_{i,t} - c_{i-1,t})^2, \\ &\quad t = 1, \ldots, \mathrm{Np}-1,\end{aligned}$$

$$L_0(0, z_0, c_0) = \sum_{i=1}^{I}(\rho^{\epsilon}_{i,0}\Delta t\Delta x_i + q_{i,0}\Delta t) + \lambda_{bs}\sum_{i=2+I}^{2I}(c_{i,0} - c_{i-1,0})^2.$$

Then,

$$h_0(c) = \Phi_0(\mathrm{Np}, z_{\mathrm{Np}}, c_{\mathrm{Np}}) + \sum_{t=0}^{\mathrm{Np}-1} L_0(t, z_t, c_t), \tag{5.23}$$

where $h_0(c)$ denotes the objective function. Similarly, the constraints (5.20) and (5.6) can be rewritten as below:

$$\begin{aligned}h_k(c) &= -\rho^{\epsilon}_{i,t}, \;\; 1 \le k \le I(\mathrm{Np}+1),\; i = 1\ldots, I,\; t = 0, \ldots, \mathrm{Np}, \\ h_k(c) &= \rho^{\epsilon}_{i,t} - \rho_{max,i},\; I(\mathrm{Np}+1)+1 \le k \le 2I(\mathrm{Np}+1), \\ &\quad i = 1\ldots, I,\; t = 0, \ldots, \mathrm{Np}, \\ h_k(c) &= -q_{i,t},\; 2I(\mathrm{Np}+1)+1 \le k \le 3I(\mathrm{Np}+1), \\ &\quad i = 1\ldots, I,\; t = 0, \ldots, \mathrm{Np}, \\ h_k(c) &= q_{i,t} - q_{max,i},\; 3I(\mathrm{Np}+1)+1 \le k \le 4I(\mathrm{Np}+1), \\ &\quad i = 1\ldots, I,\; t = 0, \ldots, \mathrm{Np}.\end{aligned}$$

Note that these constraint functions are in the same standard form as the objective function. Thus, a unified method can be used to calculate the gradients of the objective and constraint functions. Define

$$c = [(c_0)^\top, (c_1)^\top, \ldots, (c_{\mathrm{Np}-1})^\top]^\top.$$

Let the decision vector c be perturbed by $\eta\widehat{c}$, where $\eta > 0$ denotes a real number and $\widehat{c}$ refers to a random but fixed disturbance of c given by

$$\widehat{c} = [(\widehat{c}_0)^\top, (\widehat{c}_1)^\top, \ldots, (\widehat{c}_{\mathrm{Np}-1})^\top]^\top.$$

Then, we have

$$c(\eta) = c + \eta\widehat{c} = [(c_0(\eta))^\top, (c_1(\eta))^\top, \ldots, (c_{\mathrm{Np}-1}(\eta))^\top]^\top,$$

where

$$c_t(\eta) = c_t + \eta\widehat{c}_t, t = 0, \ldots, \mathrm{Np} - 1.$$

Therefore, the state of the system will be disturbed, and the objective and constraint functions will also be disturbed.

Let

$$z_t(\eta) = z_t(c(\eta)), t = 1, \ldots, \mathrm{Np}. \tag{5.24}$$

Then,

$$z_{t+1}(\eta) = \mathcal{F}(t, z_t(\eta), c_t(\eta)). \tag{5.25}$$

The variation of the state for $t = 0, 1, \ldots, \mathrm{Np} - 1$ is:

$$\Delta z_{t+1} = \frac{dz_{t+1}(\eta)}{d\eta}|_{\eta=0} = \frac{\partial\mathcal{F}(t, z_t, c_t)}{\partial z_t}\Delta z_t + \frac{\partial\mathcal{F}(t, z_t, c_t)}{\partial c_t}\widehat{c}_t \tag{5.26}$$

with

$$\Delta z_0 = 0. \tag{5.27}$$

For the $k-$th function ($k = 0$ refers to the objective function), we get

$$\begin{aligned}\frac{\partial h_k(c)}{\partial c}\widehat{c} &= \lim_{\eta\to 0}\frac{h_k(c(\eta)) - h_k(c)}{\eta} \equiv \frac{dh_k(c(\eta))}{d\eta}\mid_{\eta=0} \\ &= \sum_{t=0}^{\mathrm{Np}-1}[\frac{\partial L_k(t,z_t,c_t)}{\partial z_t}\Delta z_t + \frac{\partial L_k(t,z_t,c_t)}{\partial c_t}\widehat{c}_t] + \frac{\partial \Phi_k(z_{\mathrm{Np}})}{\partial z_{\mathrm{Np}}}\Delta z_{\mathrm{Np}}.\end{aligned} \tag{5.28}$$

For each $k = 0, 1, \ldots, 4I(\mathrm{Np}+1)$, define the Hamiltonian

$$\mathcal{H}_k(t, z_t, c_t, \lambda_{t+1}^k) = L_k(t, z_t, c_t) + (\lambda_{t+1}^k)^\top \mathcal{F}(t, z_t, c_t),$$

where $\lambda_t^k \in \mathbb{R}^{2I}, t = \mathrm{Np}, \mathrm{Np}-1, \ldots, 1$, represents the co-state sequence for the $k-$th standard constraint. Then, by (5.28), we get that

$$\begin{aligned}\frac{\partial h_k(c)}{\partial c}\widehat{c} &= \frac{\partial \Phi_k(z_{\mathrm{Np}})}{\partial z_{\mathrm{Np}}}\Delta z_{\mathrm{Np}} + \sum_{t=0}^{\mathrm{Np}-1}\{\frac{\partial \mathcal{H}_k(t,z_t,c_t,\lambda_{t+1}^k)}{\partial z_t}\Delta z_t \\ &\quad -(\lambda_{t+1}^k)^\top\frac{\partial \mathcal{F}(t,z_t,c_t)}{\partial z_t}\Delta z_t + \frac{\partial \mathcal{H}_k(t,z_t,c_t,\lambda_{t+1}^k)}{\partial c_t}\widehat{c}_t \\ &\quad -(\lambda_{t+1}^k)^\top\frac{\partial \mathcal{F}(t,z_t,c_t)}{\partial c_t}\widehat{c}_t\}.\end{aligned}$$

By (5.26) and (5.27), it holds

$$\Delta z_{t+1} = \frac{\partial \mathcal{F}(t,z_t,c_t)}{\partial z_t}\Delta z_t + \frac{\partial \mathcal{F}(t,z_t,c_t)}{\partial c_t}\widehat{c}_t. \tag{5.29}$$

Using the following system of difference equations, the co-state λ_t^k can be determined:

$$(\lambda_t^k)^\top = \frac{\partial \mathcal{H}_k(t,z_t,c_t,\lambda_{t+1}^k)}{\partial z_t}, t = \mathrm{Np}-1, \mathrm{Np}-2, \ldots, 1, \tag{5.30}$$

with

$$(\lambda_{\mathrm{Np}}^k)^\top = \frac{\partial \Phi_k(z_{\mathrm{Np}})}{\partial z_{\mathrm{Np}}}. \tag{5.31}$$

By virtue of (5.29), (5.30), (5.31), (5.24), (5.25) and the arbitrariness of $\widehat{c}$, the following gradient formula follows readily:

$$\frac{\partial h_k(c)}{\partial c} = [\frac{\partial \mathcal{H}_k(0,z_0,c_0,\lambda_1^k)}{\partial c_0}, \ldots, \frac{\partial \mathcal{H}_k(\mathrm{Np}-1, z_{\mathrm{Np}-1}, c_{\mathrm{Np}-1}, \lambda_{\mathrm{Np}}^k)}{\partial c_{\mathrm{Np}-1}}]. \tag{5.32}$$

Now the gradient calculation is summarized as the following theorem.

Theorem 5.1. *Taking Problem* (DMP$^\epsilon$) *into account, for each* $k = 0, 1, \ldots, 4I(\mathrm{Np}+1)$*, the gradient of* $h_k(c)$ *with respect to the control variable* c *is denoted by (5.32), where* $c = [(c_0)^\top, (c_1)^\top, \ldots, (c_{\mathrm{Np}-1})^\top]^\top$.

5.4.2 Algorithm

Problem DMP$^\epsilon$ is in essence a nonlinear mathematical optimization problem with control vector c. Many approaches can be used to solve it, such as SQP. To apply these optimization algorithms, for each $c \in \mathcal{U}$, the values and corresponding gradients of objective function $h_0(c)$ and constraint functions $h_k(c)$, $k = 0, 1, \ldots, 4I(\mathrm{Np}+1)$, are required. Detailed computation is shown in Algorithm 2.

Algorithm 2 : Algorithm to Calculate Gradients

1: Initialization: For a given $c \in \mathcal{U}$ and initial condition (5.22).
2: Output: Calculate the solution $z_{t+1}(c)$, $t = 0, 1, \ldots, \mathrm{Np}-1$ of system (5.21) forward in time from $t = 0$ to $t = \mathrm{Np}-1$.
3: **for** $k = 0$ to $4I(\mathrm{Np}+1)$ **do**
4: Compute $\Phi_k(\mathrm{Np}, z_{\mathrm{Np}}, c_{\mathrm{Np}})$.
5: **for** $t = 0$ to $\mathrm{Np}-1$ **do**
6: Compute $L_k(t, z_t, c_t)$.
7: **end for**
8: Calculate $\sum_{t=0}^{\mathrm{Np}-1} L_k(t, z_t, c_t)$.
9: Compute $h_k(c) = \Phi_k(\mathrm{Np}, z_{\mathrm{Np}}, c_{\mathrm{Np}}) + \sum_{t=0}^{\mathrm{Np}-1} L_k(t, z_t, c_t)$.
10: **for** $t = \mathrm{Np}$ to 1 **do**
11: Resolve the co-state system (5.30) and (5.31) backward in time. Let $\lambda_t^k(c)$ denote the solution.
12: **end for**
13: Compute the gradient of $h_k(c)$ based on (5.32).
14: **end for**

In Algorithm 2, Step 2 is to calculate the value of $z_t(c)$ with respect to each given c. Then, $h_k(c)$ can be obtained in Step 9. Furthermore, the co-state system, (5.30) and (5.31), is computed backward in time from $t = \mathrm{Np}$ to $t = 1$ to acquire $\lambda_t^k(c)$. Finally, Step 13 obtains the corresponding gradients by (5.32).

5.5 Numerical Application

In this section, the proposed approximation model is verified in the microscopic simulated environment SUMO (Simulation of Urban Mobility) for various de-

mands scenarios. We compare the obtained results with baselines.

5.5.1 Case one

A The application framework

For this case, we select one road network, which is in the Perth vicinity and is one of the most congested freeway roads in the city of Perth, Australia. The road layout is shown in Figure 5.3. The extracted section has a total length of around 7 km with 5 on-ramps and 2 off-ramps, and is divided into 14 cells with each cell 500 m. For cells $i = 3, 4, 5, 11, 12$, a ramp metering control is utilize to limit the in-flow of vehicles. Off-ramps are located at cell $i = 2, 10$.

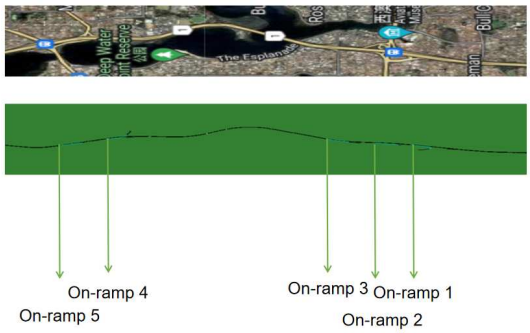


Figure 5.3: Road layout

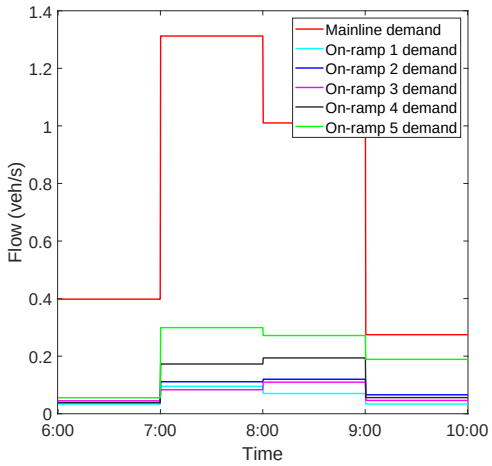


Figure 5.4: Mainline demand and on-ramp demands

The required data is collected through the simulation with no control on various scenarios. We randomly generate 100 scenarios with different traffic demand flows. The duration of each simulation is set as 4 hours (i.e. 14400 seconds). The simulation period is divided into peak and off-peak hours. For peak hours, the

in-flows for the main road and the on-ramps are generated stochastically within the range of $[3000, 6000]$ and $[200, 1000]$, respectively. For off-peak hours, the total in-flows for the main road and the on-ramps would vary within the range of $[1000, 3000]$ and $[100, 400]$, respectively. 90 scenarios are utilized for obtaining the fundamental diagrams and parameter A. We randomly choose a scenario to be the test case in the last 10 scenarios. Corresponding demand flows of main road and on-ramps are shown in Fig. 5.4. For the out-going flow $s_{i,t}$ leaving from off-ramps, let $s_{i,t} = 0.1 f_{i,t}$. We use the least square method to compute model parameters and summarize them in Table 5.1.

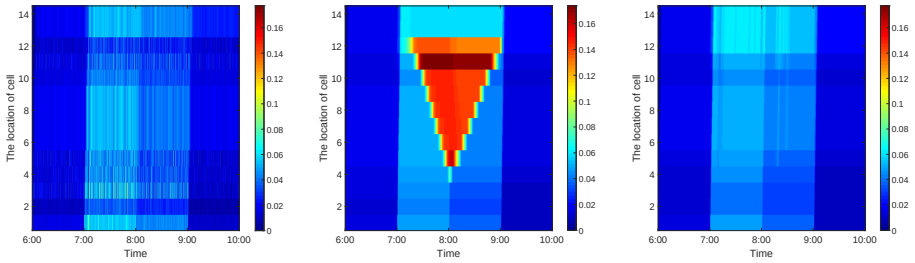
Table 5.1: Model parameters

Name of parameter	Value	Name of parameter	Value
$A_i, i = 1, \ldots, 14$	1.04	$v_i, i = 1, \ldots, 14$	27.778m/s
$w_i, i = 1, \ldots, 14$	5.2m/s	$\Delta x_i, i = 1, \ldots, 14$	500m
$r_{max,j}, j = 1, \ldots, 5$	0.5veh/s	$q_{max,i}, i = 1, \ldots, 5$	60veh
$C_i, i = 1, \ldots, 12$	1.708veh/s	$C_i, i = 13, 14$	2.278veh/s
$\rho_{max,i}, i = 1, \ldots, 12$	0.52veh/m	$\rho_{max,i}, i = 13, 14$	0.39veh/m
$\rho_{cr,i}, i = 1, \ldots, 12$	0.062veh/m	$\rho_{cr,i}, i = 13, 14$	0.082veh/m
α	0.34	θ	0.2
Δt	15s	I	14
Np	17	Nc	8

B Results and Analysis

The evolutions of density obtained from simulation, using our model with no control, only RM, only VSL, and both VSL and RM are in Fig. 5.5, respectively. Note that there is a difference between the simulation and our model with no control. This reason is that the proposed model highly dependent on the parameters obtained and the some assumptions. Due to limited data, the obtained model based on these parameters cannot fully reproduce the simulated model. By observing, we find all models can show the change of traffic condition such as free flow state and congestion state. Compared the results, our model with both VSL and RM control outperforms those with no control or with only RM or VSL.

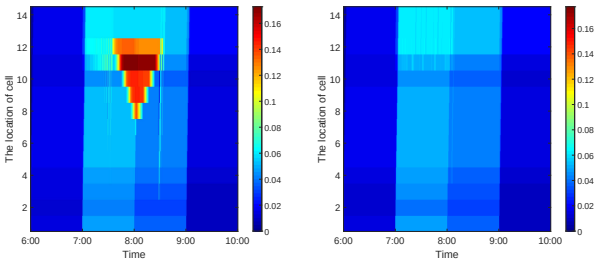
The corresponding results of the total time spent and ramp delay are shown



(a) Density (veh/m) for simulation

(b) Density (veh/m) for no control

(c) Density (veh/m) for RM



(d) Density (veh/m) for VSL

(e) Density (veh/m) for RM and VSL

Figure 5.5: Traffic density for simulation, our model with no control, only RM, only VSL and both VSL and RM

in Table 5.2. Obviously, our model only with RM, VSL and RM and VSL can optimize the time spent and improve the traffic jams. By comparison, collaborative implementation of our approximate model of RM and VSL is better than other models. Compared with results obtained from simulation with no control, our approximate model with both RM and VSL being implemented can reduce 29.36% and 95.265% of the total time spent and ramp delay, respectively. Moreover, we find that holding more vehicles on on-ramps to keep main road decongested might not improve the total performance. This result is consistent with case one.

5.5.2 Case two

A The application framework

In this application, we carry out a practical study using OpenStreeMap to extract the road layout and simulating in the SUMO simulation environment, refer to Fig. 5.6 for the road layout. The selected road is in the Shapingba vicinity, which is one of the most congested freeway roads in the city of Chongqing, China. The extracted section has a total length of around 24 km with 6 on-ramps and 4 off-ramps, and is divided into 30 cells with each cell 800 m, where cells $i = 5, \ldots, 21$, have two lanes and others have three lanes. For each ramp $i = 4, 16, 17, 19, 22, 25,$

Table 5.2: Total time spent and ramp delay with Np = 17, Nc = 8

Name of model	Total time spent (h)	Ramp delay (h)	Percentage of reduction for total time spent	Percentage of reduction for ramp delay
Simulation	1255.655h	434.3h	0%	0%
No control	1251.744h	0h	0.311%	100%
Only RM	994.756h	144.673h	20.778%	66.689%
Only VSL	1000.716h	0h	20.303%	100%
VSL and RM	886.999h	20.562h	29.36%	95.265%

there is a ramp metering to control the in-flow of vehicles. Off-ramps are located at cell $i = 4, 15, 21, 24$.

The required data for the proposed approximation model is sampled through the simulation with no control on varied scenarios. We generate 60 scenarios with different traffic flows. The duration of each simulation is set as 4 hours (i.e. 14400 seconds). The simulation period is divided into peak and off-peak hours. For off-peak hours, the total in-flows for the main road and the on-ramps are generated randomly within the range of $[1000, 2000]$ and $[100, 400]$, respectively. For peak hours, the in-flows for the main road and the on-ramps would vary within the range of $[3000, 5000]$ and $[200, 800]$, respectively. 50 scenarios are used for obtaining the fundamental diagrams and parameter A. We stochastically choose a scenario to be the test case in the last 10 scenarios. Related demand flows of main road and on-ramps are presented in Fig. 5.7. For the out-going flow $s_{i,t}$ leaving from off-ramps, we let $s_{i,t} = 0.1 f_{i,t}$. We summarize model parameters in Table 5.3.

