# Stochastic Traffic Flow Modeling and Optimal Congestion Pricing

by

Li Yang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Industrial and Operations Engineering)
in The University of Michigan
2012

Doctoral Committee:

Professor Romesh Saigal, Chair
Professor Judy Jin
Assistant Professor Gabor Orosz
Assistant Professor Siqian Shen

To my parents and my wife, Rui Mu

For their love, support and encouragement

# ACKNOWLEDGEMENTS

I would like to sincerely thank my advisor, Professor Romesh Saigal, for sharing his vast knowledge and vision in the past five years of my study, as well as his continuous guidance and encouragement throughout the creation of this dissertation. This dissertation could not be finished without his support and encouragement. I also would like to thank Professor Judy Jin, Professor Gabor Orosz and Professor Siqian Shen for their careful review, support and suggestions of this dissertation.

My gratitude also goes to all my fellow graduate students and the staff members in the Department of Industrial and Operations Engineering. They have provided an open and friendly environment where I could enjoy every moment of my PhD study. I appreciate their generous support and great inspiration and I have learnt and benefited significantly from each individual of them.

Last, but not least, I would like to express my hearty gratitude to my wife, Rui Mu, for her unconditional love and support that sustained me through this critical stage of life. I also would like to thank my parents for their faith, love and concern. I owe my gratitude to all my family members and friends who have made this dissertation possible.

# TABLE OF CONTENTS

**DEDICATION** . . . . . ii

**ACKNOWLEDGEMENTS** . . . . . iii

**LIST OF FIGURES** . . . . . vii

**LIST OF TABLES** . . . . . ix

**ABSTRACT** . . . . . x

**CHAPTER**

**I. Introduction** . . . . . 1

1.1 Motivation . . . . . 1
1.2 Research Background and Literature Review . . . . . 4
1.2.1 Traffic Flow Modeling and Prediction . . . . . 4
1.2.2 On-line Traffic State Estimation . . . . . 6
1.2.3 Optimal Congestion Pricing . . . . . 7
1.3 Research Objectives . . . . . 8
1.4 Organization of the Dissertation . . . . . 10

**II. Stochastic Traffic Flow Modeling** . . . . . 11

2.1 Introduction . . . . . 11
2.2 General Macroscopic Traffic Flow Models . . . . . 13
2.2.1 The Fundamental Flow Relationship . . . . . 14
2.2.2 The Conservation Law in Traffic Flow Models . . . 15
2.2.3 The Speed-Density Relationship . . . . . 16
2.3 The Stochastic Traffic Flow Modeling . . . . . 18
2.3.1 The Model . . . . . 18
2.3.2 Godunov's Scheme . . . . . 22
2.3.3 Simulation-Based Traffic Prediction . . . . . 30
2.4 Case Study . . . . . 32

2.4.1 Speed-Density Function Calibration . . . . . 33
2.4.2 Forcing Function Calibration . . . . . 35
2.4.3 Comparison of the Stochastic Model with the Deterministic Model . . . . . 40
2.5 Conclusion . . . . . 42

**III. Stochastic Filtering For Real Time Traffic Prediction** . . . . . 43

3.1 Introduction . . . . . 43
3.2 Problem Formulation . . . . . 46
3.3 State Space Models and Filtering . . . . . 48
3.3.1 State Space Representation . . . . . 48
3.3.2 Kalman Filter . . . . . 48
3.3.3 Extended Kalman Filter . . . . . 50
3.3.4 Unscented Kalman Filter . . . . . 51
3.3.5 Particle Filter . . . . . 53
3.4 State Space Formalization of the Traffic Flow Model . . . . . 56
3.4.1 State Transition Equation . . . . . 56
3.4.2 Measurement Equation . . . . . 58
3.5 Online Calibration of the Traffic Flow Model . . . . . 59
3.5.1 Joint Estimation . . . . . 60
3.5.2 Dual Estimation . . . . . 61
3.6 Numerical Results . . . . . 62
3.6.1 Traffic State Estimation . . . . . 62
3.6.2 Traffic State and Model Parameters Estimation . . . 66
3.6.3 Filtering Based Traffic Prediction . . . . . 69
3.7 Conclusion . . . . . 71