B Results and Analysis

Fig. 5.8 presents the evolution of density obtained from the simulation, using our model with no control, only RM, only VSL, and both VSL and RM, respectively. From Figs. 5.8(a)-5.8(e), it is obvious that all models can show the change of traffic condition such as free flow state and congestion state. By observing, we find that the models with only RM, only VSL and both VSL and RM can decrease the density especially for congested conditions, and hence mitigate traffic jams. Obviously, our model with both VSL and RM control outperforms those with no

Table 5.3: Model parameters

Name of parameter	Value	Name of parameter	Value
$A_i, i = 1, \ldots, 30$	1.09	$v_i, i = 1, \ldots, 30$	33.333m/s
$w_i, i = 1, \ldots, 30$	6.414m/s	$\Delta x_i, i = 1, \ldots, 30$	800m
$C_i, i = 5, \ldots, 21$	1.42veh/s	$C_i, i = 1, 2, 3, 4, 22, \ldots, 30$	2.13veh/s
$r_{max,j}, j = 1, \ldots, 6$	0.5veh/s	$q_{max,i}, i = 1, \ldots, 6$	60veh
$\rho_{max,i}, i = 5, \ldots, 21$	0.264veh/m	$\rho_{max,i}, i = 1, 2, 3, 4, 22, \ldots, 30$	0.396veh/m
$\rho_{cr,i}, i = 5, \ldots, 21$	0.043veh/m	$\rho_{cr,i}, i = 1, 2, 3, 4, 22, \ldots, 30$	0.064veh/m
α	0.5	θ	0.1
Np	36	Nc	6
Δt	20s	I	30

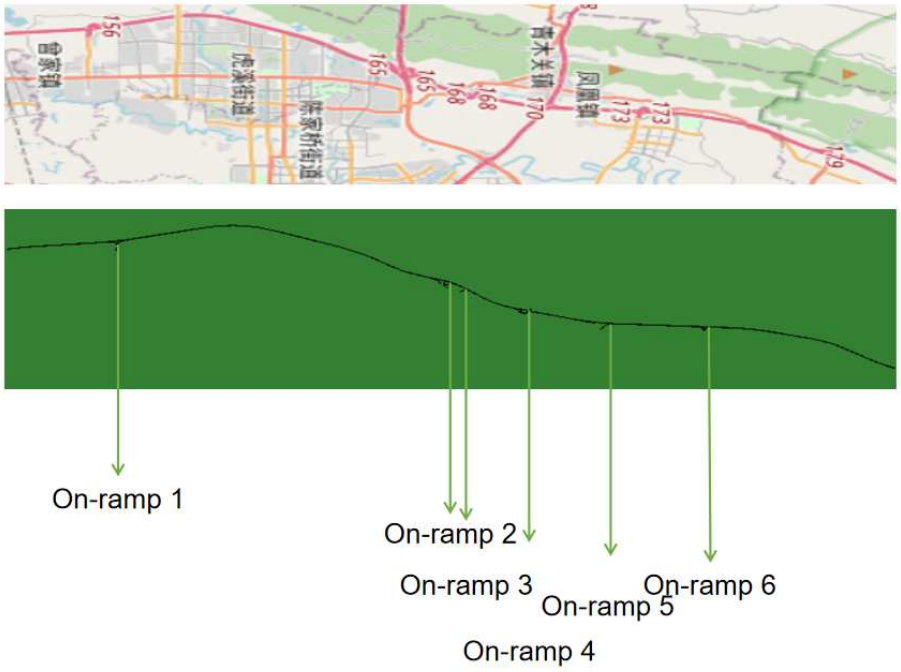


Figure 5.6: Road layout

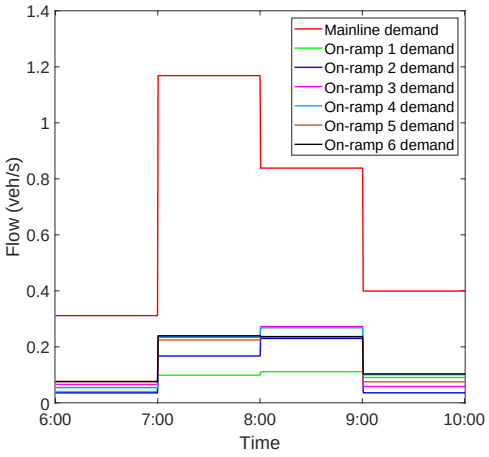


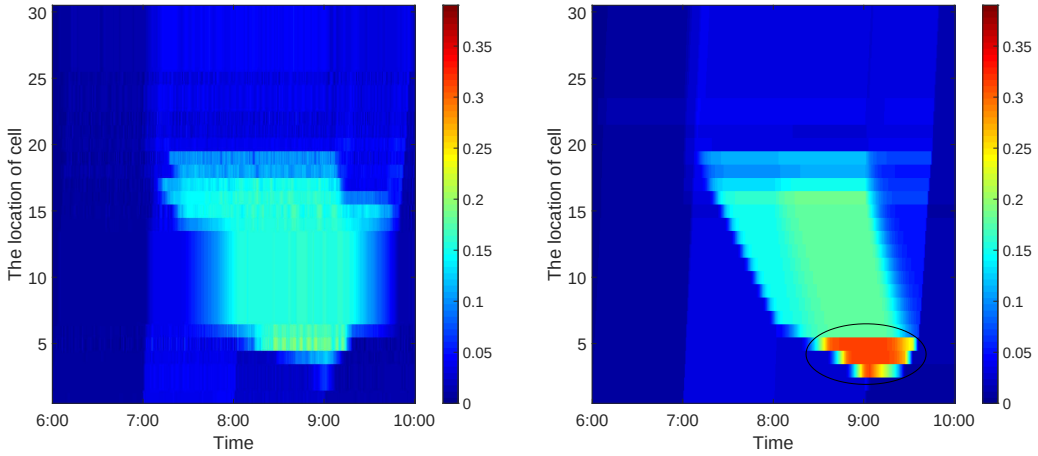
Figure 5.7: Mainline demand and on-ramp demands

control or with only RM or VSL.

Fig. 5.9 presents the evolution of flow obtained from simulation, using our model with no control, only RM, only VSL, and both RM and VSL, respectively. Observing Figs. 5.9(a)-5.9(e), we conclude that all models can show the change trend of traffic flow. Moreover, our proposed model with only RM, only VSL, and both RM and VSL can overcome the flow holding-back problem, which implies that our proposed approximated model is efficient. In comparison with models for only RM, only VSL, and both RM and VSL, our model with both RM and VSL has the best performance. This is consistent with the results in Fig. 5.8.

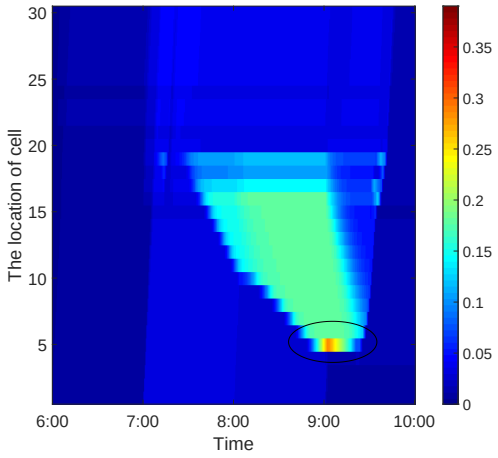
The evolution of queue length and ramp metering for our model with only RM, and with both RM and VSL are shown in Figs. 5.10 and 5.11. From Figs. 5.10(a)-5.10(b), we find that the queue length of our model is maximum in peak hours and is minimum in the off-peak hours, which is reasonable. In our model, the mainstream of the freeway will be congested when traffic flow exceeds the capacity, and thus ramps will meter the inflows from on-ramps to the main road of freeways to mitigate traffic congestion on freeway mainstream. Less vehicles entering the main road may produce the queue vehicles waiting on on-ramps. This evolution continues until the ramp queue length reaches its maximum due to limited capacity space on on-ramps, and then keep maximum ramp queue length until the density on main road reduces. This trend will last around two hours, namely peak hours, which is consistent with practical situation.

Fig. 5.12 shows the VSL rate and the free flow speed for our model involving only VSL and our model involving both RM and VSL. By observing, we find that the models with only VSL and both RM and VSL show a similar trend, which is that VSL is implemented in peak periods. This is reasonable that the

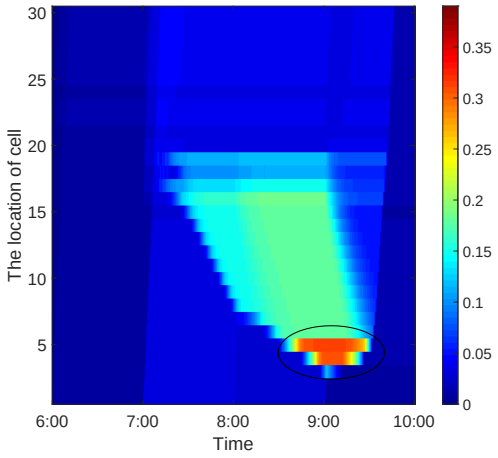


(a) Density (veh/m) for simulation

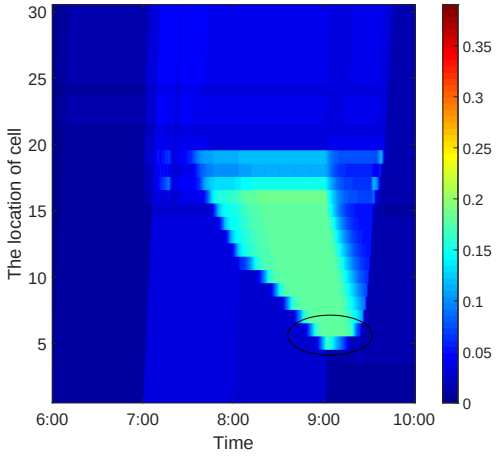
(b) Density (veh/m) for no control



(c) Density (veh/m) for RM

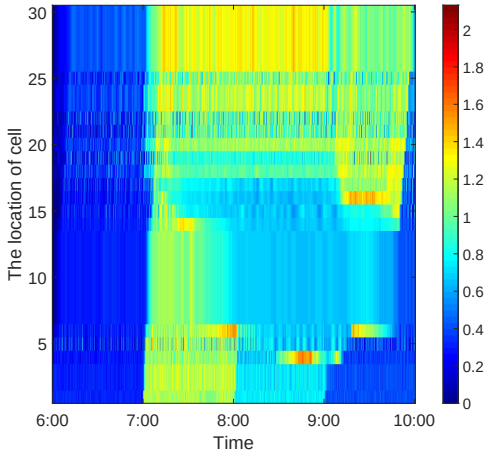


(d) Density (veh/m) for VSL

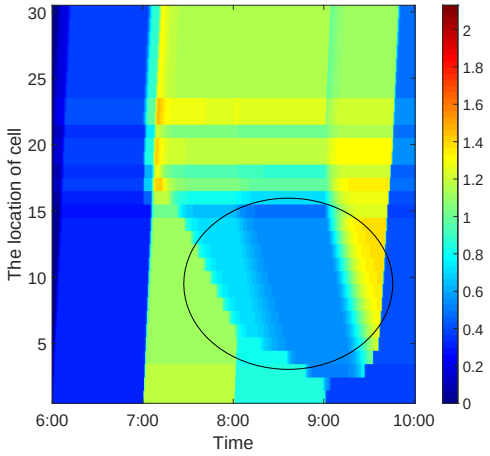


(e) Density (veh/m) for RM and VSL

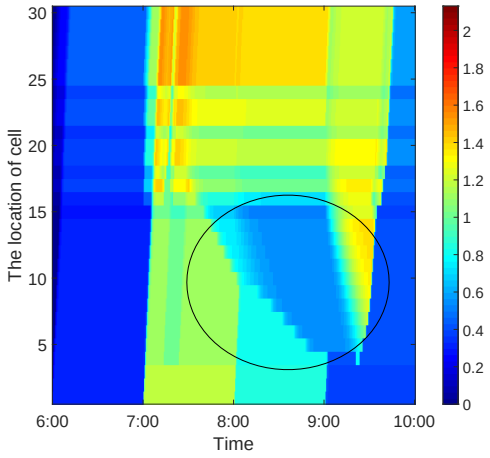
Figure 5.8: Traffic density for simulation, our model with no control, only RM, only VSL and both VSL and RM



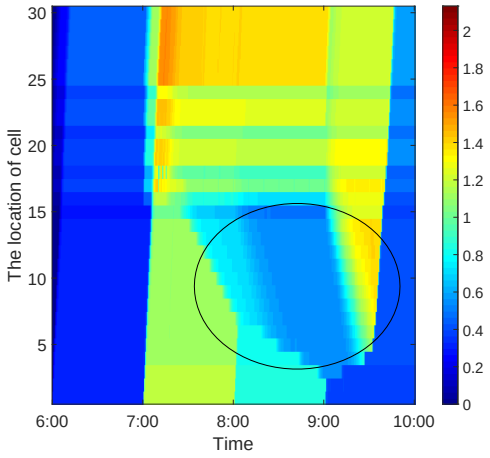
(a) Flow (veh/s) for simulation



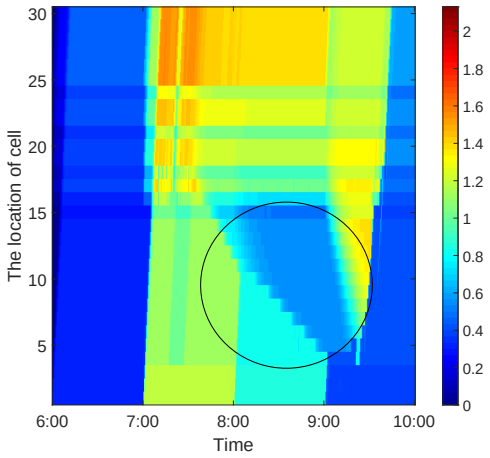
(b) Flow (veh/s) for no control



(c) Flow (veh/s) for RM

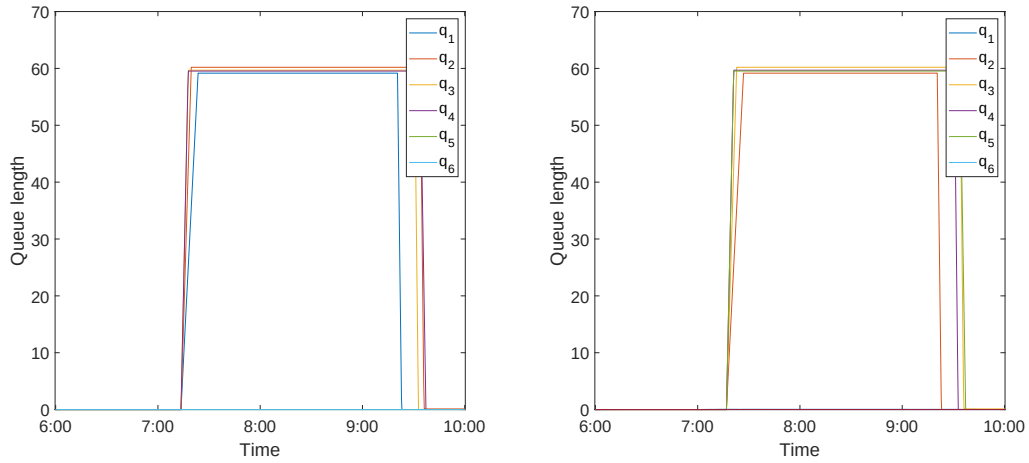


(d) Flow (veh/s) for VSL



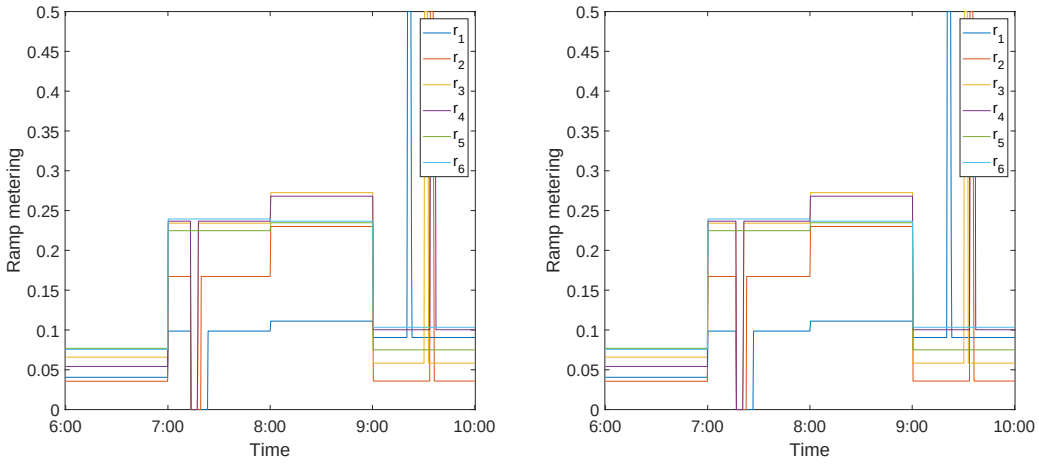
(e) Flow (veh/s) for RM and VSL

Figure 5.9: Flow for simulation, our model with no control, only RM, only VSL and both RM and VSL



(a) Queue length (veh) for RM (b) Queue length (veh) for RM and VSL

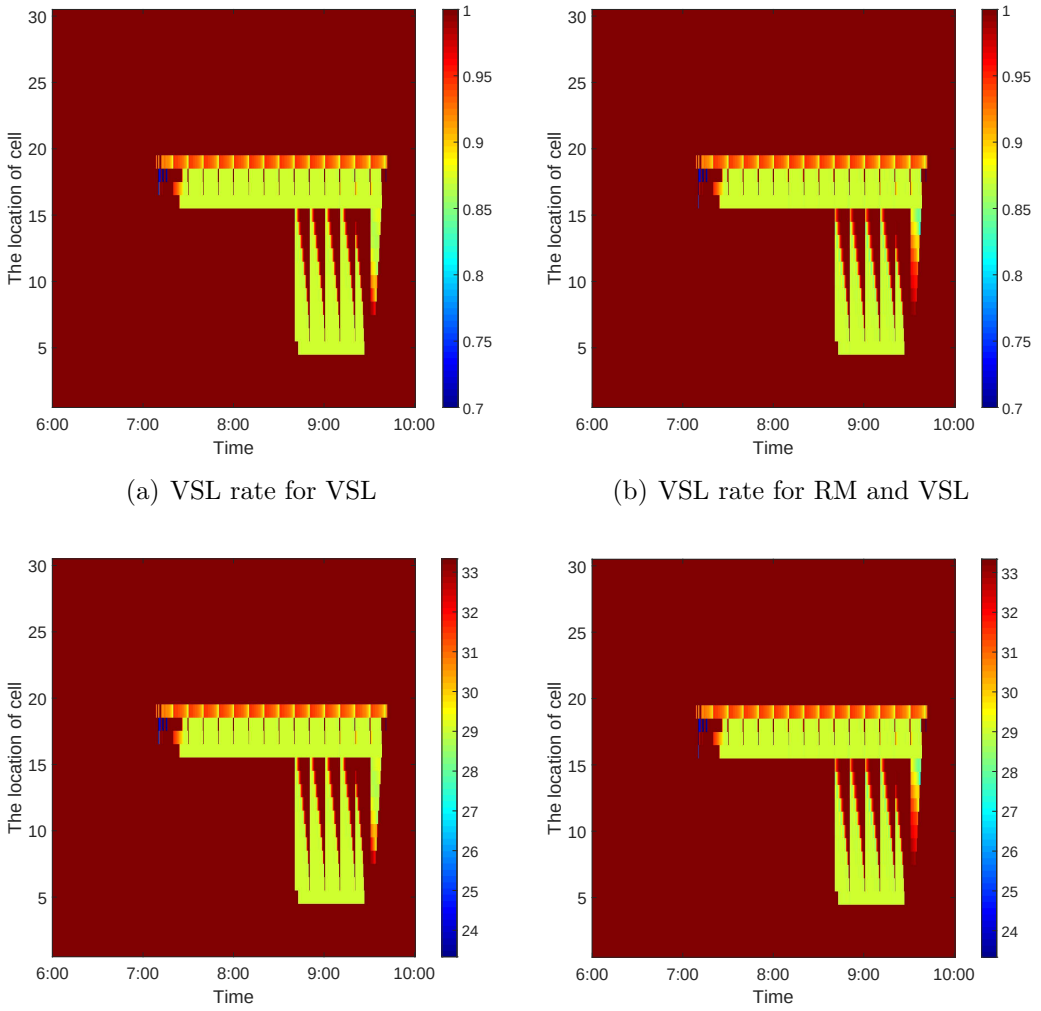
Figure 5.10: Queue length for our model with only RM, and both RM and VSL



(a) Ramp metering (veh/s) for RM (b) Ramp metering (veh/s) for RM and VSL

Figure 5.11: Ramp metering for our model with only RM, and both RM and VSL

out-going flow will increase by adjusting the VSL rate in congestion sate, but the maximum out-going flow does not be improved no matter any VSL rate is implemented. By comparison, it can be clearly seen that the highest efficiency is achieved by simultaneous adoption of both RM and VSL.



(a) VSL rate for VSL

(b) VSL rate for RM and VSL

(c) Free flow speed (m/s) for VSL

(d) Free flow speed (m/s) for RM and VSL

Figure 5.12: VSL rate and free flow speed for our model with only VSL, and both RM and VSL

Table 5.4 shows the results related to the total time spent and ramp delay. By comparison, collaborative implementation of our approximate model of RM and VSL is better than other models. Compared with results obtained from simulation with no control, our approximate model with both RM and VSL being implemented can reduce 11.564% and 53.21% of the total time spent and ramp delay, respectively. This indicates that our model is effect and useful, and RM and VSL are efficient control strategies for mitigating traffic congestion. Furthermore,

if we hold more vehicle on on-ramps, the congestion on main road can be improved but the total time spent is bigger compared with other models.

Table 5.4: Total time spent and ramp delay with Np = 36, Nc = 6

Name of model	Total time spent (h)	Ramp delay (h)	Percentage of reduction for total time spent	Percentage of reduction for ramp delay
Simulation	5295.806h	1109.444h	0%	0%
No control	5206.285h	0h	1.69%	100%
Only RM	4866.275h	532.361h	8.111%	52.016%
Only VSL	5019.833h	0h	5.211%	100%
VSL and RM	4683.424h	519.106h	11.564%	53.21%

5.6 Conclusion

In this work, we developed a optimal control model (DMP) by incorporating the effect of RM and VSL. The proposed model was developed on the basis of CTM with consideration of the capacity drop. To describe the influence of VSL, we proposed a linear model by adding an auxiliary variable which is estimated using obtained data. Due to the complexity and difficulty in computation of the 'min' operators, a smoothing method is used to obtain an approximate smooth discrete optimal control problem (DMP$^\epsilon$). Furthermore, the supporting theoretical analysis is carried out. Particularly, we have shown that the optimal solutions of the approximate problem (DMP$^\epsilon$) converge to that of Problem (DMP) as the parameter ϵ approaches to zero. A customised MPC method was designed to resolve Problem (DMP$^\epsilon$) to implement a feedback control law. To calibrate and validate the efficiency of the proposed model (DMP$^\epsilon$), two experimental applications are verified using the simulation data obtained in SUMO simulation environment.

CHAPTER 6

Deep Koopman Traffic Modeling for Freeway Ramp Metering

6.1 Introduction

In this Chapter, we design a data-driven modelling method with neural networks, denoted by deep Koopman model, to learn a finite-dimensional approximation of the Koopman operator. To consider the sequential relations of the ramps and main roads on the freeway, a long short-term memory network is applied. Furthermore, a model predictive controller with the trained deep Koopman model is proposed for the real-time control of the ramp metering on the freeway. To validate the performance of the proposed approach, experiments based on the simulation in the SUMO environment are conducted. The results validate the effectiveness of the developed approach on both the dynamics prediction and the real-time control of the ramp metering.

Traditionally, the control strategies of ramp metering control are designed as fixed or traffic responsive. The fixed control approach is easy to implement, but the main shortcoming is that the fixed parameters do not respond to the changing traffic states, which greatly reduces the performance of approach [179]. The traffic responsive strategies can be divided into as rule-based and model-based methods. The most widely used rule-based approaches, the ALINEA [142] and its variants [143,145,180], aim at regulating the occupancy approaching a predefined target value. However, the rule-based methods control the metering rate according to traffic conditions passively. Rule natures greatly limit the performance of approaches, especially in a rapidly changing traffic environment. Furthermore, the control parameters in the rule-based controllers depend on human prior knowledge [179]. On the other hand, the model-based approaches optimize the control strategies based on the conventional data collected by loop detectors [118] (such

as flow, density, speed). The MPC [162] framework considering the interaction between ramp control and future traffic conditions is often used to forecast the traffic evolution and realize the active traffic control [181]. However, the models used to describe traffic dynamics may result in nonlinear control problems, which generates difficulties in computing the optimal metering rate [174,182–184]. Some studies [73, 118, 120, 121] propose the CTM-based approaches, which simplify the macroscopic fundamental diagrams with a piecewise-point linear model. Because of reformulating as a linear framework, these approaches could obtain optimal solutions. However, such a simplification could result in the inaccurate modeling of non-linear dynamics. Therefore, it is still challenging to model the non-linear dynamics system in ramp metering problem.

With the development of big data technologies in recent years, real-time traffic data can be obtained through the sensors deployed on vehicles and road infrastructures. Compared with the conventional loop detector data, real-time sensory data has the ability to better describe the underlying traffic dynamics [185–187]. Machine learning techniques, for example, deep learning (DL) and reinforcement learning (RL), have been applied to learn the freeway traffic flow dynamics and ramp metering controlling strategies [16–18, 188]. However, the deep neural network-based traffic models in these studies suffer from a lack of interpretability [189]. Therefore, they might have unknown sensitive modes and be influenced by the model uncertainty. More importantly, due to the non-linearity of the activation functions in the deep neural networks, it is hard to integrate the learned dynamical models in the design of well-adjusted controllers, such as MPC and Linear Quadratic Regulator (LQR) [190]. Furthermore, deep neural networks are sometimes over-parameterized and computational time-consuming [191].

Recently, the Koopman operator theory, which provides a data-driven framework for learning non-linear dynamical systems with the theoretical guarantee, has drawn a lot of attentions from researchers in the field of control [189,192]. The Koopman operator is an invariant linear operator acting on the infinite dimensional lifted observable space of the state of the dynamical system [87, 104, 193]. The spectral properties of Koopman operator are helpful to describe the intrinsic properties of complex system dynamics by decomposition it into spatial-temporal coherent structures [104]. The Koopman operator could formulate a linear dynamical system instead of a black-box mapping of deep neural networks, which is more suitable to be used with controllers such as the linear MPC [189]. Therefore, the idea of the Koopman operator is theoretically widely applicable and useful for modelling dynamical systems that can not be accurately approximated by local

linearized models [104].

Various approaches have been designed to approximate the properties of the Koopman operator including the Koopman eigenvalues, eigenfunctions and modes. Three main classic algorithms that approximate these properties with finite-dimensional matrix have been widely studied: Ulam Galerkin Method [194, 195], Generalized Laplace Analysis (GLA) [102, 196, 197], and Dynamic mode Decomposition (DMD) [100, 101, 103]. All these algorithms are data-driven and require no explicit formulas, while none of them can approximate all properties of the Koopman operator. For example, the Ulam Galerkin Method can be applied to approximate directly both the Koopman eigenfunctions and eigenvalues [195], while DMD is used to approximate the Koopman modes and eigenvalues [100,103]. Despite this disadvantage, they have been successfully applied for the analysis of non-linear dynamics, such as fluid flows [101, 198], buildings and power systems [92, 95], and traffic management [21, 199]. Based on the DMD method, an extended dynamic mode decomposition algorithm (EDMD) [104] has been proposed to approximate all three properties of Koopman operator with a dataset of snapshot pairs and a dictionary of observable functions. In EDMD, the observable functions that map the system state into a lifted space are normally manually selected. In studies [104, 200], the standard procedure of function selection is to choose a commonly used observable function, such as the radial basis function or the polynomial function. However, such approach does not take any empirical information or data of dynamics into consideration.

Deep neural networks have been demonstrated as an alternative approximator of the Koopman operator, which directly learn the observable functions [189]. Inspired by these works, we propose a Deep Koopman method, a variant of the EDMD, which uses neural networks to approximate the Koopman operator with finite-dimensional feature space. This is different from the aforementioned traffic models [16–18, 188]. More specifically, an encoder-decoder network is designed to learn the representation of the system state in the lifted space as well as the mapping process of converting features from the lifted space back to its original space. The LSTM unit is used to process the sequential data of the traffic flow on the freeway. To realize a reliable real-time controlling of the ramp metering on the freeway, we design a linear MPC controller to produce an online optimal control with the objective of maximizing the total throughput of leaving vehicles within a time frame. A few studies of combining the Koopman operator and the MPC have been conducted [189, 206]. In this work, the simulation study is conducted in SUMO environment due to its real-time visualization, the inclusion

of various vehicle behaviour models, and the supportive built-on computational frameworks for DL. The applications of SUMO can be referred to [201–203].

The chapter is organized as follows: Section 6.2 shows the modelling of dynamics of freeway traffic flows and the ramp metering control problem. Section 6.3 develops the Deep Koopman model for the freeway traffic flow dynamics and Section 6.4 describes the MPC controller with the dynamical system approximated by the Deep Koopman model. Numerical studies and conclusions are given in Section 6.5 and Section 6.6, respectively.

6.2 The freeway traffic model and ramp metering controls

We consider a multi-lane freeway road consisting of M main road segments $\mathcal{M} = \{m_1 \ ... \ m_M\}$, N on-ramps $\mathcal{N} = \{n_1 \ ... \ n_N\}$, and L off-ramps $\mathcal{L} = \{l_1 \ ... \ l_L\}$. Fig. 6.1(a) illustrates the freeway network, where each on-ramp has a ramp metering controller adjusting the traffic flow merging into the main road. At each time step $t \in \{0 \ 1 \ ... \ T\}$, the state $x_t \in \mathcal{X} \subseteq \mathbb{R}^n$ of the freeway traffic is defined by the vehicle out-flow through each off-ramp $o_{t,i}, i \in \mathcal{L}$ and the main road $o_{t,i}, i = m_M$, the vehicle queue length at each on-ramp $q_{t,i}, i \in \mathcal{N}$, and the number of vehicles on each main road segment $q_{t,i}, i \in \mathcal{M}$, where $n = M + N + L + 1$. The sequence of state features should be consistent with the flow direction of the main road segments and ramps, such as Fig. 6.1(b). We define the sequence of main road segments and on-ramps as $\mathcal{J} = \{j_1 \ ... \ j_J\}$, where $j_j \in \mathcal{M} \cup \mathcal{N}$. The decision variables of the ramp metering control are defined as the proportion of green light duration over the decision interval t for each on-ramp, denoted by $a_{t,i}, i \in \mathcal{N}$. Except for the features in the traffic state x_t, the estimated in-flow demand of vehicles for each on-ramp would also affect the control variables $a_{t,i}$. Therefore, we also include the demand of each on-ramp, $d_{t,i}, i \in \mathcal{N}$ and in-flow of main road $d_{t,i}, i = m_1$ as a fixed variable in the system control input $u_t \in \mathcal{U} \subseteq \mathbb{R}^m$, where $m = 2N + 1$. In this section, we describe the dynamics system of the freeway traffic as a discrete-time control model:

$$\begin{aligned} x_{t+1} &= f(x_t, u_t), \\ x_t &= [o_{t,l_1}, ..., o_{t,l_L}, o_{t,m_M}, q_{t,j_1}, ..., q_{t,j_J}]^\top, \\ u_t &= [a_{t,n_1}, ..., a_{t,n_N}, d_{t,n_1}, ..., d_{t,n_N}, d_{t,m_1}]^\top, \end{aligned} \tag{6.1}$$

Based on the nonlinear system (6.1), we define an optimal control problem as

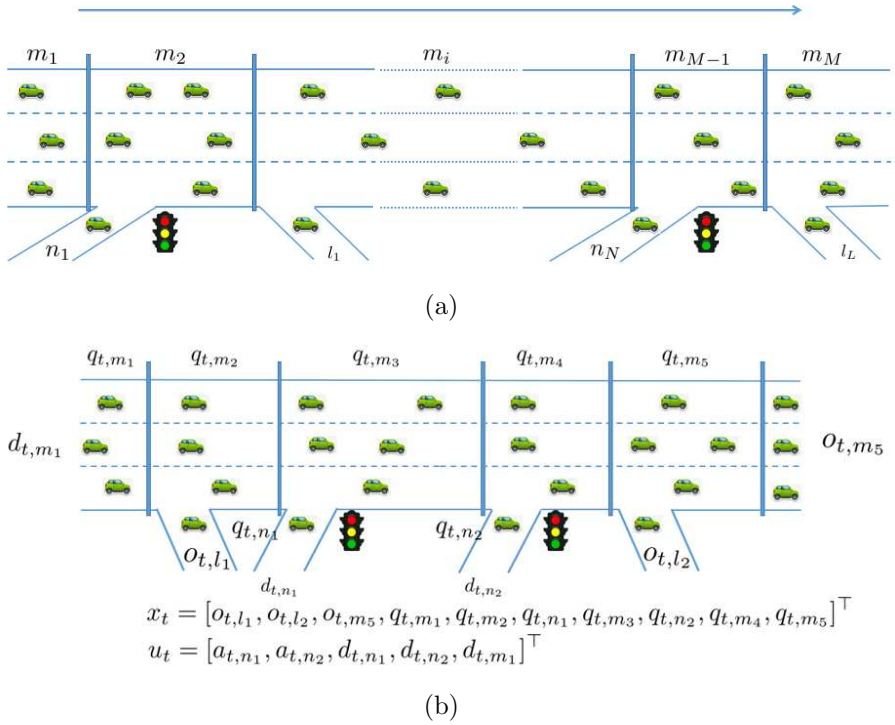


Figure 6.1: Schematic diagrams: (a) ramp metering for a freeway, (b) an example of describing the state x_t and control u_t with ramp metering

follows:

$$\max_{u_t, t=0,\ldots,T-1} \mathrm{T} = \sum_{t=0}^{T} v_t^\top x_t - \lambda \sum_{t=0}^{T-2} ||u_{t+1} - u_t||^2, \tag{6.2}$$

$$\text{s.t.}$$

$$x_{t+1} = f(x_t, u_t), t = 0, \ldots, T-1, \tag{6.3}$$

$$x_{min} \leq x_t \leq x_{max}, t = 0, \ldots, T, \tag{6.4}$$

$$u_{min} \leq u_t \leq u_{max}, t = 0, \ldots, T-1, \tag{6.5}$$

where the first linear part in (6.2) denotes the throughput of the freeway road including vehicles leaving through the main road and the off-ramps, where $v_t \in \mathbb{R}^n$. Let $v_{t,i} = 1, i = 1, \ldots, N+1$, others are 0. In our design of the ramp metering control, we anticipate the control variables to be as stable as possible. Therefore, the second term in (6.2) is a penalty term that penalizes the variation of control variables within two consecutive time steps, and the coefficient λ is a pre-defined hyperparameter adjusting the weight of penalty. (6.4) and (6.5) define the upper and lower bounds of state variables x and control inputs u, respectively.

One of the main challenges of resolving the problem (6.2) - (6.5) is the modelling of the non-linear dynamics system presented in (6.3). Thus, we introduce

a Deep Koopman model to learn a linear approximation with convergence guarantees of the non-linear dynamics system of the freeway traffic based on the Koopman operator.

6.3 Deep Koopman Model for the freeway traffic flow dynamics

In this section, we first introduce the forced system described using the Koopman operator. Then, we outline the extended dynamic mode decomposition (EDMD) method developed based on the Koopman operator theory. Finally, a data-driven Deep Koopman model is presented to construct an approximation of the appropriate observable functions.

6.3.1 The Koopman Operator with forced system

The Koopman operator is originally developed to describe the inherent properties of an uncontrolled non-linear dynamical system through a linear dynamical evolution. With a slight variation, we can employ the Koopman operator for the controlled dynamical system. In this work, we take a non-linear dynamical system for the freeway traffic flow into account:

$$x_{t+1} = f(x_t, u_t). \tag{6.6}$$

Equation (6.6) is a discrete-time freeway traffic flow system that evolves in the light of an unknown nonlinear law, where $x_t \in \mathcal{X}$ defined as Equation (6.1) is the state variable at time step t; $u_t \in \mathcal{U}$ defined as Equation (6.1) is the control input variable at time step t. Let $\mathbf{u}_t = (u_l)_{l=t}^{\infty} \in \overline{\mathcal{U}}$ and $\mathcal{G}\mathbf{u}_t = \mathbf{u}_{t+1}$ with $\mathcal{G}$ being the left shift operator, where $\overline{\mathcal{U}}$ denotes the space of all sequences $(u_l)_{l=0}^{\infty}$ with $u_l \in \mathcal{U}$. Then, we define the Koopman operator of the dynamics on the extended state $[x, \mathbf{u}]$ as follows:

$$\mathcal{K}g(x_t, \mathbf{u}_t) = g(x_{t+1}, \mathbf{u}_{t+1}) = g(f(x_t, u_t), \mathcal{G}\mathbf{u}_t), \tag{6.7}$$

where $\mathcal{K} : \mathcal{H} \to \mathcal{H}$ refers to the Koopman operator, $g : \mathcal{X} \times \overline{\mathcal{U}} \to \mathbb{R}^{\mathrm{N}}$, $g \in \mathcal{H}$ is an observable function, $\mathcal{H}$ is a Banach space. The Koopman operator $\mathcal{K}$ is a composition map of g and f. Note that, $\mathcal{K}$ is infinite-dimensional and linear even though the dynamics $f(\cdot)$ is non-linear. Due to the linear characteristics of $\mathcal{K}$,

we can have the eigendecomposition of $\mathcal{K}$ in the standard form:

$$\mathcal{K}\varphi_j(x_t, \mathbf{u}_t) = \lambda_j \varphi_j(x_t, \mathbf{u}_t), j = 1, 2, \dots \tag{6.8}$$

where $\varphi_j(x_t, \mathbf{u}_t)$ and λ_j are the eigenfunction and the associated eigenvalue of $\mathcal{K}$, respectively. Obviously, the operator is spanned by eigenfunctions that are governed by the state and the control input. A vector-valued observable g can be rewritten in terms of these eigenfunctions φ_j as:

$$g(x_t, \mathbf{u}_t) = (\mathcal{K}^t g)(x_0, \mathbf{u}_0) = \sum_{j=1}^{\infty} \lambda_j^t \varphi_j(x_0, \mathbf{u}_0)\nu_j, \tag{6.9}$$

where $\mathcal{K}^t$ denotes t times Koopman operator and ν_j denotes the Koopman mode [100] corresponding to the eigenvalue λ_j and x_0 is the initial state. It is noted that not all vector-valued observables can be formulated as Equation (6.9). For some systematic observables, additional parts may be required to explain the rest of the spectrum of Koopman operators, as shown in [204, 205].

6.3.2 Extended Dynamic Mode Decomposition

The extended dynamic mode decomposition (EDMD) [104] is an approach of finding the finite-dimensional approximation of the Koopman operator. The EDMD uses a variety of basis functions, such as radial basis functions (RBF) with different kernel centers and widths, to represent the observable functions and the least square method to calculate a finite-dimensional approximation $\mathbf{K} \in \mathbb{R}^{\mathrm{N}\times\mathrm{N}}$ of $\mathcal{K}$. Note it is impossible if we use the least square method to compute an approximation of the Koopman operator in a finite time because $[x, \mathbf{u}]$ is in general infinite-dimensional. To obtain the finite-dimensional approximation $\mathbf{K}$, we choose the lifting function g_i in a special way for any $i = \{1, 2, \dots, \mathrm{N}\}$, which is written of the form

$$g_i(x_t, \mathbf{u}_t) = \psi_i(x_t) + h_i(\mathbf{u}_t), \tag{6.10}$$

where $\psi_i : \mathcal{X} \to \mathbb{R}$ is in general nonlinear but $h_i : \overline{\mathcal{U}} \to \mathbb{R}$ is linear. Without loss of generality, we can assume that the vector-valued observable function $g = [g_1 \ \dots \ g_{\mathrm{N}}]^\top$ is of the form

$$g(x_t, \mathbf{u}_t) = [\psi^\top(x_t) \ \widetilde{u}_t^\top]^\top, \tag{6.11}$$

where $\psi(x_t) = [\psi_1(x_t)\ \psi_2(x_t)\ \dots\ \psi_l(x_t)]^\top$ with $l = \mathrm{N} - m$, and $\widetilde{u}_t \in \mathbb{R}^m$ is the first component of the sequence $\mathbf{u}_t$, In order not to abuse the notation, we use u_t to denote $\widetilde{u}_t$ throughout the Chapter. Based on Equations (6.7) and (6.11), it gives rise

$$\mathcal{K}g(x_t, \mathbf{u}_t) = \mathbf{K}[\psi^\top(x_t)\ u_t^\top]^\top + r(x_t, u_t), \tag{6.12}$$

where the function $r(x_t, u_t)$ is a residual term that describes the gap between the N−dimensional approximation of the observable space and the actual lifted space of the Koopman operator. We optimize the cost function $||r(x_t, u_t)||_2^2$ to determine $\mathbf{K}$. Because we do not care about the predictive future value of control sequence, we can disregard the last m components of $[\psi^\top(x_{t+1})\ u_{t+1}^\top]^\top$. Let $\Psi(x_t) = [\psi^\top(x_t)\ u_t^\top]^\top$, we write the observables dynamics with the Koopman operator as:

$$\begin{aligned} \psi(x_{t+1}) &= \mathrm{A}\Psi(x_t), \\ \widetilde{x}_t &= \mathcal{C}\psi(x_t), \end{aligned} \tag{6.13}$$

where $\mathrm{A} = [\mathcal{A}\ \mathcal{B}] \in \mathbb{R}^{l\times \mathrm{N}}$, $\mathcal{A} \in \mathbb{R}^{l\times l}$, $\mathcal{B} \in \mathbb{R}^{l\times m}$. A is the former l rows of $\mathbf{K}$, and $\mathcal{C} \in \mathbb{R}^{n\times l}$ is the mapping from the lifted space to the original space, and $\widetilde{x}_t \in \mathbb{R}^n$ is the reconstruction state of x_t. Hence, we can obtain the solution of $\mathcal{A}, \mathcal{B}$ by optimizing the cost function

$$\min_{\mathcal{A},\mathcal{B}} ||r||_2^2 = \min_{\mathcal{A},\mathcal{B}} \sum_{t=0}^{T} ||\psi(x_{t+1}) - \mathcal{A}\psi(x_t) - \ \mathcal{B}u_t||_2^2. \tag{6.14}$$

Furthermore, its solution can be obtained as:

$$[\mathcal{A}\ \mathcal{B}] = \mathcal{S}\mathcal{V}^\top(\mathcal{V}\mathcal{V}^\top)^\dagger, \tag{6.15}$$

where $\mathcal{S} = [\psi(x_1)\ \psi(x_2)\ \dots\ \psi(x_{T+1})]$, $\mathcal{V} = [\Psi(x_0)\ \Psi(x_1)\ \dots\ \Psi(x_T)]$, and † is the Moore Penrose pseudoinverse. The matrix $\mathcal{C}$ can be computed by minimizing the following problem:

$$\min_{\mathcal{C}} \sum_{t=0}^{T} ||x_t - \mathcal{C}\psi(x_t)||_2^2, \tag{6.16}$$

with the solution:

$$\mathcal{C} = \mathcal{R}\mathfrak{W}^{\top}(\mathfrak{W}\mathfrak{W}^{\top})^{\dagger}, \tag{6.17}$$

where $\mathcal{R} = [x_0\ x_1\ \ldots\ x_T]$, and $\mathfrak{W} = [\psi(x_0)\ \psi(x_1)\ \ldots\ \psi(x_T)]$. In particular, if the set of lifting functions $\{\psi_1\ \ldots\ \psi_l\}$ includes the state variable x, i.e., after possible reordering, $\psi_i(x) = x_i$ for any $i \in \{1, \ldots, n\}$. In this case, we have that $\mathcal{C} = [I\ \mathbf{0}]$, where I denotes the identity matrix of size n. More details regarding to the calculation of $\mathrm{A}, \mathcal{C}$ can be referred to literatures [104, 206].