**IV. Dynamic Pricing For Managed Toll Lanes** . . . . . 74

4.1 Introduction . . . . . 74
4.2 Distance-Based Dynamic Pricing Model . . . . . 78
4.2.1 Infrastructure . . . . . 78
4.2.2 The Underlying Traffic Flow Model on the General Lane . . . . . 79
4.2.3 Demand Function . . . . . 83
4.2.4 Flow Conservation on the Toll Lane . . . . . 86
4.2.5 The Constraint . . . . . 88
4.2.6 Complete Model Formulation . . . . . 89
4.3 Analysis of the Stochastic Control Model . . . . . 90
4.3.1 Complexity Analysis . . . . . 90
4.3.2 Solution . . . . . 92
4.4 Numerical Case Study . . . . . 101
4.5 Conclusion . . . . . 106

**V. Conclusions and Future Work** . . . . . 109

5.1 Conclusions and Summary of Contributions . . . . . 109
5.2 Future Work . . . . . 111

**BIBLIOGRAPHY** . . . . . 113

# LIST OF FIGURES

**Figure**


| | | |
|---|---|---|
| 1.1 | Scheme of the research objective | 9 |
| 2.1 | Speed and flux functions for Piecewise speed-density function | 18 |
| 2.2 | The shock wave | 24 |
| 2.3 | The rarefaction wave | 26 |
| 2.4 | Godunov's scheme for homogeneous conservation law | 27 |
| 2.5 | Godunov decomposition | 30 |
| 2.6 | The map of the highway | 33 |
| 2.7 | Regression results of the Piecewise speed-density function | 36 |
| 2.8 | Calibrated parameters of the Piecewise speed-density function | 37 |
| 2.9 | Regression results of the forcing function at station 3 between 7:15am and 7:19am | 39 |
| 2.10 | Evolution of the forcing function parameters over the day at station 2 | 40 |
| 2.11 | Comparison of RMSE between the deterministic model and the stochastic model at all stations | 41 |
| 3.1 | Evolution of the forcing function parameters | 64 |
| 3.2 | The result of traffic state estimation given known model parameters | 65 |
| 3.3 | RMSE comparison of the density estimation given known model parameters | 66 |

3.4 Evolution of true and estimated traffic state in cells . . . . . 68

3.5 RMSE comparison of the density estimation given unknown model parameters . . . . . 69

3.6 Evolution of true and estimated model parameters . . . . . 70

3.7 Evolution of true and predicted traffic state in cells . . . . . 72

3.8 RMSE comparison of traffic state prediction for different filtering algorithms . . . . . 73

4.1 General lane and managed toll lane . . . . . 78

4.2 Decomposition of dynamic toll pricing problem . . . . . 93

4.3 Parameters in the stochastic traffic flow model used for simulation. . 103

4.4 Density evolution on the general lane . . . . . 104

4.5 Speed evolution on the general lane . . . . . 105

4.6 Optimal toll prices on each individual toll entrance . . . . . 106

4.7 Volume evolution on the toll lane . . . . . 107

4.8 Cumulative revenue collected by each individual toll entrance . . . . 108

# LIST OF TABLES

**Table**


2.1 Speed-density functions . . . . . 17

2.2 Fitting errors of speed-density functions . . . . . 35

3.1 Parameters in the experiment for traffic state estimation. . . . . 63

3.2 Parameters in the experiment for traffic state and model parameters estimation. . . . . 67

4.1 Overview of the existing HOT facilities in the United States . . . . 76

4.2 Parameters in the numerical experiment . . . . . 104

# ABSTRACT

Stochastic Traffic Flow Modeling and Optimal Congestion Pricing

by

Li Yang

Chair: Romesh Saigal

Congestion in surface transportation networks causes serious economical and environmental problems in urban areas of the United States and around the world. Due to the continuous deterioration of urban traffic conditions in recent years, increasing numbers of intelligent transportation system applications have been developed and implemented to help administrators manage the highway and make the usage of the whole traffic network more efficient. The success of most intelligent transportation system applications is dependent on an accurate prediction of the future traffic state, because such accurate prediction can help the decision maker choose the right strategy and/or provide reliable travel information to the drivers. This dissertation explores an innovative stochastic traffic flow model to better predict the future traffic state on the highway, and provides a framework to investigate some potential applications, including dynamic congestion pricing.