6.3.3 The Deep Koopman Method

Compared to the classic EDMD, the Deep Koopman model uses a deep neural network to generate the observable subspace of the Koopman operator. The proposed Deep Koopman model adopts an encoder-decoder structure, where the encoder ψ^e transforms the system inputs to a lifted space and the decoder ψ^d converts the lifted system features back to the original space. We can represent the dynamical system as:

$$\begin{aligned} \psi^e(x_{t+1}, w^e) &= \widetilde{\mathrm{A}}\Phi(x_t, w^e, u_t), \\ \widetilde{x}_t &= \psi^d(\psi^e(x_t, w^e), w^d), \end{aligned} \tag{6.18}$$

where $\widetilde{\mathrm{A}} = [\widetilde{\mathcal{A}}\ \widetilde{\mathcal{B}}] \in \mathbb{R}^{l\times \mathrm{N}}$, $\Phi(x_t, w^e, u_t) = [\psi^e(x_t, w^e)^{\top}\ u_t^{\top}]^{\top} \in \mathbb{R}^{\mathrm{N}}$, ψ^e is the encoder with parameter w^e, and ψ^d is the decoder with parameter w^d. Fig. 6.2 shows the detail process of the Deep Koopman model.

For the ramp metering control problem on the freeway, there exists spatial dependencies for elements of state x. In this case, a LSTM network is used to process the sequential data in the encoder, i.e., $\psi^e(x, w^e)$, which lifts the primal state x to the higher-dimension vector-valued observables. Fig. 6.3 shows the structure of the LSTM. The LSTM is an architecture of recurrent neural network (RNN) consisting of feedback connections. A LSTM has a forget gate $\mathfrak{f}_k$, an input gate $\mathfrak{i}_k$ and an output gate $\mathfrak{o}_k$. The forget gate utilizes the output $\mathfrak{h}_{k-1}$ in the last step and the current state x_k to compute the part of the cell state $\mathfrak{c}_{k-1}$ retaining in the current evaluation. Similarly, the input gate determines the update of the cell state. Finally, the output gate uses the newly updated cell state to determine the output. Algorithm 3 shows the process of LSTM, where $\otimes$ stands for the Hadamard product and σ represents the logistic sigmoid function. $\mathcal{W}$ and $\mathfrak{b}$ denote the weights and biases.

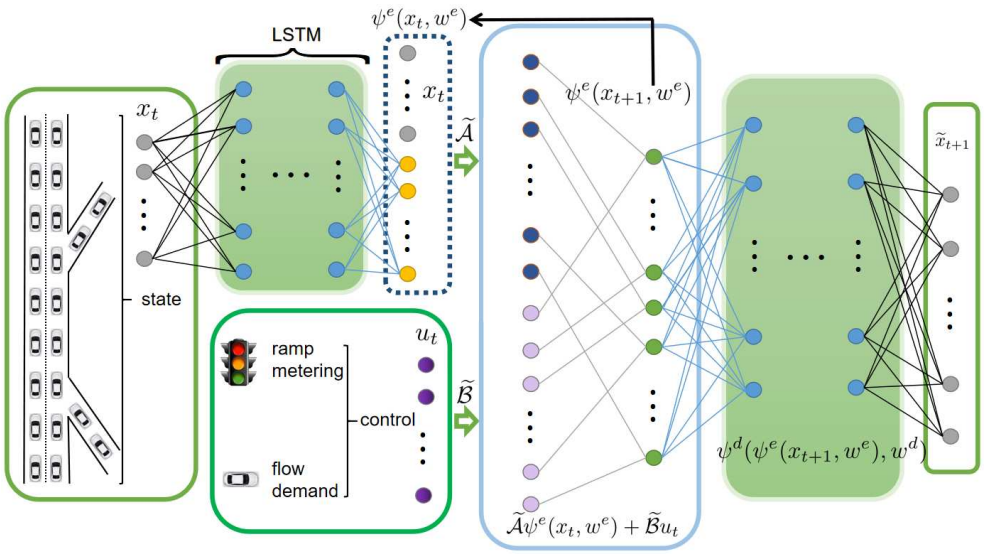


Figure 6.2: The diagram of the proposed DKM. The original state x_t is lifted with the encoder, i.e.,$\psi^e(x_t, w^e)$. Then $\psi^e(x_t, w^e), u_t$ form the lifted state for constructing the linear evolution in the vector-valued observables. The freeway traffic flow dynamics can be recovered via a decoder $\psi^d(\psi^e(x_{t+1}, w^e), w^d)$ from the vector-valued observables

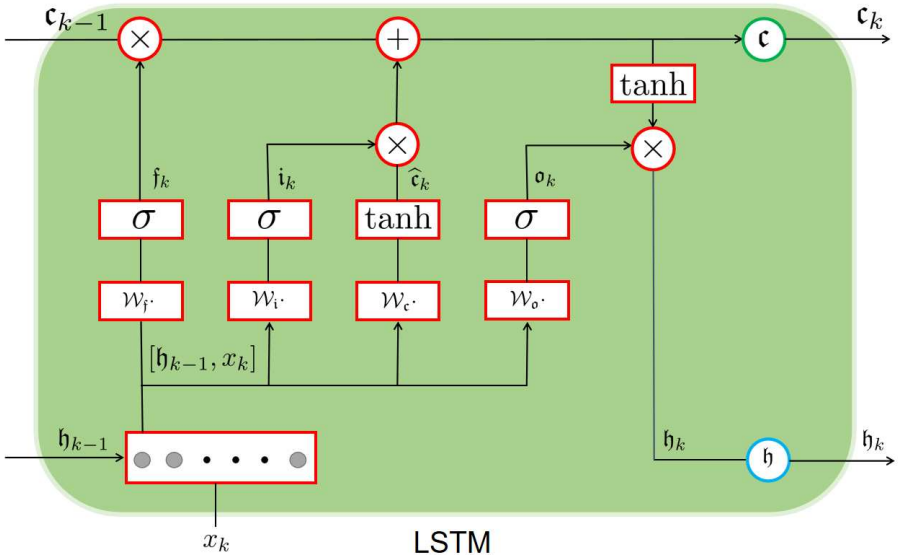


Figure 6.3: The diagram of the LSTM

Algorithm 3 : Compute the output sequence of a LSTM network

Input: Sequence x_1, x_2,..., x_K,
Output: Sequence $\mathfrak{h}_1$, $\mathfrak{h}_2$,..., $\mathfrak{h}_K$,
Set $\mathfrak{h}_0 = 0$
Set $\mathfrak{c}_0 = 0$
for $k \leftarrow k$ to K do **do**
    $\mathfrak{f}_k = \sigma(\mathcal{W}_\mathfrak{f}[\mathfrak{h}_{k-1}\ x_k] + \mathfrak{b}_\mathfrak{f})$
    $\mathfrak{i}_k = \sigma(\mathcal{W}_\mathfrak{i}[\mathfrak{h}_{k-1}\ x_k] + \mathfrak{b}_\mathfrak{i})$
    $\widehat{\mathfrak{c}}_k = \tanh(\mathcal{W}_\mathfrak{c}[\mathfrak{h}_{k-1}\ x_k] + \mathfrak{b}_\mathfrak{c})$
    $\mathfrak{c}_k = \mathfrak{f}_k \otimes \mathfrak{c}_{k-1} + \mathfrak{i}_k \otimes \widehat{\mathfrak{c}}_k$
    $\mathfrak{o}_k = \sigma(\mathcal{W}_\mathfrak{o}[\mathfrak{h}_{k-1}\ x_k] + \mathfrak{b}_\mathfrak{o})$
    $\mathfrak{h}_k = \mathfrak{o}_k \otimes \tanh(\mathfrak{c}_{k-1})$
end for

Assuming K is the sequence length of the state inputs in the encoder, the vector-valued observables can be estimated by the last layer hidden state:

$$\psi^e(x, w^e) = [x^\top\ \mathfrak{h}_K^\top]^\top. \tag{6.19}$$

We utlize a feed-forward neural network to construct the decoder, i.e., ψ^d, which includes K' fully connected layers to recover the original state from the higher-dimensional observables. In particular, the output at any hidden layer $k \in \{1, ..., K'\}$ can be written as

$$\mathfrak{h}_k^d = \sigma_k^d(\mathcal{W}_k^d \mathfrak{h}_{k-1}^d + \mathfrak{b}_k^d). \tag{6.20}$$

The predicted state can be estimated by the output of the last layer of the decoder:

$$\widetilde{x}_t = \psi^d(\psi^e(x_t, w^e), w^d) = \sigma(\mathcal{W}_{K'}^d \mathfrak{h}_{K'-1}^d + \mathfrak{b}_{K'}^d). \tag{6.21}$$

Based on Equations (6.19) and (6.21), we can have the expressions of the encoder and decoder for dynamics system in (6.18). Moreover, parameters $\mathcal{W}$ and $\mathfrak{b}$ can be determined by optimizing a fixed loss function during training.

To approximate the freeway traffic flow dynamics in a longer time window, a multi-step prediction error rather than the single-step prediction error is considered. To formulate the multi-step loss function, we first present the state prediction in an M time-step interval:

$$x_{t+M} = \psi^d(\widetilde{\mathrm{A}}^M \Phi(x_t, w^e, u_t), w^d) + r(x, M), \tag{6.22}$$

where $\widetilde{\mathrm{A}}^M\Phi(x_t, w^e, u_t)$ denotes the M-step ahead state beginning from x_t with the expression of:

$$\begin{aligned}\widetilde{\mathrm{A}}^M\Phi(x_t, w^e, u_t) &= \psi^e(x_{t+M}, w^e) \\ &= \widetilde{\mathcal{A}}\psi^e(x_{t+M-1}, w^e) + \widetilde{\mathcal{B}}u_{t+M-1} \\ &= \widetilde{\mathcal{A}}^M\psi^e(x_t, w^e) + \sum_{s=1}^{M}\widetilde{\mathcal{A}}^{s-1}\widetilde{\mathcal{B}}u_{t+M-s}.\end{aligned} \tag{6.23}$$

Now, we define the main components of the loss function for the M time steps:

$$\mathfrak{L}_1 = \frac{1}{M}\sum_{s=1}^{M}||x_{t+s} - \psi^d(\widetilde{\mathrm{A}}^s\Phi(x_t, w^e, u_t), w^d)||_2^2, \tag{6.24}$$

$$\mathfrak{L}_2 = \frac{1}{M}\sum_{s=1}^{M}||\psi^e(x_{t+s}, w^e) - \widetilde{\mathrm{A}}^s\Phi(x_t, w^e, u_t)||_2^2, \tag{6.25}$$

$$\mathfrak{L}_3 = \frac{1}{M}\sum_{s=1}^{M}||x_s - \psi^d(\psi^e(x_s, w^e), w^d)||_2^2, \tag{6.26}$$

where Equation (6.24) describes the prediction error, Equation (6.25) describes the predicted error in the lifted vector-valued observable space, and Equation (6.26) shows the reconstruction error. Based on the definitions of above loss functions, we can write the optimization problem as follows:

$$\begin{aligned}\min_{w^e, w^d, \widetilde{\mathcal{A}}, \widetilde{\mathcal{B}}}\mathfrak{L} &= \alpha_1\mathfrak{L}_1 + \alpha_2\mathfrak{L}_2 + \alpha_3\mathfrak{L}_3 + \alpha_4||w^e||_2^2 \\ &\quad +\alpha_5||w^d||_2^2,\end{aligned} \tag{6.27}$$

where $\alpha_i, i = 1, 2, \ldots, 5$, are the weight parameters, and the last two terms of $\mathfrak{L}$ are regularization terms for preventing from over-fitting. Algorithm 4 shows the training process of the Deep Koopman model.

With the Deep Koopman model trained in Algorithm 4, and the approximated dynamics for freeway traffic flow can be written as:

$$\psi^e(x_{t+1}, w^e) = \widetilde{\mathcal{A}}\psi^e(x_t, w^e) + \widetilde{\mathcal{B}}u_t, \tag{6.28}$$

$$\widetilde{x}_t = \psi^d(\psi^e(x_t, w^e), w^d). \tag{6.29}$$

Algorithm 4 The Deep Koopman Method (DKM)

Initialization: w^e, w^d, $\widetilde{\mathcal{A}}$, $\widetilde{\mathcal{B}}$, M. $Epoch = 0$, $Epoch_{max}$, $\alpha_i, i = 1, \ldots, 5$, batch size b_s, a small constant $\epsilon > 0$;
Train: trained w^e, w^d, $\widetilde{\mathcal{A}}$ and $\widetilde{\mathcal{B}}$;
while $Epoch > Epoch_{max}$ or $|\mathfrak{L}| \leq \epsilon$ **do**
    Reset the training episodes;
    while *Epoch* is not Terminated **do**
        Sample a batch data sequence of state x and control u;
        Compute the vector-valued observables $\psi^e(x, w^e)$ with (6.19) and reconstruction states $\widetilde{x} = \psi^d(\psi^e(x, w^e), w^d)$ with (6.21);
        Calculate the multi-step vector-valued observables $\widetilde{\mathrm{A}}^s\Phi(x_0, w^e, u_0)$ with (6.23) and predicted states $\widetilde{x}_s = \psi^d(\widetilde{\mathrm{A}}^s\Phi(x_0, w^e, u_0), w^d)$, where $s = 1, 2, \ldots, M$;
        Compute the loss function $\mathfrak{L}$ with (6.27);
        Update w^e, w^d, $\widetilde{\mathcal{A}}$ and $\widetilde{\mathcal{B}}$ via solving problem (6.27);
    end while
    $Epoch = Epoch + 1$
end while

6.4 MPC-Deep-Koopman for Freeway Traffic Flow Systems with Ramp Metering

The main idea of MPC controller aims to optimize a specified optimization problem of the control input at each time step of the closed-loop operator. In this section, we design a MPC controller based on the linear system approximated by the Deep Koopman model (6.28) to solve the non-linear system in (6.6). Its computational complexity is comparable to that of a MPC controller for a linear system with original system state space. We formulate the optimization problem

at each time step t of the closed loop as:

$$\begin{aligned} \max_{u_{t,p}} \mathrm{T}_t &= \sum_{p=0}^{\mathrm{Np}} \widetilde{v}_p^\top z_{t,p} - \lambda \sum_{p=0}^{\mathrm{Np}-2} ||u_{t,p+1} - u_{t,p}||^2, \end{aligned} \tag{6.30}$$

$$\begin{aligned} \text{s.t.} & \\ z_{p+1} &= \widetilde{\mathcal{A}} z_{t,p} + \widetilde{\mathcal{B}} u_{t,p}, p = 0, \ldots, \mathrm{Np} - 1, \end{aligned} \tag{6.31}$$

$$b_{min,p} \leq E_p z_{t,p} \leq b_{max,p}, p = 0, \ldots, \mathrm{Np}, \tag{6.32}$$

$$\begin{aligned} u_{min,p} &\leq u_{t,p} \leq u_{max,p}, p = 0, \ldots, \mathrm{Np} - 1, \end{aligned} \tag{6.33}$$

$$z_{t,0} = \psi^e(x_t, w^e), \tag{6.34}$$

where Np is the predictive horizon, $\widetilde{v}_p \in \mathbb{R}^l$, $E_p \in \mathbb{R}^{d \times l}$, and $u_{min,p}$ and $u_{max,p}$ denotes the lower and upper bounds of control input $u_{t,p}$, respectively, and $z_{t,0}$ is the lifted state obtained using the Deep Koopman method at time step t. Note that the optimization problem (6.30) is parametrized by the current state x_t of the non-linear system (6.34). We can obtain a feedback controller by solving the problem T_t. We denote u_t^* as the optimal solution and reformulate the problem T_t in a so-called dense form:

$$\begin{aligned} \max_{u_t} \mathrm{T}_t &= u_t^\top R u_t + h^\top u_t + c^\top z_{t,0}, \\ \text{s.t.} & \\ b_{min} &\leq E z_{t,0} + D u_t \leq b_{max}, \\ u_{min} &\leq u_t \leq u_{max}, \\ z_{t,0} &= \psi^e(x_t, w^e), \end{aligned} \tag{6.35}$$

where $u_t = [u_{t,0}^\top, \ldots, u_{t,\mathrm{Np}-1}^\top]^\top \in \mathbb{R}^{m\mathrm{Np}}$; $R \in \mathbb{R}^{m\mathrm{Np} \times m\mathrm{Np}}$; $h \in \mathbb{R}^{m\mathrm{Np}}$; $c = \sum_{p=0}^{\mathrm{Np}} \widetilde{v}_p^\top \widetilde{\mathcal{A}}^p$ $\in \mathbb{R}^l$; $E \in \mathbb{R}^{d(\mathrm{Np}+1) \times l}$; $D \in \mathbb{R}^{d(\mathrm{Np}+1) \times m\mathrm{Np}}$; $b_{min} = [b_{min,0}^\top \ \ldots \ b_{min,\mathrm{Np}}^\top]^\top \in \mathbb{R}^{d(\mathrm{Np}+1)}$; $b_{max} = [b_{max,0}^\top \ \ldots \ b_{max,\mathrm{Np}}^\top]^\top \in \mathbb{R}^{d(\mathrm{Np}+1)}$; $u_{min} = [u_{min,0}^\top \ \ldots \ u_{min,\mathrm{Np}-1}^\top]^\top \in \mathbb{R}^{m\mathrm{Np}}$;

and $u_{max} = [u_{max,0}^\top \ \dots \ u_{max,\mathrm{Np}-1}^\top]^\top \in \mathbb{R}^{m\mathrm{Np}}$. In particular,

$$R = \begin{bmatrix} -\lambda * e & \lambda * e & \dots & \mathbf{0} & \mathbf{0} \\ \lambda * e & -2\lambda * e & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \dots & -2\lambda * e & \lambda * e \\ \mathbf{0} & \mathbf{0} & \dots & \lambda * e & -\lambda * e \end{bmatrix},$$

where $e = \mathbf{I} \in \mathbb{R}^{m\times m}$ is a unit matrix,

$$h = \left[\sum_{p=1}^{\mathrm{Np}} \widetilde{v}_p^\top \widetilde{\mathcal{A}}^{p-1}\widetilde{\mathcal{B}} \ \sum_{p=2}^{\mathrm{Np}} \widetilde{v}_p^\top \widetilde{\mathcal{A}}^{p-2}\widetilde{\mathcal{B}} \ \dots \ \widetilde{v}_{\mathrm{Np}}\widetilde{\mathcal{B}}\right]^\top,$$

$$E = \begin{bmatrix} E_0 \\ E_1\widetilde{\mathcal{A}} \\ \vdots \\ E_{\mathrm{Np}}\widetilde{\mathcal{A}}^{\mathrm{Np}} \end{bmatrix},$$

and

$$D = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ E_1\widetilde{\mathcal{B}} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \dots & \mathbf{0} \\ E_{\mathrm{Np}}\widetilde{\mathcal{A}}^{\mathrm{Np}-1}\widetilde{\mathcal{B}} & E_{\mathrm{Np}}\widetilde{\mathcal{A}}^{\mathrm{Np}-2}\widetilde{\mathcal{B}} & \dots & E_{\mathrm{Np}}\widetilde{\mathcal{B}} \end{bmatrix}.$$

Note that the dimension of matrix R and the number of constraints d do not rely on the dimension of vector-valued observables z. We can solve the linear MPC problem on the same predictive horizon if we obtain $z_{t,0} = \psi^e(x_t, w^e)$. Fig. 6.4 shows the diagram of MPC-Deep-Koopman to compute the control sequence. Thus, we summarized the algorithm of the closed-loop operation of vector-valued observables as follows:

Algorithm 5 : MPC-Deep-Koopman to compute the control sequence of Problem (6.35)

for $t = 0, 1, \ldots, T$ **do**
  Let $z_{t,0} = \psi^e(x_t, w^e)$;
  Minimize Problem T_t to obtain an optimal solution u_t^*;
  Let $u_t = u_t^*(1:m)$;
  Update $\widetilde{x}_{t+1}$ with system (6.28).
  Obtain the practical x_{t+1} with SUMO by implementing the control u_t.
end for

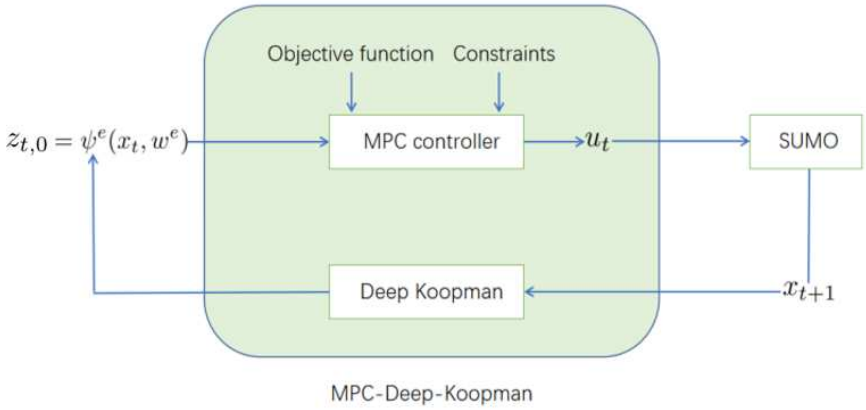


Figure 6.4: The diagram of MPC-Deep-Koopman to compute the control sequence

6.5 Numerical Studies

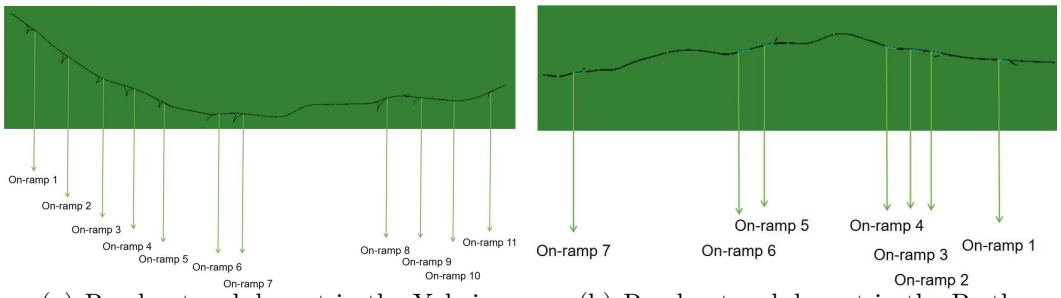
In this section, our proposed Deep Koopman model is verified in the microscopic simulated environment SUMO for various demand scenarios. We consider the comparisons with baselines.

6.5.1 Experiment Design

To validate the efficiency of our developed Deep Koopman model, we conduct two case studies in the SUMO simulation environment. SUMO is an open-source program of traffic flow simulation. We extract two freeway road networks, which are located in the Yubei and Perth vicinity, separately. They are one of the most congested freeway roads in the city of Chongqing, China, and Perth, Australia. Fig. 6.5 shows the actual map of the selected freeway road including its nearby layout. The selected freeway roads have a total length of around 27 km and 12 km, starting from Jinxing road and ending at Modern road with 11 on-ramps and 8 off-ramps, and beginning from Fiona Wood Road and ending at Narrows Bridge with 7 on-ramps and 4 off-ramps, separately. At every on-ramp, there is

(a) Road network layout in the Yubei (b) Road network layout in the Perth

Figure 6.5: Road network layout



(a) Road network layout in the Yubei (b) Road network layout in the Perth

Figure 6.6: Illustration of the SUMO simulation model for the investigated freeway traffic network

a ramp meter controlling the in-flow of vehicles. Fig. 6.6 illustrates the freeway traffic models in the SUMO simulation.

The training data for the proposed Deep Koopman model is sampled through the freeway traffic simulation with randomized controls on different scenarios. We generate 100 scenarios with varied traffic flows. The duration of each simulation is set as 4 hours (i.e. 14400 seconds) and the control interval is set as 60s. The simulation period is divided into peak and off-peak hours. For off-peak hours, the total in-flows for the main road and the on-ramps are generated randomly within the range of $[1000, 4000]$ and $[1, 400]$, respectively. For peak hours, the in-flows for the main road and the on-ramps would vary within the range of $[3000, 8000]$ and $[200, 800]$, respectively. 80 scenarios are used for the training data collection resulting in a total of 19200 data triplets, i.e., $(x_t,\ u_t,\ x_{t+1})$, while the other 20 scenarios are used for validation. Random controls are sampled uniformly within the range of $[u_{min}, u_{max}]$. During the simulation, we convert the control variable $a_{t,i} \in u_t, i \in \mathcal{N}$ for each on-ramp i to the duration of green light: $\Delta t = 60 * a_{t,i}$. In our experiment, all models are complied in Python 3.7 running on a computer with Intel Xeon E5-1650 v4, 3.6GHz CPU.

6.5.2 Evaluation metrics and baselines

To comprehensively evaluate the performance of the proposed Deep Koopman model, baseline approaches are selected for comparison. Firstly, we assess the performances of predicting the dynamics system between the proposed Deep Koopman model and extended dynamic mode decomposition (EDMD) [31]. For EDMD, we adopt the Gaussian radial basis function (RBF) [207] as the basis function $\psi(x)$:

$$\psi(x) = e^{(\epsilon||x-C||)^2}, \tag{6.36}$$

where ϵ denotes the shape parameter and $C \in \mathbb{R}^l$ is the center points. In our experiment, the kernel centers are selected using K-means clustering [208]. Secondly, we estimate the performance of the classic traffic control models including ALINEA regulator [142] and CTM-based deterministic optimization (CTM-DO) [73]. Thirdly, we validate the performance of different approaches on ramp metering control. Besides the EDMD, we design a linear control method [209] which approximates the dynamics system with a linear model on the original state. To illustrate the effectiveness of our proposed Deep Koopman model, we also evaluate the performances of traffic simulation with a fixed control and with no ramp meters. The fixed control means that each ramp meter would follow a fixed duration for each phase. In our experiment, we set the durations of green light, yellow light and red light as 39s, 3s and 18s, respectively. Note that models (CTM-DO, EDMD, Linear and Deep Koopman) are tested in MPC framework.

As indicated in Section 6.5.1, we generate 4800 testing data from 20 testing scenarios and the mean squared error (MSE) of prediction are utilized to assess the performance of traffic state prediction. The performance of traffic control is evaluated based on four metrics including the number of leaving vehicles per decision interval, the average travel time of vehicles, the total throughput of traffic out-flow, and the deviation of control variables.

6.5.3 Example one

A Parameters setting

In our experiment, we set the dimension of the lifted space as 1000. While training the Deep Koopman model, we run the mini-batch training with a batch size of 32 for 100 epochs. Xavier initialization [210] is adopted for initializing the weights of neural networks in the proposed Deep Koopman model. The details of the associated parameters are shown in Table 6.1. In particular, we define the matrix

$E_p \in \mathbb{R}^{9\times l}$ as follows:

$$E_p = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & \dots & 0 \end{bmatrix}, \ p = 0, \dots, \mathrm{Np}.$$

Table 6.1: Model parameters for Deep Koopman

Name of parameter	Value	Name of parameter	Value
$\alpha_1, \alpha_2, \alpha_3$	1	α_4, α_5	10^{-9}
$\widetilde{v}_{i,p}, i = 1, \dots, 9$	1	$\widetilde{v}_{i,p}, i = 10, \dots, 43$	0
$b_{min,ip}, i = 1, \dots, 9$	0veh/m	$b_{max,ip}, i = 1, \dots, 8$	33veh/m
$b_{max,ip}, i = 9$	66veh/m	$u_{min,ip}, i = 1 \dots, 11$	20%
$u_{max,ip}, i = 1 \dots, 11$	70%	$\triangle T$	60s
λ	10^{-4}	Np	17m
T	240	Nc	1m

For implementing the ALINEA regulator and the CTM model, we set their corresponding parameters as shown in Tables 6.2 and 6.3:

Table 6.2: Model parameters for ALINEA regulator

Name of parameter	Value	Name of parameter	Value
$o_{cr,i}, i = 1, \dots, 11, 14, 15$	10.35veh/m	$o_{cr,i}, i = 12, 13, 16, \dots, 23$	6.9veh/m
r_{max}	0.5veh/s	K_p	0.02veh/s

Table 6.3: Model parameters for CTM

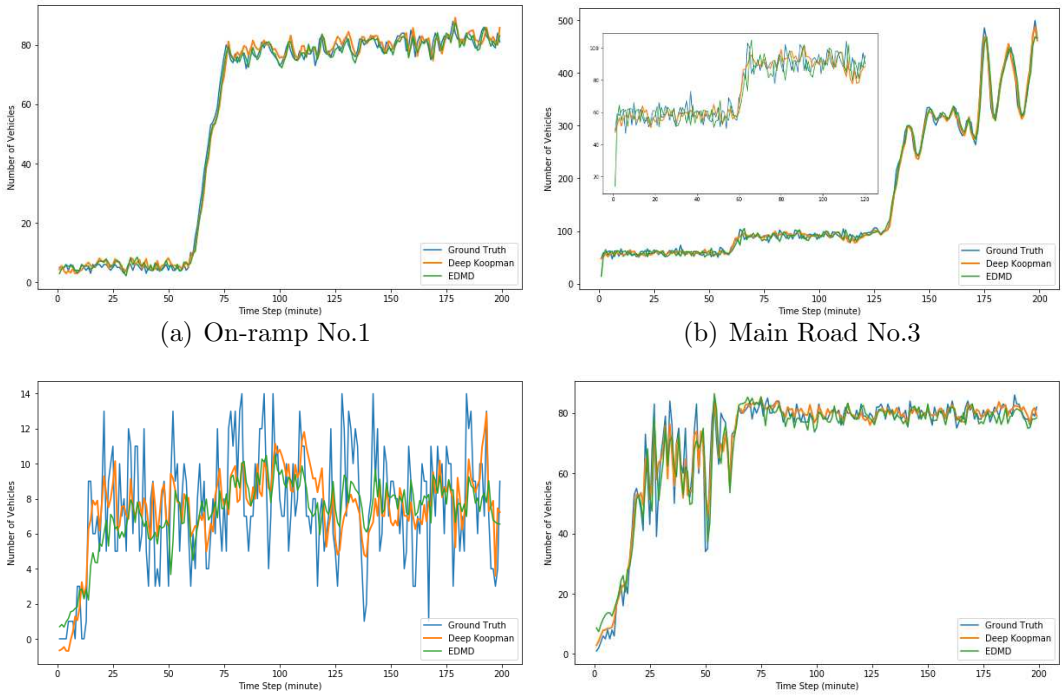
Name of parameter	Value	Name of parameter	Value
$Q_i, i = 1, \ldots, 11, 14, 15$	1.92veh/s	$Q_i, i = 12, 13, 16, \ldots, 23$	1.28veh/s
$\rho_{max,i}, i = 1, \ldots, 11, 14, 15$	0.4veh/m	$\rho_{max,i}, i = 12, 13, 16, \ldots, 23$	0.27veh/m
w	5.81m/s	v_f	27.78m/s
ΔT	5s	I	23
Np	324	Nc	12
r_{max}	0.5veh/s	q_{max}	60veh

Note that Np and Nc denote the prediction horizon and control horizon, respectively; ΔT is the simulation duration at each step; T represents total total simulation step; o_{cr} is the critical occupancy; r_{max} is the maximum ramp metering rate; Q denotes the capacity; ρ_{cr} is the critical density; ρ_{max} represents the jam density; w refers to the backward wave speed; v_f is the free flow speed; I represents the number of cells and q_{max} is the maximum queue length.

B Results and analysis

1) Traffic prediction: We first calculate the mean squared error of traffic state prediction: MSE $= \sum_{t=1}^{240} \sum_{i=1}^{43} (x_{t,i} - \widetilde{x}_{t,i})^2$. The MSE values of EDMD and Deep Koopman are 59.062 and 47.293, respectively. Obviously, the Deep Koopman model outperforms the EDMD with 24.885% improvement. Fig. 6.7 shows the comparison of traffic state prediction between EDMD and the proposed Deep Koopman for the scenario 1. The blue line is the actual traffic state. More specifically, Figs. 6.7(a)-6.7(d) show the predictions of the real-time number of vehicles on the on-ramp 1 and the main road segment 3, and the number of the leaving vehicles from the off-ramp 7 and the main road. We can observe that the Deep Koopman model has a higher accuracy compared to the EDMD. This implies that the proposed Deep Koopman model could better predict future states of the freeway traffic flow and have a more precise reconstruction of the state from the lifted space.

2) Traffic control performance analysis: To assess the performance of the controller, we present the average travel time (ATT) of the vehicles, the total throughput of vehicles (TTV) and the deviation of the control variables (DCV)



(a) On-ramp No.1 (b) Main Road No.3

(c) Off-ramp No.7 (d) Leaving from Main Road

Figure 6.7: The Comparison of State Prediction between EDMD and the proposed Deep Koopman for the Scenario 1

for four testing scenarios, as shown in Table 6.4, where ATT can be retrieved from the SUMO directly, TTV $= \sum_{t=0}^{240} v_t^\top x_t$ and DCV $= \sum_{t=0}^{238} \lambda ||u_{t+1} - u_t||^2$. It is obvious that all the baseline control methods could reduce the ATT and improve the TTV compared to No control strategy. For the traditional control methods, the ALINEA does not show a significant improve compared to the Fixed Control strategy, while the MPC-CTM-DO model is better performed on the ATT for all scenarios. However, their results on ATT are no better than that of MPC-EDMD and MPC-Deep-koopman. For example, the proposed MPC-Deep-Koopman model achieves a 18.55% lower ATT than the MPC-CTM-DO for the scenario 1, and reduces the ATT by 23.956% for the scenario 3, and by at least 21% for the other 3 cases compared to No Control strategy. The metric of TTV shows the performance of the controllers on optimizing the first part of the objective function in (6.30), while DCV represents the second part. While analyzing the results of TTV, the control strategy produced by the MPC-Deep-Koopman model allows 6391 more vehicles completing their trips during the simulation compared to the No Control approach for the scenario 2. There are 5000 vehicles more completing the travel for other 3 instances. It is worth mentioning that ALINEA and MPC-CTM-DO outperform the MPC-linear in scenario 3. The D-

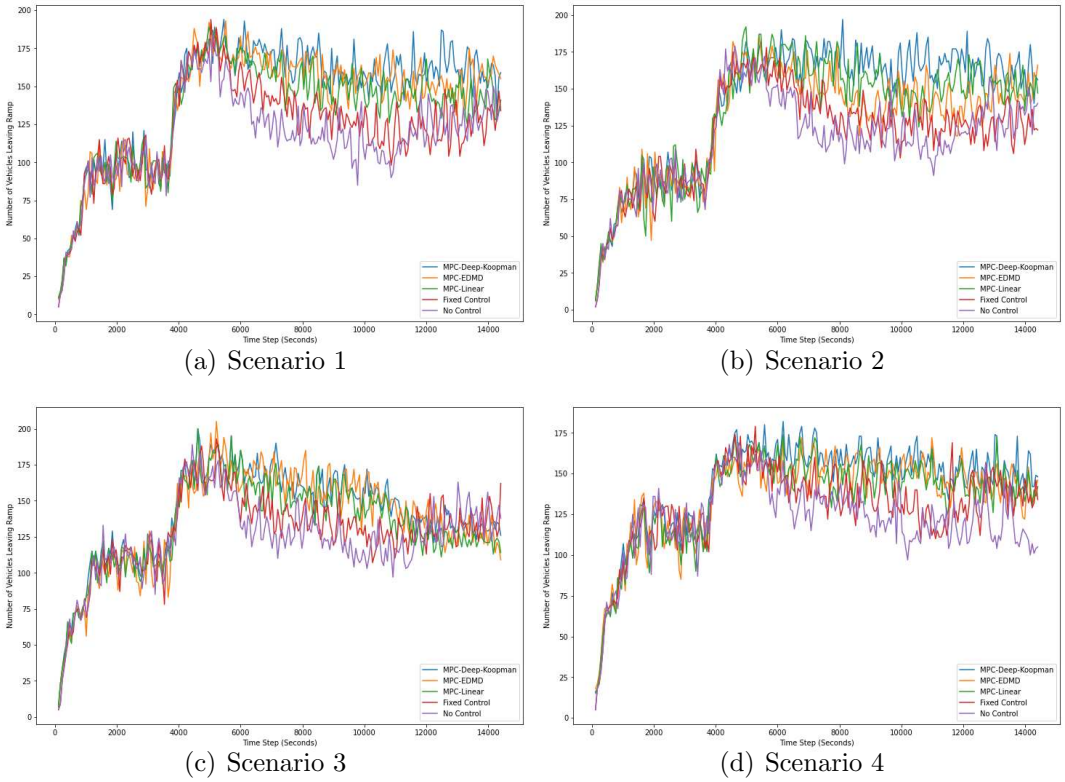
CV of the MPC-EDMD is better than the MPC-Linear model for all instances, while the proposed MPC-Deep-Koopman model has the best performance and ALINEA performs the worst on DCV. This is reasonable because ALINEA computes the control strategy based on the current state without considering future changes. Moreover, the controls of MPC-Deep-Koopman model are more stable and robust. We can conclude that our proposed MPC-Deep-Koopman model is more efficient compared to the baselines. The controller integrated with the MPC-Deep-Koopman is capable of mitigating traffic jams so that improving the freeway management.