Part I presents an innovative macroscopic stochastic traffic flow model and the off-line calibration algorithm for this model. It also develops the numerical algorithm for future traffic state prediction based on this model. The model is validated by using real highway data. The empirical results show that this model outperforms, in

terms of prediction accuracy, the traditional macroscopic traffic flow model.

Part II is devoted to the development of on-line parameter calibration and traffic state estimation algorithms for the proposed innovative stochastic traffic flow model. The algorithms are tested on synthetic data, and the numerical results show that the algorithms are able to capture the change of model parameters and improve the prediction accuracy of the traffic flow model.

Part III formulates a mathematical model for the problem of optimal distance-based dynamic congestion pricing. This mathematical model utilizes the proposed stochastic traffic flow model as the underlying dynamics of the traffic flow. It develops the optimal dynamic pricing strategy under a specified objective and a numerical case study is presented to illustrate the solution.

# CHAPTER I

# Introduction

## 1.1 Motivation

Congestion in surface transportation networks causes serious economical and environmental problems in urban areas of the United States and around the world. For instance, it has been estimated that 32% of the daily travel in major US urban areas occurred under congested traffic conditions (*FHWA*, 2008). Also *Schrank and Lomax* (2009) showed that congestion caused urban Americans to travel extra 4.2 billion hours and to purchase additional 2.8 billion gallons of fuel thus incurring the congestion cost of 87.2 billion dollars annually, 50% more than the cost incurred a decade ago.

Due to the continuous deterioration of urban traffic conditions, increasing numbers of Intelligent Transportation System (ITS) applications have been developed and implemented to help administrators manage the highway and make the usage of the whole traffic network more efficient. However these applications do not currently take full advantage of available traffic flow data. For instance, while vehicle navigation models can factor in current conditions when making routing recommendations, they do not typically consider conditions that may actually be experienced by a driver once he reaches a particular roadway link. This results in navigation systems that cannot truly be used to mitigate congestion, as the route recommendation may often

only move the congestion from one location to another if too many drivers opt for the same route. This may result in not only potential increase in delays experienced by the travelers but also more frequent stop and go and reduced fuel efficiency.

The success of most ITS applications is dependent on an accurate prediction of the future traffic state. Such accurate prediction can help the decision maker choose the right strategy and/or provide reliable travel information to the drivers. For example, in traffic flow control applications, such as ramp metering and congestion pricing, if future traffic state can be precisely predicted, the utilization of available highway capacity will be improved by letting an appropriate number of vehicles enter the highway. Travel time estimation is another example of the value of traffic state prediction. Based on the prediction of future traffic evolution, the travel time can be calculated in real time. This information can assist the drivers make right decisions on route choice and departure time. Therefore traffic state prediction has connections to ITS applications and we expect the prediction accuracy to have a significant impact on the performance of these applications.

While there has been extensive research on the development of traffic prediction algorithms, those developed so far primarily seek to find network equilibrium by factoring link flow capacities and resulting travel times. Most of these algorithms cannot be applied in real-time without drastically increasing the underlying size of the problem, thus making the formulation computationally intractable.

Nowadays, with the extensive usage of the detection and surveillance devices on the roads, a massive amount of traffic data has been collected and become available for research purposes. In many cities in the United States, the administrators of the transportation department have built such databases and the databases can provide historical and real time traffic data to the public. These databases provide researchers an opportunity to build and validate models for the purpose of traffic state prediction.

In this dissertation, we propose to develop and study a stochastic macroscopic

traffic flow model that can be used to predict the near future traffic conditions in real-time setting. This represents a paradigm change in the state of the art, where most macroscopic traffic models considered are deterministic and do not have the ability to dynamically calibrate its parameters in real-time. The stochasticity of the model allows the consideration of the inherent stochastic nature of driver behavior, different vehicle mixes, the effects of weather conditions, and various other effects at random locations and time. In addition, the ability of real-time parameter calibration will make the model more robust to varying underlying conditions. The understanding of the congestion phenomenon and its impact on the individual vehicle performance and system operations can be further analyzed.