Table 6.4: Performance Comparison of Average Objectives (ATT: Average Travel Time, TTV: Total Throughput, DCV: Deviation of Control)

Approach	Scenario 1			Scenario 2			Scenario 3			Scenario 4		
	ATT	TTV	DCV	ATT	TTV	DCV	ATT	TTV	DCV	ATT	TTV	DCV
ALINEA	1833.23	28951	0.36130	1774.16	27643	0.33570	1688.34	31383	0.37520	1566.15	31122	0.37640
MPC-CTM-DO	1790.06	31328	0.31730	1739.92	27968	0.28430	1578.85	31355	0.29720	1496.82	32613	0.27830
MPC-EDMD	1571.74	32613	0.20779	1514.45	30948	0.20336	1459.89	32674	0.22930	1370.05	32723	0.15006
MPC-Linear	1626.27	32290	0.28700	1554.33	32195	0.21649	1549.06	31057	0.32817	1433.83	31608	0.32998
Fixed Control	1810.53	29403	0	1774.38	28662	0	1763.53	30582	0	1640.80	31029	0
No Control	1893.01	27998	-	1829.19	27560	-	1792.48	28903	-	1696.09	29106	-
MPC-Deep-Koopman	**1457.96**	**34333**	**0.09384**	**1433.62**	**33951**	**0.13973**	**1363.08**	**34267**	**0.15235**	**1295.33**	**34615**	**0.11362**

3) Number of leaving vehicles: To further visualize and validate the efficiency of the MPC-Deep-Koopm, we monitor the real-time number of leaving vehicles from the freeway to assess the performance of controllers on improving the traffic throughput. We present the results of 4 different scenarios obtained by the proposed MPC-Deep-Koopman model and other baselines, as shown in Fig. 6.8. As we can observe, the performance of No Control is the worst for all 4 scenarios. This is because the increase of traffic flow on the freeway could lead to serious traffic jams, especially when no control actions are applied. It is obvious that the number of leaving vehicles increases after 450s in simulation when a controller is adopted. Compared to the MPC-Linear model, MPC-EDMD and the Fixed Control, the proposed MPC-Deep-Koopman model has the most significant improvement on the real-time throughput, especially in the scenarios 1, 2 and 4. The results imply that the developed MPC-Deep-Koopman model is more effective in improving the traffic operation.

4) Ramp control variation: We show the results of ramp metering control for approaches including the MPC-Linear model, the MPC-EDMD and the MPC-Deep-Koopman in scenarios 1 and 3. Figs. 6.9(a)-6.9(b) show the control variables for the on-ramp 1. We find the variation of controls showing an opposite

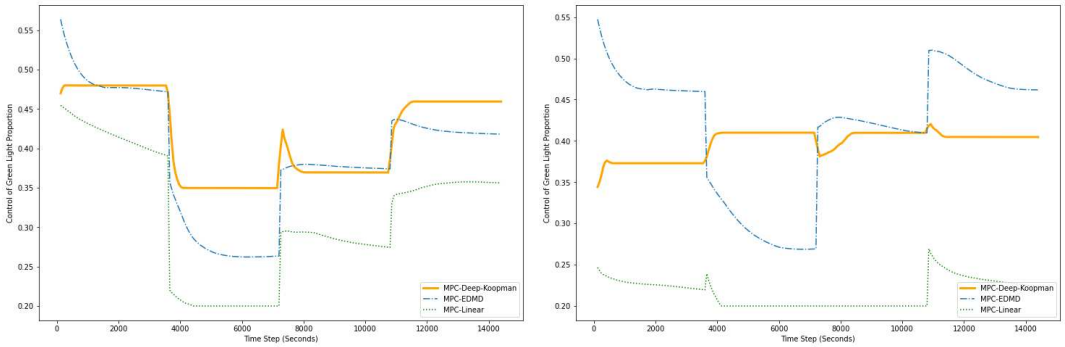


(a) Scenario 1 (b) Scenario 2

(c) Scenario 3 (d) Scenario 4

Figure 6.8: The Comparison of Number of Leaving Vehicles between baseline approaches and the proposed MPC-Deep-Koopman model

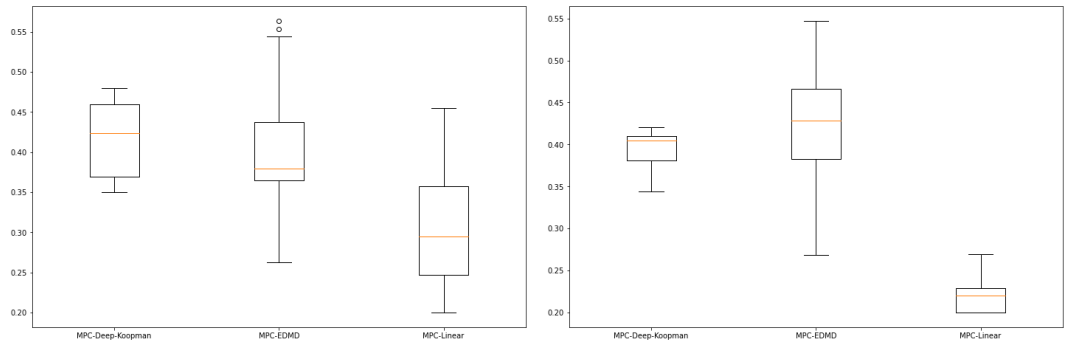
trend with the changing of the traffic demand flow. As shown in Fig. 6.9(a), we can see that the higher demand results in shorter duration of green light. This is consistent with the actual situation, because reducing the green phase duration of traffic light of on-ramps can stop vehicles from merging to the main road from on-ramps so that mitigating the traffic jams on the main road. Fig. 6.10 shows the comparison of the fluctuations of the ramp control for the on-ramp 1. Red lines in Figs. 6.10(a)-6.10(b) represent the median of ramp metering control variables. The bottom and top lines are the minimum and maximum values, respectively. The bottom and top edges of quadrangle refer to the first quarter and the third quarter of the quantile, respectively. By comparison, we can observe that the MPC-EDMD produces the controls with a median value around 0.37 and the MPC-Linear model has the controls with a median value around 0.28, and the median value of our controller is around 0.42 in the scenario 1. The MPC-Deep-Koopman and MPC-EDMD achieve similar values on instance 3, but the MPC-Linear model has a lower median value. In addition, the MPC-EDMD and the MPC-Linear model have greater fluctuations than the MPC-Deep-Koopman model, which implies that the MPC-Deep-Koopman model has a better performance on optimizing the deviation of controls.



(a) Scenario 1 (b) Scenario 3

Figure 6.9: The Comparison of Ramp Control for On-ramp No.1

5) Computational time: In this section, we present the actual computation time for each MPC closed loop for the ALINEA, MPC-CTM-DO, MPC-Linear model, the MPC-EDMD and our proposed MPC-Deep Koopman, as shown in Table 6.5. As we can observe, the ALINEA and MPC-Linear are the fastest approaches as they have simpler modeling of dynamics, while the MPC-CTM-DO outperforms the MPC-EDMD and MPC-Deep Koopman. Moreover, the computation time of MPC-Deep-Koopman is 0.8852s, which means that the running time is less than 1s at each time step when the optimal control strategy



(a) Scenario 1 (b) Scenario 3

Figure 6.10: The Comparison of Ramp Control Fluctuation for On Ramp No.1

is obtained. This is acceptable for our problem and can satisfy the practical requirements.

Table 6.5: Computation time for baselines and MPC-Deep-Koopman for each MPC computation with Np = 17m, Nc = 1m

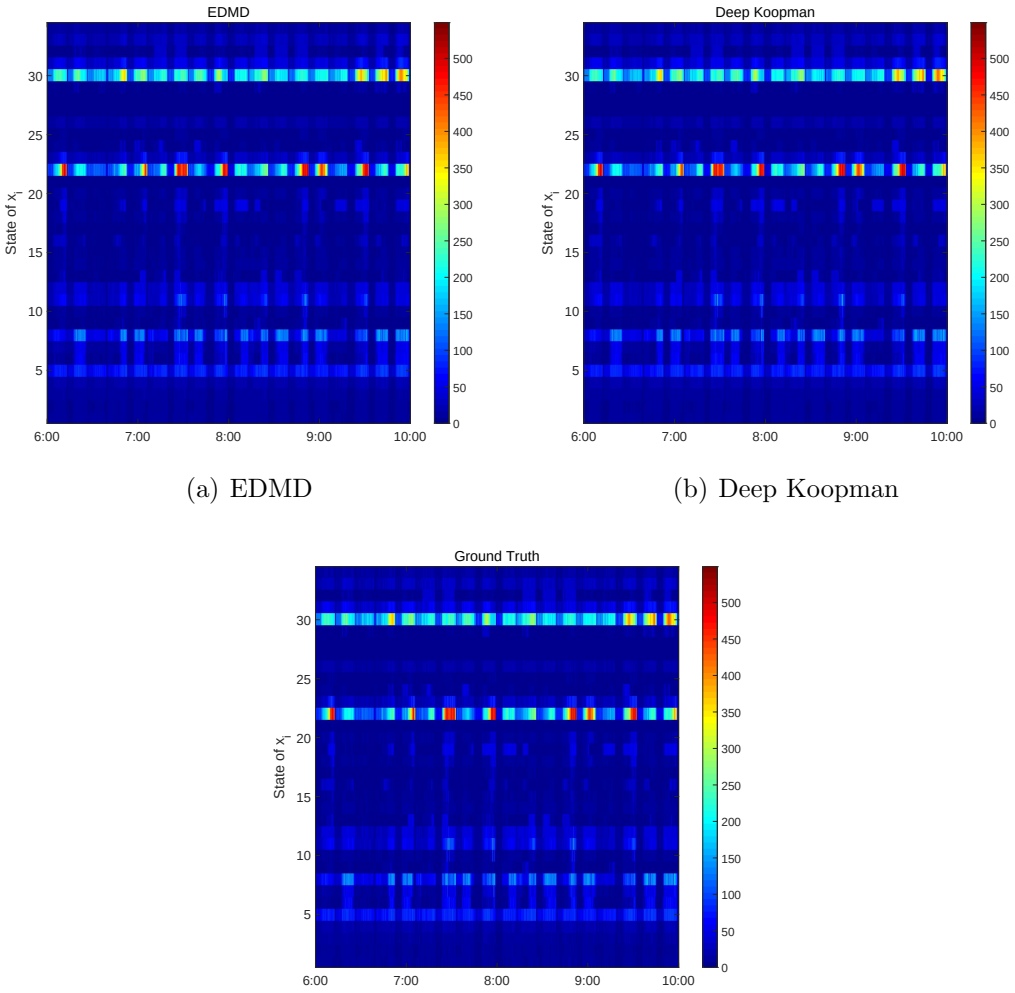
Approaches	Computation Time (s)
ALINEA	0.2388
MPC-CTM-DO	0.6992
MPC-Linear	0.3581
MPC-EDMD	2.7641
MPC-Deep-Koopman	0.8852

6.5.4 Example two

A Results and analysis

1) Traffic prediction: We first show traffic state prediction of x for EDMD, the Deep Koopman and the Ground truth in Fig. 6.11. Obviously, the Deep Koopman model and DEMD can predict the traffic state evolution. By comparison, we observe that the Deep Koopman model has a higher accuracy compared to the EDMD. This implies that the proposed Deep Koopman model could better predict future states of freeway traffic flows and have a more precise reconstruction of the state from the lifted space.

2) Number of leaving vehicles: We monitor the leaving vehicles from the freeway main road to validate the efficiency of controllers. We show the results

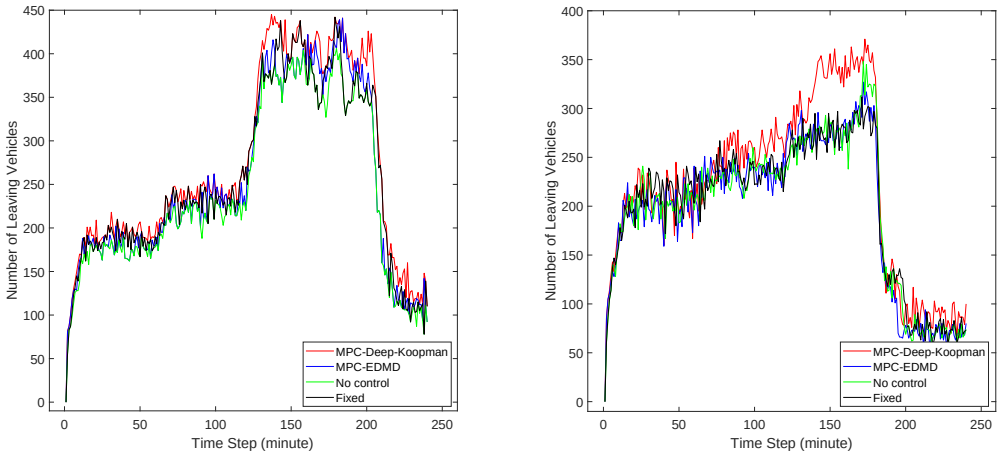


(a) EDMD

(b) Deep Koopman

(c) Ground truth

Figure 6.11: The Comparison of State Prediction between EDMD, the proposed Deep Koopman and Ground truth



(a) Scenario one (b) Scenario two

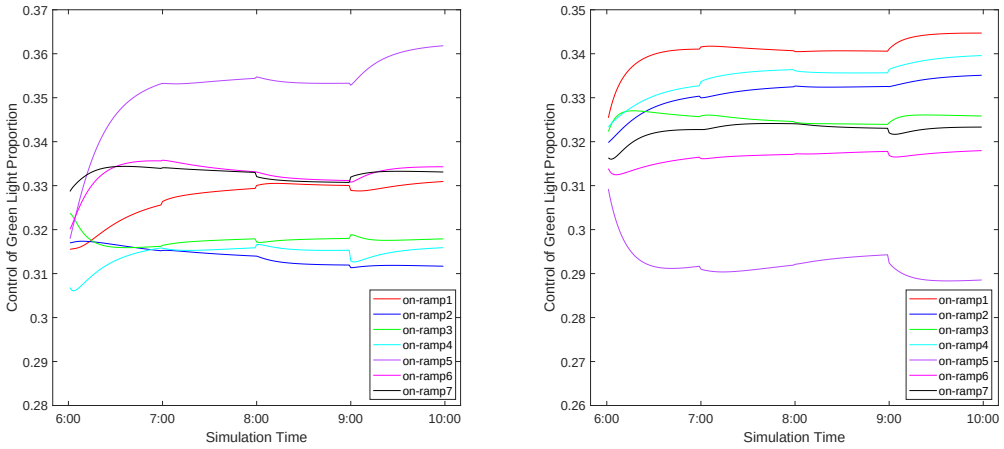
Figure 6.12: The Comparison of Number of Leaving Vehicles between baseline approaches and the proposed MPC-Deep-Koopman model

of 2 different scenarios obtained by the proposed MPC-Deep-Koopman model and other basis lines, as shown in Fig. 6.12. By observing, we find that the performance of No Control is the worst for all scenarios. Compared to the MPC-EDMD and the Fixed Control, the proposed MPC-Deep-Koopman model has the most significant improvement on the real-time throughput, especially in the scenario 2. Among the baseline approaches, the Fixed Control has the least improvement on the number of leaving vehicles. The result indicates that the proposed MPC-Deep-Koopman model is more effective.

3) Ramp control: We show the results of ramp metering control for the MPC-Deep-Koopman model in two different scenarios. Figs. 6.13(a)-6.13(b) show the control evolution for all on-ramps. By observing, we find that the change trend of control variable in on-ramp 5 is significant for different scenarios, but other control metering has tiny change and is mainly concentrated in the interval $[0.31, 0.34]$. This indicates that the MPC-Deep-Koopman model is more applicable, which is consistent with the results of example one.

6.6 Conclusion

In this Chapter, we study a ramp metering control problem on the freeway traffic. Because of the high non-linearity and complexity of the freeway traffic flow dynamics, we use the Koopman operator to model the evolution of the dynamical system due to its interpretability and linearity. To learn a finite-dimensional approximation of the Koopman operator, we propose a Deep Koopman model



(a) Scenario one (b) Scenario two

Figure 6.13: The evolution of Ramp Control of the proposed MPC-Deep-Koopman

based on the neural networks with an encoder-decoder structure. LSTM units are used to process the sequential traffic data in the encoder. Compared to the EDMD, the Deep Koopman model is a data-driven approach of learning the lifted state of dynamics without any prior information of the basis function selection. To provide a real-time control of the ramp metering, we design a model predictive control with pre-trained Deep Koopman model to maximize the total throughput of vehicles on the freeway. SUMO environment is used for the experimental studies to assess the effectiveness of the proposed approach. By Comparison with the EDMD with RBF basis functions, the proposed Deep Koopman model has shown a better precision in the predictions of the dynamical states. We also compare the performance of the real-time control of ramp metering between the MPC-Deep-Koopman, MPC-EDMD, and classic traffic control models. The MPC-Deep-Koopman model outperforms the baselines on all metrics, which demonstrates the effectiveness of the proposed approach. The future research will focus on the ramp metering problem on the freeway with dynamical speed limits.

CHAPTER 7

Conclusions and future work

7.1 Conclusions

In this thesis, we study two kinds of freeway optimization control models taking into account of ramp metering and variable speed limits. The former is modelled on the basis of the macroscopic traffic flow models, namely the CTM. Based on this model, we study three different optimization control models. Firstly, we propose an optimization model with stochastic demand flows to implement ramp metering. Secondly, we design a ramp metering optimal control model with smooth objective and constraint functions that handles the flow holding-back problem. Finally, we develop a smooth model that considers capacity drop, ramp metering and variable speed limits. The latter is a data-driven model on the basis of the Koopman operator. The main contributions are presented in two aspects as follows:

(1) The proposed models have theoretical guarantees.

(i) With available partial information of random traffic demands, a distributionally robust chance constrained optimization model is formulated as a semi-definite programming using the Worst-Case Conditional Value-at-Risk (WCVaR) constraints to approximate distributionally robust chance constraints. This transformation has the theoretical guarantee. Moreover, an optimal control is obtained by optimizing the approximated problem with the objective of total delay of main road and on-ramps.

(ii) Two novel smooth optimization models on the basis of the CTM were proposed to handling the flow holding-back problem resulted from the impractical fundamental diagram if the 'min operators are relaxed into inequality constraints directly. The former only takes the freeway ramp metering into consideration, and the latter considers both ramp metering and variables speed limits under

the capacity drop. There are theoretical analyses for both proposed models. Theoretical results show that the solution derived from the proposed method converges that of primal CTM when ϵ approaches to zero.

(iii) A data-driven model in the framework of Koopman operator is utilized to model the evolution of the freeway dynamical system because of its interpretability and linearity. To learn a finite-dimensional approximation of the Koopman operator, we propose a Deep Koopman model based on the neural networks with an encoder-decoder structure. LSTM units are used to process the sequential traffic data in the encoder.

(2) The numerical results obtained from the study have validated the efficiency of our proposed models.

(i) The numerical results of three various optimization control models, i.e., deterministic optimization model, robust optimization model and the proposed model, were showed and compared. These results indicate that our proposed model is efficient to control the total delay of the system when the considered dynamical system encounters uncertainties in a series of scenarios. The results also demonstrate that the proposed model is more effective in reducing traffic congestion of the system.

(ii) The proposed two smooth optimal control models can effectively solve the flow holding-back problem under practical and synthetic applications. Moreover, both RM and VSL are efficient control strategies of mitigating traffic congestion. Furthermore, the performance is the best when RM and VSL are implemented cooperatively.

(iii) A real-time control of ramp metering was provided based on the proposed Deep Koopman model. SUMO environment was used for the experimental studies to assess the performance of the developed approach. Compared to the EDMD with RBF basis functions, the proposed Deep Koopman model has shown a better precision in the predictions of the states. We also assessed the efficiency of the real-time control of ramp metering between the Deep Koopman model, EDMD, Linear model, and Fixed control. The Deep Koopman model outperforms other baselines on all metrics, which demonstrates the effectiveness of the proposed model.

7.2 Future work

The future work may involve different directions, we briefly summarize it as follows:

(i) In the objective one, we only consider simple case that the dynamical freeway system has uncertain external demand flows with partial information under triangle fundamental diagrams. However, the obtained parameters from triangle fundamental diagrams, for example, critical density and jam density significantly impact the performance of the proposed model, especially for limited data. Hence, considering the set-valued fundamental diagram under the distributionally robust optimization is necessary.

(ii) Even though the objectives two and three solve the flow holding-back problem caused by inequality relaxation making use of smooth functions to approximate the "min" operators, but the considered scenario is simple and the problem only has one goal. Therefore, designing multi-objective optimization model such as the traffic safety and the pollution reduction under complex scenarios is ideal choice. Moreover, route guidance should also be taken into consideration.

(iii) Objective four only uses ramp metering to alleviate traffic congestion, but it can not solve the congestion problem completely. Hence, it is necessary to propose a freeway optimal control model of the ramp metering with dynamical speed limits and route guidance under the data-driven framework. Furthermore, large-scale numerical applications are necessary.

Bibliography

[1] Cambridge Systematics. Traffic congestion and reliability: Trends and advanced strategies for congestion mitigation. Technical report, United States. Federal Highway Administration, 2005.