One potential application of the stochastic traffic flow model is the dynamic pricing of tolls for a managed toll lane which incorporates the future traffic state prediction. Managed lanes are defined as the highway lanes that are operated under fixed or real-time dynamic strategies to achieve a variety of objectives. The objectives include but not limited to improving facility utilization efficiency, controlling congestion levels and increasing the return on investment. Strategies for managing the lanes can be roughly classified into three categories: eligibility control, access control and pricing. A managed lane implementing pricing strategy is also called a managed toll lane. In the United States, increasing investments are being made in managed toll lanes because of their obvious effectiveness in maintaining the level of service on the road and their revenue generating ability, which helps pay off the investors. For example, a High Occupancy Toll (HOT) lane is in operation in Alameda County, California along a 14 mile stretch of southbound I680. The toll is dynamic, and it is $1.00 between 5 am and 6:45 am; at approximately 6:45 am, it is increased to $1.75 until 11 am when it drops down to $0.30 for the rest of the day. Vehicles with two or more occupants can enter the lane without paying toll. In this dissertation, we will formulate a general mathematical model for dynamic toll pricing problem and investigate the optimal

pricing strategy based on distance traveled and the time saved.

In conclusion, there are many ITS applications, whose performance can be improved if the stochastic model can successfully predict the future traffic condition. This dissertation will investigate such a traffic flow model and its application on optimal dynamic congestion pricing.

## 1.2 Research Background and Literature Review

This section conducts a comprehensive review of related literature on traffic flow modeling, on-line traffic state estimation and optimal congestion pricing for managed toll lanes.

### 1.2.1 Traffic Flow Modeling and Prediction

Some purely data-driven models have already been developed to improve the traffic state prediction accuracy. For example, a hybrid method combining Kohonen maps with ARIMA time series models is developed to forecast short-term traffic flow by *Van Der Voort et al.* (1996). A multivariate time series state-space model using highway loop detector data has also been investigated and it suggests that different model specifications should be set for different time periods of the day (*Stathopoulos and Karlaftis*, 2003). Other than time series models, neural network is also used for traffic prediction. It offers an attractive alternative because neural network could model undefined and complex nonlinear surface (*Smith and Demetsky*, 1994). A framework of traffic prediction model combining wavelet transform, neural network and fuzzy logic has been developed and the wavelet transform is found useful to eliminate noise caused by random travel conditions (*Xiao et al.*, 2003).

These purely data-driven models do not rely on any knowledge of transportation. Sometimes these data-driven models can work very well by capturing the patterns from the historical data. However due to the lack of the understanding on how the

traffic flow evolves, these models are not reliable. There is another category of models, which are built exclusively for traffic state prediction. These models investigate and simulate the real traffic evolution, so that the prediction can be made by projecting the traffic evolution according to the traffic flow model. The key of such prediction models is the underlying traffic flow model. The prediction is obtained by projecting the traffic evolution as determined by the traffic flow model. According to the level of detail with which they simulate the traffic evolution, these traffic flow models can be classified into three categories: microscopic models, mesoscopic models and macroscopic models.

Microscopic models study the driver's behavior and the interaction between individual vehicles at a very high level of detail. This approach models the movement of individual vehicles based on assumptions about the drivers' driving behavior. Researchers have developed some popular simulator modules. For example, *Yang and Koutsopoulos* (1996) developed a microscopic traffic simulator (MITSIM) for modeling traffic networks with advanced traffic control, route guidance and surveillance system. It simulates individual vehicles by using models capturing car following, lane changing behavior and drivers' route choice decision. Another simulator called Corridor traffic simulator (CORSIM) is developed and maintained by FHWA (*Halati et al.*, 1997), it is one of the most frequently used software packages for the analysis of traffic. Parallel microscopic simulation (PARAMICS) is a super parallel computing simulator developed by Quadstone Limited, a Scottish company (*Cameron and Duncan*, 1996). It is a stochastic simulation model, which is very comprehensive and powerful for a wide variety of traffic modeling applications. VISSIM, developed by a German company Planung Transport Verkehr (PTV), is one of the most sophisticated simulators; It provides significant improvements in driver behavior and incorporates multi-modal transit operation models (*Choa et al.*, 2004). The simulators introduced above are the most commonly used software packages for microscopic traffic flow simulation.