[2] Louis A Pipes. An operational analysis of traffic dynamics. *Journal of applied physics*, 24(3):274–281, 1953.

[3] Masako Bando, Katsuya Hasebe, Akihiro Nakayama, Akihiro Shibata, and Yuki Sugiyama. Dynamical model of traffic congestion and numerical simulation. *Physical review E*, 51(2):1035, 1995.

[4] Rui Jiang, Qingsong Wu, and Zuojin Zhu. Full velocity difference model for a car-following theory. *Physical Review E*, 64(1):017101, 2001.

[5] Kai Nagel and Michael Schreckenberg. A cellular automaton model for freeway traffic. *Journal de physique I*, 2(12):2221–2229, 1992.

[6] Katsuhiro Nishinari and Daisuke Takahashi. Analytical properties of ultradiscrete burgers equation and rule-184 cellular automaton. *Journal of Physics A: Mathematical and General*, 31(24):5439, 1998.

[7] Ofer Biham, A Alan Middleton, and Dov Levine. Self-organization and a dynamical transition in traffic-flow models. *Physical Review A*, 46(10):R6124, 1992.

[8] Michael James Lighthill and Gerald Beresford 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*, 229(1178):317–345, 1955.

[9] Paul I Richards. Shock waves on the highway. *Operations research*, 4(1):42–51, 1956.

[10] Harold J Payne. Model of freeway traffic and control. *Mathematical Model of Public System*, pages 51–61, 1971.

[11] Rui Jiang, Qing-Song Wu, and Zuo-Jin Zhu. A new continuum model for traffic flow and numerical tests. *Transportation Research Part B: Methodological*, 36(5):405–419, 2002.

[12] Yisheng Lv, Yanjie Duan, Wenwen Kang, Zhengxi Li, and Fei-Yue Wang. Traffic flow prediction with big data: a deep learning approach. *IEEE Transactions on Intelligent Transportation Systems*, 16(2):865–873, 2014.

[13] Yuankai Wu, Huachun Tan, Lingqiao Qin, Bin Ran, and Zhuxi Jiang. A hybrid deep learning based traffic flow prediction method and its understanding. *Transportation Research Part C: Emerging Technologies*, 90:166–180, 2018.

[14] Zhiyong Cui, Kristian Henrickson, Ruimin Ke, and Yinhai Wang. Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting. *IEEE Transactions on Intelligent Transportation Systems*, 21(11):4883–4894, 2019.

[15] Xueyan Yin, Genze Wu, Jinze Wei, Yanming Shen, Heng Qi, and Baocai Yin. Deep learning on traffic prediction: Methods, analysis and future directions. *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[16] Fuwen Deng, Jiandong Jin, Yu Shen, and Yuchuan Du. Advanced self-improving ramp metering algorithm based on multi-agent deep reinforcement learning. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 3804–3809. IEEE, 2019.

[17] Saeed Ghanbartehrani, Anahita Sanandaji, Zahra Mokhtari, and Kimia Tajik. A novel ramp metering approach based on machine learning and historical data. *Machine Learning and Knowledge Extraction*, 2(4):379–396, 2020.

[18] Yue Zhou, Kaan Ozbay, Pushkin Kachroo, and Fan Zuo. Ramp metering for a distant downstream bottleneck using reinforcement learning with value function approximation. *Journal of Advanced Transportation*, 2020, 2020.

[19] Allan M Avila. *Applications of Koopman Operator Theory to Highway Traffic Dynamics*. University of California, Santa Barbara, 2017.

[20] Esther Ling, Lillian Ratliff, and Samuel Coogan. Koopman operator approach for instability detection and mitigation in signalized traffic. In

2018 21st International Conference on Intelligent Transportation Systems (ITSC), pages 1297–1302. IEEE, 2018.

[21] Esther Ling, Liyuan Zheng, Lillian J Ratliff, and Samuel Coogan. Koopman operator applications in signalized traffic systems. *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[22] Mohammad Saifuzzaman and Zuduo Zheng. Incorporating human-factors in car-following models: a review of recent developments and research needs. *Transportation research part C: emerging technologies*, 48:379–403, 2014.

[23] Peter G Gipps. A behavioural car-following model for computer simulation. *Transportation Research Part B: Methodological*, 15(2):105–111, 1981.

[24] Mark Brackstone and Mike McDonald. Car-following: a historical review. *Transportation Research Part F: Traffic Psychology and Behaviour*, 2(4):181–196, 1999.

[25] Xiao Wang, Rui Jiang, Li Li, Yilun Lin, Xinhu Zheng, and Fei-Yue Wang. Capturing car-following behaviors by deep learning. *IEEE Transactions on Intelligent Transportation Systems*, 19(3):910–920, 2017.

[26] BD Greenshields, JR Bibbins, WS Channing, and HH Miller. A study of traffic capacity. In *Highway research board proceedings*, volume 1935. National Research Council (USA), Highway Research Board, 1935.

[27] EL Ericksen. Traffic performance at urban intersections. *Traffic Quarterly*, 1(3), 1947.

[28] Harold Greenberg. An analysis of traffic flow. *Operations research*, 7(1):79–85, 1959.

[29] Gordon Frank Newell. Nonlinear effects in the dynamics of car following. *Operations research*, 9(2):209–229, 1961.

[30] RT Underwood, Volume Speed, and Density Relationships. Quality and theory of traffic flow. *Yale Bureau of Highway Traffic*, 141:188, 1961.

[31] Jennifer Drake, Joseph Schofer, and ADJR May. A statistical analysis of speed-density hypotheses. *Traffic Flow and Transportation*, 1965.

[32] PK Munjal and Louis Albert Pipes. Propagation of on-ramp density perturbations on unidirectional two-and three-lane freeways. *Transportation Research/UK/*, 1971.

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

[34] Boris S Kerner and P Konhäuser. Structure and parameters of clusters in traffic flow. *Physical Review E*, 50(1):54, 1994.

[35] Matthias Herrmann and Boris S Kerner. Local cluster effect in different traffic flow models. *Physica A: Statistical Mechanics and Its Applications*, 255(1-2):163–188, 1998.

[36] Gerald Beresford Whitham. *Linear and nonlinear waves*. John Wiley & Sons, 2011.

[37] Markos Papageorgiou. A hierarchical control system for freeway traffic. *Transportation Research Part B: Methodological*, 17(3):251–261, 1983.

[38] R Kühne. Macroscopic freeway model for dense traffic-stop-start waves and incident detection. *Transportation and traffic theory*, 9:21–42, 1984.

[39] Peter Berg and Andrew Woods. Traveling waves in an optimal velocity model of freeway traffic. *Physical review E*, 63(3):036107, 2001.

[40] Carlos F Daganzo. Requiem for second-order fluid approximations of traffic flow. *Transportation Research Part B: Methodological*, 29(4):277–286, 1995.

[41] AATM Aw and Michel Rascle. Resurrection of" second order" models of traffic flow. *SIAM journal on applied mathematics*, 60(3):916–938, 2000.

[42] H Michael Zhang. A non-equilibrium traffic model devoid of gas-like behavior. *Transportation Research Part B: Methodological*, 36(3):275–290, 2002.

[43] Yu Xue and Shi-Qiang Dai. Continuum traffic model with the consideration of two delay time scales. *Physical Review E*, 68(6):066123, 2003.

[44] Panos G Michalopoulos, Dimitrios E Beskos, and Yasuji Yamauchi. Multilane traffic flow dynamics: some macroscopic considerations. *Transportation Research Part B: Methodological*, 18(4-5):377–395, 1984.

[45] Edward N Holland and Andrew W Woods. A continuum model for the dispersion of traffic on two-lane roads. *Transportation Research Part B: Methodological*, 31(6):473–485, 1997.

[46] Jorge A Laval and Carlos F Daganzo. Lane-changing in traffic streams. *Transportation Research Part B: Methodological*, 40(3):251–264, 2006.

[47] Haijun Huang, Tieqiao Tang, and Ziyou Gao. Continuum modeling for two-lane traffic flow. *Acta Mechanica Sinica*, 22(2):131–137, 2006.

[48] GCK Wong and SC Wong. A multi-class traffic flow model–an extension of lwr model with heterogeneous drivers. *Transportation Research Part A: Policy and Practice*, 36(9):827–841, 2002.

[49] Peng Zhang, Ru-Xun Liu, SC Wong, and Shi-Qiang Dai. Hyperbolicity and kinematic waves of a class of multi-population partial differential equations. *European Journal of Applied Mathematics*, 17(2):171–200, 2006.

[50] Peng Zhang, Sze Chun Wong, and Chi-Wang Shu. A weighted essentially non-oscillatory numerical scheme for a multi-class traffic flow model on an inhomogeneous highway. *Journal of Computational Physics*, 212(2):739–756, 2006.

[51] Sylvie Benzoni-Gavage and Rinaldo M Colombo. An-populations model for traffic flow. *European Journal of Applied Mathematics*, 14(5):587–612, 2003.

[52] Stéphane Chanut and Christine Buisson. Macroscopic model and its numerical solution for two-flow mixed traffic with different speeds and lengths. *Transportation research record*, 1852(1):209–219, 2003.

[53] Takashi Nagatani. Modified kdv equation for jamming transition in the continuum models of traffic. *Physica A: Statistical Mechanics and its Applications*, 261(3-4):599–607, 1998.

[54] Boris S Kerner and Peter Konhäuser. Cluster effect in initially homogeneous traffic flow. *Physical review E*, 48(4):R2335, 1993.

[55] Zhu Hui-Bing. Lattice models of traffic flow considering drivers delay in response. *Chinese Physics B*, 18(4):1322, 2009.

[56] Yi-Rong Kang and Di-Hua Sun. Lattice hydrodynamic traffic flow model with explicit drivers physical delay. *Nonlinear Dynamics*, 71(3):531–537, 2013.

[57] Zhang Min, Sun Di-Hua, and Tian Chuan. An extended two-lane traffic flow lattice model with drivers delay time. *Nonlinear Dynamics*, 77(3):839–847, 2014.

[58] H_X Ge, SQ Dai, Yu Xue, and LY Dong. Stabilization analysis and modified korteweg–de vries equation in a cooperative driving system. *Physical Review E*, 71(6):066119, 2005.

[59] Tao Wang, Zi-You Gao, and Xiao-Mei Zhao. Multiple flux difference effect in the lattice hydrodynamic model. *Chinese Physics B*, 21(2):020512, 2012.

[60] Sun Di-Hua, Tian Chuan, and Liu Wei-Ning. A traffic flow lattice model considering relative current influence and its numerical simulation. *Chinese Physics B*, 19(8):080514, 2010.

[61] Jianzhong Chen, Zhongke Shi, Lei Yu, and Zhiyuan Peng. Nonlinear analysis of a new extended lattice model with consideration of multi-anticipation and driver reaction delays. *Journal of Computational and Nonlinear Dynamics*, 9(3), 2014.

[62] Takashi Nagatani. Jamming transitions and the modified korteweg–de vries equation in a two-lane traffic flow. *Physica A: Statistical Mechanics and Its Applications*, 265(1-2):297–310, 1999.

[63] Takashi Nagatani. Jamming transition in a two-dimensional traffic flow model. *Physical Review E*, 59(5):4857, 1999.

[64] HM Zhang. A finite difference approximation of a non-equilibrium traffic flow model. *Transportation Research Part B: Methodological*, 35(4):337–365, 2001.

[65] HY Lee, H-W Lee, and D Kim. Origin of synchronized traffic flow on highways and its dynamic phase transitions. *Physical Review Letters*, 81(5):1130, 1998.

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

[67] Carlos F Daganzo. The cell transmission model, part ii: network traffic. *Transportation Research Part B: Methodological*, 29(2):79–93, 1995.

[68] Carlos F Daganzo. A finite difference approximation of the kinematic wave model of traffic flow. *Transportation Research Part B: Methodological*, 29(4):261–276, 1995.

[69] Carlos F Daganzo. The lagged cell-transmission model. 1999.

[70] Wai Yuen Szeto. Enhanced lagged cell-transmission model for dynamic traffic assignment. *Transportation Research Record*, 2085(1):76–85, 2008.

[71] Laura Muñoz, Xiaotian Sun, Roberto Horowitz, and Luis Alvarez. Traffic density estimation with the cell transmission model. In *Proceedings of the 2003 American Control Conference, 2003.*, volume 5, pages 3750–3755. IEEE, 2003.

[72] Laura Muñoz, Xiaotian Sun, Roberto Horowitz, and Luis Alvarez. Piecewise-linearized cell transmission model and parameter calibration methodology. *Transportation Research Record*, 1965(1):183–191, 2006.

[73] Gabriel Gomes and Roberto Horowitz. Optimal freeway ramp metering using the asymmetric cell transmission model. *Transportation Research Part C: Emerging Technologies*, 14(4):244–262, 2006.

[74] Gabriel Gomes, Roberto Horowitz, Alex A Kurzhanskiy, Pravin Varaiya, and Jaimyoung Kwon. Behavior of the cell transmission model and effectiveness of ramp metering. *Transportation Research Part C: Emerging Technologies*, 16(4):485–513, 2008.

[75] Xi-Qun Chen, Wei-Jun Xie, Jing Shi, and Qi-Xin Shi. Perturbation and stability analysis of the multi-anticipative intelligent driver model. *International Journal of Modern Physics C*, 21(05):647–668, 2010.

[76] Hong K Lo. A novel traffic signal control formulation. *Transportation Research Part A: Policy and Practice*, 33(6):433–448, 1999.

[77] Hong K Lo. A cell-based traffic control formulation: strategies and benefits of dynamic timing plans. *Transportation Science*, 35(2):148–164, 2001.

[78] Rasool Mohebifard and Ali Hajbabaie. Dynamic traffic metering in urban street networks: Formulation and solution algorithm. *Transportation research part C: emerging technologies*, 93:161–178, 2018.

[79] Rasool Mohebifard and Ali Hajbabaie. Optimal network-level traffic signal control: A benders decomposition-based solution algorithm. *Transportation Research Part B: Methodological*, 121:252–274, 2019.

[80] Maria Kontorinaki, Anastasia Spiliopoulou, Claudio Roncoli, and Markos Papageorgiou. First-order traffic flow models incorporating capacity drop: Overview and real-data validation. *Transportation Research Part B: Methodological*, 106:52–75, 2017.

[81] Jiancheng Long, Ziyou Gao, Xiaomei Zhao, Aiping Lian, and Penina Orenstein. Urban traffic jam simulation based on the cell transmission model. *Networks and Spatial Economics*, 11(1):43–64, 2011.

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

[83] Hong K Lo, Elbert Chang, and Yiu Cho Chan. Dynamic network traffic control. *Transportation Research Part A: Policy and Practice*, 35(8):721–744, 2001.

[84] JianCheng Long, ZiYou Gao, HuaLing Ren, and AiPing Lian. Urban traffic congestion propagation and bottleneck identification. *Science in China Series F: Information Sciences*, 51(7):948–964, 2008.

[85] Andy HF Chow and Hong K Lo. Sensitivity analysis of signal control with physical queuing: Delay derivatives and an application. *Transportation Research Part B: Methodological*, 41(4):462–477, 2007.

[86] Vinayak V Dixit and Essam Radwan. Hurricane evacuation: origin, route, and destination. *Journal of Transportation Safety & Security*, 1(1):74–84, 2009.

[87] Bernard O Koopman. Hamiltonian systems and transformation in hilbert space. *Proceedings of the national academy of sciences of the united states of america*, 17(5):315, 1931.

[88] I Antoniou, L Dmitrieva, Yu Kuperin, and Yu Melnikov. Resonances and the extension of dynamics to rigged hilbert space. *Computers & Mathematics with Applications*, 34(5-6):399–425, 1997.

[89] I Antoniou, K Gustafson, and Z Suchanecki. On the inverse problem of statistical physics: from irreversible semigroups to chaotic dynamics. *Physica A: Statistical Mechanics and its Applications*, 252(3-4):345–361, 1998.

[90] I Antoniou, VA Sadovnichii, and SA Shkarin. Time operators and shift representation of dynamical systems. *Physica A: Statistical Mechanics and its Applications*, 269(2-4):299–313, 1999.

[91] Enoch Yeung, Andrew Ng, Jongmin Kim, Zachary Z Sun, and Richard M Murray. Modeling the effects of compositional context on promoter activity in an e. coli extract based transcription-translation system. In *53rd IEEE Conference on Decision and Control*, pages 5405–5412. IEEE, 2014.

[92] Yoshihiko Susuki and Igor Mezic. Nonlinear koopman modes and coherency identification of coupled swing dynamics. *IEEE Transactions on Power Systems*, 26(4):1894–1904, 2011.

[93] Yoshihiko Susuki and Igor Mezic. Nonlinear koopman modes and a precursor to power system swing instabilities. *IEEE Transactions on Power Systems*, 27(3):1182–1191, 2012.

[94] A Mounir Boudali, Peter J Sinclair, Richard Smith, and Ian R Manchester. Human locomotion analysis: Identifying a dynamic mapping between upper and lower limb joints using the koopman operator. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1889–1892. IEEE, 2017.

[95] Bryan Eisenhower, Tobias Maile, Martin Fischer, and Igor Mezic. Decomposing building system data for model validation and analysis using the koopman operator. In *Proceedings of the National IBPSAUSA Conference, New York, USA*, 2010.

[96] Michael Georgescu, Bryan Eisenhower, and Igor Mezi. Creating zoning approximations to building energy models using the koopman operator. *Proceedings of SimBuild*, 5(1):40–47, 2012.

[97] Mohammad N Murshed and M Monir Uddin. Time delay coordinate based dynamic mode decomposition of a compressible signal. In *2019 22nd International Conference on Computer and Information Technology (ICCIT)*, pages 1–5. IEEE, 2019.

[98] Indranil Nayak and Fernando L Teixeira. Dynamic mode decomposition for prediction of kinetic plasma behavior. In *2020 International Applied Computational Electromagnetics Society Symposium (ACES)*, pages 1–2. IEEE, 2020.

[99] S Ananthakrishnan, P Geetha, and KP Soman. Temperature forecasting using dynamic mode decomposition. In *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*, pages 1590–1596. IEEE, 2020.

[100] Clarence W Rowley, Igor Mezić, Shervin Bagheri, Philipp Schlatter, and Dan S Henningson. Spectral analysis of nonlinear flows. *Journal of fluid mechanics*, 641:115–127, 2009.

[101] Peter J Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of fluid mechanics*, 656:5–28, 2010.

[102] Marko Budišić, Ryan Mohr, and Igor Mezić. Applied koopmanism. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(4):047510, 2012.

[103] Jonathan H Tu. *Dynamic mode decomposition: Theory and applications*. PhD thesis, Princeton University, 2013.

[104] Matthew O Williams, Ioannis G Kevrekidis, and Clarence W Rowley. A data–driven approximation of the koopman operator: Extending dynamic mode decomposition. *Journal of Nonlinear Science*, 25(6):1307–1346, 2015.

[105] Zlatko Drmac, Igor Mezic, and Ryan Mohr. Data driven modal decompositions: analysis and enhancements. *SIAM Journal on Scientific Computing*, 40(4):A2253–A2285, 2018.

[106] Milan Korda and Igor Mezić. On convergence of extended dynamic mode decomposition to the koopman operator. *Journal of Nonlinear Science*, 28(2):687–710, 2018.

[107] Enoch Yeung, Soumya Kundu, and Nathan Hodas. Learning deep neural network representations for koopman operators of nonlinear dynamical systems. In *2019 American Control Conference (ACC)*, pages 4832–4839. IEEE, 2019.

[108] Yiqiang Han, Wenjian Hao, and Umesh Vaidya. Deep learning of koopman representation for control. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pages 1890–1895. IEEE, 2020.

[109] Joseph Allanson Wattleworth. *PEAK-PERIOD CONTROL OF A FREEWAY SYSTEM-SOME THEORETICAL CONSIDERATIONS.* Northwestern University, 1963.

[110] Li Shin Yuan and John B Kreer. Adjustment of freeway ramp metering rates to balance entrance ramp queues. *Transportation Research/UK/*, 5, 1971.

[111] Jin J Wang and Adolf D May. Computer model for optimal freeway on-ramp control. *Highway Research Record*, 469:16–25, 1973.

[112] Y Iida, T Hasegawa, Y Asakura, and CF Shao. A formulation of on-ramp traffic control system with route guidance for urban expressway. *IFAC Proceedings Volumes*, 23(2):229–236, 1990.

[113] Albert Messmer and Markos Papageorgiou. Metanet: A macroscopic simulation program for motorway networks. *Traffic engineering & control*, 31(9), 1990.

[114] Markos Papageorgiou and Apostolos Kotsialos. Freeway ramp metering: An overview. *IEEE transactions on intelligent transportation systems*, 3(4):271–281, 2002.

[115] Markos Papageorgiou, Christina Diakaki, Vaya Dinopoulou, Apostolos Kotsialos, and Yibing Wang. Review of road traffic control strategies. *Proceedings of the IEEE*, 91(12):2043–2067, 2003.

[116] Apostolos Kotsialos and Markos Papageorgiou. Nonlinear optimal control applied to coordinated ramp metering. *IEEE Transactions on control systems technology*, 12(6):920–933, 2004.

[117] Agachai Sumalee, RX Zhong, TL Pan, and WY Szeto. Stochastic cell transmission model (sctm): A stochastic dynamic traffic model for traffic state surveillance and assignment. *Transportation Research Part B: Methodological*, 45(3):507–533, 2011.

[118] Andy HF Chow and Ying Li. Robust optimization of dynamic motorway traffic via ramp metering. *IEEE Transactions on Intelligent Transportation Systems*, 15(3):1374–1380, 2014.

[119] Alex A Kurzhanskiy and Pravin Varaiya. Guaranteed prediction and estimation of the state of a road network. *Transportation research part C: emerging technologies*, 21(1):163–180, 2012.

[120] Claudio Roncoli, Markos Papageorgiou, and Ioannis Papamichail. Traffic flow optimisation in presence of vehicle automation and communication systems–part i: A first-order multi-lane model for motorway traffic. *Transportation Research Part C: Emerging Technologies*, 57:241–259, 2015.

[121] Claudio Roncoli, Markos Papageorgiou, and Ioannis Papamichail. Traffic flow optimisation in presence of vehicle automation and communication systems–part ii: Optimal control for multi-lane motorways. *Transportation Research Part C: Emerging Technologies*, 57:260–275, 2015.

[122] Yu Han, Yufei Yuan, Andreas Hegyi, and Serge P Hoogendoorn. Linear quadratic mpc for integrated route guidance and ramp metering. In *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, pages 1150–1155. IEEE, 2015.

[123] Yu Han, Andreas Hegyi, Yufei Yuan, Serge Hoogendoorn, Markos Papageorgiou, and Claudio Roncoli. Resolving freeway jam waves by discrete first-order model-based predictive control of variable speed limits. *Transportation Research Part C: Emerging Technologies*, 77:405–420, 2017.

[124] Farzam Tajdari, Claudio Roncoli, and Markos Papageorgiou. Feedback-based ramp metering and lane-changing control with connected and automated vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[125] Anastasia Spiliopoulou, Diamantis Manolis, Foteini Vandorou, and Markos Papageorgiou. Adaptive cruise control operation for improved motorway traffic flow. *Transportation research record*, 2672(22):24–35, 2018.

[126] Maria Kontorinaki, Iasson Karafyllis, and Markos Papageorgiou. Local and coordinated ramp metering within the unifying framework of an adaptive control scheme. *Transportation Research Part A: Policy and Practice*, 128:89–113, 2019.

[127] Aharon Ben-Tal, Alexander Goryashko, Elana Guslitzer, and Arkadi Nemirovski. Adjustable robust solutions of uncertain linear programs. *Mathematical programming*, 99(2):351–376, 2004.

[128] Steve Zymler, Daniel Kuhn, and Berç Rustem. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, 137(1):167–198, 2013.

[129] Wenqing Chen, Melvyn Sim, Jie Sun, and Chung-Piaw Teo. From cvar to uncertainty set: Implications in joint chance-constrained optimization. *Operations research*, 58(2):470–485, 2010.

[130] Gunes Dervisoglu, Gabriel Gomes, Jaimyoung Kwon, Roberto Horowitz, and Pravin Varaiya. Automatic calibration of the fundamental diagram and empirical observations on capacity. In *Transportation Research Board 88th Annual Meeting*, volume 15, pages 31–59. Citeseer, 2009.

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

[132] Ying Li, Andy HF Chow, and Daniela Lichtler Cassel. Optimal control of motorways by ramp metering, variable speed limits, and hard-shoulder running. *Transportation Research Record*, 2470(1):122–130, 2014.

[133] Giuseppe Calafiore and Marco C Campi. Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102(1):25–46, 2005.

[134] Arkadi Nemirovski and Alexander Shapiro. Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17(4):969–996, 2007.

[135] Yufei Sun, Grace Aw, Ryan Loxton, and Kok Lay Teo. Chance-constrained optimization for pension fund portfolios in the presence of default risk. *European Journal of Operational Research*, 256(1):205–214, 2017.

[136] Grani A Hanasusanto, Vladimir Roitch, Daniel Kuhn, and Wolfram Wiesemann. A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Mathematical Programming*, 151(1):35–62, 2015.

[137] R Tyrrell Rockafellar, Stanislav Uryasev, et al. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000.

[138] R Tyrrell Rockafellar and Stanislav Uryasev. Conditional value-at-risk for general loss distributions. *Journal of banking & finance*, 26(7):1443–1471, 2002.

[139] Alexander Shapiro and Anton Kleywegt. Minimax analysis of stochastic problems. *Optimization Methods and Software*, 17(3):523–542, 2002.

[140] Jos F Sturm. Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones. *Optimization methods and software*, 11(1-4):625–653, 1999.

[141] Johan Lofberg. Yalmip: A toolbox for modeling and optimization in matlab. In *2004 IEEE international conference on robotics and automation (IEEE Cat. No. 04CH37508)*, pages 284–289. IEEE, 2004.

[142] Markos Papageorgiou, Habib Hadj-Salem, Jean-Marc Blosseville, et al. Alinea: A local feedback control law for on-ramp metering. *Transportation research record*, 1320(1):58–67, 1991.

[143] Yibing Wang, Elias B Kosmatopoulos, Markos Papageorgiou, and Ioannis Papamichail. Local ramp metering in the presence of a distant downstream bottleneck: Theoretical analysis and simulation study. *IEEE Transactions on Intelligent Transportation Systems*, 15(5):2024–2039, 2014.

[144] Yuheng Kan, Yibing Wang, Markos Papageorgiou, and Ioannis Papamichail. Local ramp metering with distant downstream bottlenecks: A comparative study. *Transportation Research Part C: Emerging Technologies*, 62:149–170, 2016.

[145] Ioannis Papamichail and Markos Papageorgiou. Traffic-responsive linked ramp-metering control. *IEEE Transactions on Intelligent Transportation Systems*, 9(1):111–121, 2008.

[146] Dongbin Zhao, Xuerui Bai, Fei-Yue Wang, Jing Xu, and Wensheng Yu. Dhp method for ramp metering of freeway traffic. *IEEE Transactions on Intelligent Transportation Systems*, 12(4):990–999, 2011.

[147] Nikolas Geroliminis, Anupam Srivastava, and Panos Michalopoulos. A dynamic-zone-based coordinated ramp-metering algorithm with queue constraints for minnesota's freeways. *IEEE Transactions on Intelligent Transportation Systems*, 12(4):1576–1586, 2011.

[148] Tom Bellemans, Bart De Schutter, and Bart De Moor. Anticipative model predictive control for ramp metering in freeway networks. In *Proceedings of the 2003 American Control Conference, 2003.*, volume 5, pages 4077–4082. IEEE, 2003.

[149] Apostolos Kotsialos, Markos Papageorgiou, Morgan Mangeas, and Habib Haj-Salem. Coordinated and integrated control of motorway networks via a non-linear optimal control. *Transportation Research Part C: Emerging Technologies*, 10(1):65–84, 2002.

[150] Ajith Muralidharan and Roberto Horowitz. Optimal control of freeway networks based on the link node cell transmission model. In *2012 American Control Conference (ACC)*, pages 5769–5774. IEEE, 2012.

[151] Giacomo Como, Enrico Lovisari, and Ketan Savla. Convexity and robustness of dynamic traffic assignment and freeway network control. *Transportation Research Part B: Methodological*, 91:446–465, 2016.

[152] Christian Rosdahl, Gustav Nilsson, and Giacomo Como. On distributed optimal control of traffic flows in transportation networks. In *2018 IEEE Conference on Control Technology and Applications (CCTA)*, pages 903–908. IEEE, 2018.

[153] Domenico Bianchi, Alessandro Borri, Maria Domenica Di Benedetto, and Antonella Ferrara. Decentralized model predictive control of freeway traffic systems over lossy communication networks. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pages 1074–1079. IEEE, 2019.

[154] RX Zhong, FF Yuan, TL Pan, Andy HF Chow, CJ Chen, and Zhongzhen Yang. Linear complementarity system approach to macroscopic freeway traffic modelling: uniqueness and convexity. *Transportmetrica A: Transport Science*, 12(2):142–174, 2016.

[155] Marius Schmitt and John Lygeros. An exact convex relaxation of the freeway network control problem with controlled merging junctions. *Transportation Research Part B: Methodological*, 114:1–25, 2018.

[156] Sergueï Konstantinovitch Godunov. A difference method for numerical calculation of discontinous equations of hydrodynamics. *Math. Sb*, 47(89):271–300, 1959.

[157] JP Lebacque. The godunov scheme and what it means for rst order tra c ow models. In *Proceedings of the 13th International Symposium on Transportation and Trafc Theory, Lyon, France, July*, volume 2426. Citeseer, 1996.

[158] Craig T Lawrence and André L Tits. A computationally efficient feasible sequential quadratic programming algorithm. *Siam Journal on optimization*, 11(4):1092–1118, 2001.

[159] Yu Han, Mohsen Ramezani, Andreas Hegyi, Yufei Yuan, and Serge Hoogendoorn. Hierarchical ramp metering in freeways: an aggregated modeling and control approach. *Transportation research part C: emerging technologies*, 110:1–19, 2020.

[160] Yihang Zhang, Isik Ilber Sirmatel, Faisal Alasiri, Petros A Ioannou, and Nikolas Geroliminis. Comparison of feedback linearization and model predictive techniques for variable speed limit control. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 3000–3005. IEEE, 2018.

[161] Eduardo F Camacho and Carlos Bordons Alba. *Model predictive control*. Springer science & business media, 2013.

[162] Andreas Hegyi, Bart De Schutter, and Hans Hellendoorn. Model predictive control for optimal coordination of ramp metering and variable speed limits. *Transportation Research Part C: Emerging Technologies*, 13(3):185–209, 2005.

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

[164] Gurobi Optimization et al. Gurobi optimizer reference manual, 2020. *URL http://www. gurobi. com*, 12, 2012.

[165] C Diakaki and M Papageorgiou. Design and simulation test of coordinated ramp metering control (metaline) for a10-west in amsterdam. *Dynamic Syst. Simulat. Lab., Tech. Univ. Crete, Chania, Greece, Internal Rep*, 2:1994, 1994.

[166] Rodrigo C Carlson, Ioannis Papamichail, Markos Papageorgiou, and Albert Messmer. Optimal motorway traffic flow control involving variable speed limits and ramp metering. *Transportation science*, 44(2):238–253, 2010.

[167] Xiao-Yun Lu and Steven E Shladover. Review of variable speed limits and advisories: Theory, algorithms, and practice. *Transportation research record*, 2423(1):15–23, 2014.

[168] Mark Robinson et al. Examples of variable speed limit applications: Speed management workshop. 2000.

[169] Rongjie Yu and Mohamed Abdel-Aty. An optimal variable speed limits system to ameliorate traffic safety risk. *Transportation research part C: emerging technologies*, 46:235–246, 2014.

[170] Solomon Kidane Zegeye, Bart De Schutter, J Hellendoorn, Ewald A Breunesse, and Andreas Hegyi. A predictive traffic controller for sustainable mobility using parameterized control policies. *IEEE Transactions on Intelligent Transportation Systems*, 13(3):1420–1429, 2012.

[171] Alfréd Csikós and Balázs Kulcsár. Variable speed limit design based on mode dependent cell transmission model. *Transportation Research Part C: Emerging Technologies*, 85:429–450, 2017.

[172] José Ramón D Frejo, Ioannis Papamichail, Markos Papageorgiou, and Bart De Schutter. Macroscopic modeling of variable speed limits on freeways. *Transportation research part C: emerging technologies*, 100:15–33, 2019.

[173] Yeqing Zhang, Meiling Wang, Xing Fang, and Umit Ozguner. Unifying analytical methods with numerical methods for traffic system modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 50(6):2068–2082, 2018.

[174] Ioannis Papamichail, Apostolos Kotsialos, Ioannis Margonis, and Markos Papageorgiou. Coordinated ramp metering for freeway networks–a model-predictive hierarchical control approach. *Transportation Research Part C: Emerging Technologies*, 18(3):311–331, 2010.

[175] Markos Papageorgiou, Elias Kosmatopoulos, and Ioannis Papamichail. Effects of variable speed limits on motorway traffic flow. *Transportation Research Record*, 2047(1):37–48, 2008.

[176] K Kampitaki. *Integrated control of traffic flow using ramp metering and variable speed limits*. PhD thesis, M. Sc. thesis, Technical University of Crete, Chania, Greece, 2008.

[177] Chuanye Gu, Changzhi Wu, Kok Lay Teo, Yonghong Wu, and Song Wang. A smoothing method for ramp metering. *IEEE Transactions on Intelligent Transportation Systems*, 23(8):13358–13371, 2022.

[178] Kok Lay Teo, C Goh, and K Wong. *A unified computational approach to optimal control problems*. 1991.

[179] Mofeng Yang, Zhibin Li, Zemian Ke, and Meng Li. A deep reinforcement learning-based ramp metering control framework for improving traffic operation at freeway weaving sections. In *Proceedings of the Transportation Research Board 98th Annual Meeting, Washington, DC, USA*, pages 13–17, 2019.

[180] Emmanouil Smaragdis and Markos Papageorgiou. Series of new local ramp metering strategies: Emmanouil smaragdis and markos papageorgiou. *Transportation Research Record*, 1856(1):74–86, 2003.

[181] Bing Liu, Yu Tang, Yuxiong Ji, Yu Shen, and Yuchuan Du. A deep reinforcement learning approach for ramp metering based on traffic video data. *Journal of Advanced Transportation*, 2021, 2020.

[182] Tom Bellemans, B De Schutter, and Bart De Moor. Model predictive control for ramp metering of motorway traffic: A case study. *Control Engineering Practice*, 14(7):757–767, 2006.

[183] Cecilia Pasquale, Simona Sacone, Silvia Siri, and Bart De Schutter. A multi-class model-based control scheme for reducing congestion and emissions in freeway networks by combining ramp metering and route guidance. *Transportation Research Part C: Emerging Technologies*, 80:384–408, 2017.

[184] Apostolos Kotsialos. A varying parameter multi-class second-order macroscopic traffic flow model for coordinated ramp metering with global and local environmental objectives. *Transportation Research Part C: Emerging Technologies*, 128:103106, 2021.

[185] Bin Ran, Peter J Jin, David Boyce, Tony Z Qiu, and Yang Cheng. Perspectives on future transportation research: Impact of intelligent transportation system technologies on next-generation transportation modeling. *Journal of Intelligent Transportation Systems*, 16(4):226–242, 2012.

[186] Ibrar Yaqoob, Ibrahim Abaker Targio Hashem, Abdullah Gani, Salimah Mokhtar, Ejaz Ahmed, Nor Badrul Anuar, and Athanasios V Vasilakos. Big data: From beginning to future. *International Journal of Information Management*, 36(6):1231–1247, 2016.

[187] Pengjun Zhao and Haoyu Hu. Geographical patterns of traffic congestion in growing megacities: Big data analytics from beijing. *Cities*, 92:164–174, 2019.

[188] Theodoros Alexakis, Nikolaos Peppes, Evgenia Adamopoulou, and Konstantinos Demestichas. An artificial intelligence-based approach for the controlled access ramp metering problem. *Vehicles*, 3(1):63–83, 2021.

[189] Yongqian Xiao, Xinglong Zhang, Xin Xu, Xueqing Liu, and Jiahang Liu. Deep neural networks with koopman operators for modeling and control of autonomous vehicles. *arXiv preprint arXiv:2007.02219*, 2020.

[190] Dominik Moser, Roman Schmied, Harald Waschl, and Luigi del Re. Flexible spacing adaptive cruise control using stochastic model predictive control. *IEEE Transactions on Control Systems Technology*, 26(1):114–127, 2017.

[191] Wan Li, Jingxing Wang, Rong Fan, Yiran Zhang, Qiangqiang Guo, Choudhury Siddique, and Xuegang Jeff Ban. Short-term traffic state prediction from latent structures: Accuracy vs. efficiency. *Transportation Research Part C: Emerging Technologies*, 111:72–90, 2020.

[192] Lucian Cristian Iacob, Gerben Izaak Beintema, Maarten Schoukens, and Roland Tóth. Deep identification of nonlinear systems in koopman form. *arXiv preprint arXiv:2110.02583*, 2021.

[193] Bernard O Koopman and J v Neumann. Dynamical systems of continuous spectra. *Proceedings of the National Academy of Sciences of the United States of America*, 18(3):255, 1932.

[194] Erik M Bollt and Naratip Santitissadeekorn. *Applied and computational measurable dynamics*. SIAM, 2013.

[195] Gary Froyland, Georg A Gottwald, and Andy Hammerlindl. A computational method to extract macroscopic variables and their dynamics in multiscale systems. *SIAM Journal on Applied Dynamical Systems*, 13(4):1816–1846, 2014.

[196] Alexandre Mauroy and Igor Mezić. On the use of fourier averages to compute the global isochrons of (quasi) periodic dynamics. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(3):033112, 2012.

[197] Alexandre Mauroy, Igor Mezić, and Jeff Moehlis. Isostables, isochrons, and koopman spectrum for the action–angle representation of stable fixed point dynamics. *Physica D: Nonlinear Phenomena*, 261:19–30, 2013.

[198] Tomas W Muld, Gunilla Efraimsson, and Dan S Henningson. Flow structures around a high-speed train extracted using proper orthogonal decomposition and dynamic mode decomposition. *Computers & Fluids*, 57:87–97, 2012.

[199] Kazi Redwan Shabab, Shakib Mustavee, Shaurya Agarwal, Mohamed H Zaki, and Sajal Das. Exploring dmd-type algorithms for modeling signalised intersections. *arXiv preprint arXiv:2107.06369*, 2021.

[200] Vít Cibulka, Tomáš Haniš, and Martin Hromčík. Data-driven identification of vehicle dynamics using koopman operator. In *2019 22nd International Conference on Process Control (PC19)*, pages 167–172. IEEE, 2019.

[201] Huiyu Chen, Fan Wu, Kaizhe Hou, and Tony Z Qiu. Backpressure-based distributed dynamic route control for connected and automated vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 2022.

[202] Neetesh Kumar, Sarthak Mittal, Vaibhav Garg, and Neeraj Kumar. Deep reinforcement learning-based traffic light scheduling framework for sdn-enabled smart transportation system. *IEEE Transactions on Intelligent Transportation Systems*, 23(3):2411–2421, 2021.

[203] Tianshu Chu, Jie Wang, Lara Codecà, and Zhaojian Li. Multi-agent deep reinforcement learning for large-scale traffic signal control. *IEEE Transactions on Intelligent Transportation Systems*, 21(3):1086–1095, 2019.

[204] Igor Mezić. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dynamics*, 41(1):309–325, 2005.

[205] Igor Mezić. Analysis of fluid flows via spectral properties of the koopman operator. *Annual Review of Fluid Mechanics*, 45:357–378, 2013.

[206] Milan Korda and Igor Mezić. Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control. *Automatica*, 93:149–160, 2018.

[207] Bengt Fornberg, Elisabeth Larsson, and Natasha Flyer. Stable computations with gaussian radial basis functions. *SIAM Journal on Scientific Computing*, 33(2):869–892, 2011.

[208] John A Hartigan and Manchek A Wong. Algorithm as 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)*, 28(1):100–108, 1979.

[209] Jonathan H Tu, Clarence W Rowley, Dirk M Luchtenburg, Steven L Brunton, and J Nathan Kutz. On dynamic mode decomposition: Theory and applications. *Journal of Computational Dynamics*, 1(2):391–421, 2014.

[210] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256. JMLR Workshop and Conference Proceedings, 2010.