Mesoscopic models describe the traffic evolution at a medium level of detail. In mesoscopic models, the vehicle behavior is modeled in terms of a probability distribution, which is more aggregate than that of microscopic simulation models. Three well-know mesoscopic models are the so called headway distribution models, cluster models and gas-kinetic continuum models (*Hoogendoorn and Bovy*, 2001).

Macroscopic models describe the traffic evolution from a high level of aggregation as it focuses on aggregated system variables, such as density and volume, rather than individual vehicle behavior. LWR model is a classical macroscopic model, which was proposed by *Lighthill and Whitham* (1955) and *Richards* (1956). It is a first-order deterministic model. In 1971, *Payne* (1971) developed a second-order model, which overcomes some deficiencies of first-order models and improves the model accuracy. In 1994, *Daganzo* (1994)proposed a discretized model, which is also called cell transmission model. The traffic flow evolution is explained from a different perspective in terms of sending and receiving function in every cell. These three models are the most commonly used macroscopic models. To the best of the author's knowledge, much work has been done on deterministic macroscopic traffic models, however only few works have considered stochastic macroscopic traffic models.

### 1.2.2 On-line Traffic State Estimation

Research on traffic state estimation has been studied extensively in recent years. Kalman filter is found to be an effective method for on-line traffic state estimation and various forms of Kalman filter have been widely used in literature. *Wang and Papageorgiou* (2005) proposed a general approach for real-time traffic state estimation by using extended Kalman filter (EKF). In their approach, the unknown model parameters in the traffic flow model are also considered as state variables so that they can be estimated on-line simultaneously. In this way, the off-line calibration of model parameters can be avoided and it is more adaptive to capture the change of parame-

ters. The biggest advantage of EKF is that its computation cost is lower than many other kinds of Kalman filter algorithms. However EKF requires a linearization of the system, which is sometimes impossible for highly complicated nonlinear traffic flow models like the stochastic model developed in Chapter II. *Sun et al.* (2003) developed a sequential Monte Carlo based Kalman filter, which is called mixture Kalman filter, to estimate the traffic density at unmeasured locations. The underlying traffic flow model is a cell transmission model based on switching state-space model. The algorithm is efficient in a way that it reaches a comparable performance to EKF with a small number of sample sequences. *Hegyi et al.* (2006) applies unscented Kalman filter (UKF) for traffic state estimation. UKF was proposed by *Julier and Uhlmann* (2004). Instead of calculating the Jacobians and Hessians to approximate the mean and covariance in EKF. UKF applied the so-called sigma points to obtain the mean and covariance. UKF is easier to implement and more accurate than EKF. *Mihaylova et al.* (2007) compared the performance of UKF with the so-called particle filter (PF), and compared the accuracy and complexity based on a cell transmission model. The results showed that the PF provides better performance than UKF. Both of them are suitable for real-time traffic estimation, because both are easy to implement and do not require linearization.

### 1.2.3 Optimal Congestion Pricing

Distance-based dynamic pricing is a reasonable and effective pricing strategy, while it is also the most complicated. Although some researchers and practitioners have investigated this pricing strategy in recent years, there is no unified scheme for this problem. *Yang et al.* (2010) investigated the pricing scheme for road network from the perspective of congestion control. They propose an iterative algorithm, which can adjust the toll price based on the observation of link flows over the network without knowing the travel time and demand functions and users' value of time. *Lou et al.*

(2011) also propose a self-learning approach to determine the pricing strategy for single toll station on toll lanes, and applies simulation as the underlying traffic flow model for determining the prices that optimize an objective. *Zhang et al.* (2008) propose a feedback-based simulation algorithm to dynamically adjust the toll lane prices in order to realize the optimal traffic allocation for the overall infrastructure efficiency. The above works use microscopic simulation models in their investigations, however there are disadvantages in applying microscopic simulation model for optimal dynamic pricing; An important one is that the simulation models are more difficult to calibrate than macroscopic traffic flow models.

This dissertation will formulate the dynamic pricing problem based on a stochastic macroscopic traffic flow model. An advantage of using a macroscopic model (over a microscopic simulation model) is that it can be calibrated on-line by applying the traffic state estimation algorithm developed in chapter III. Following the formulation, we develop a methodology to obtain an optimal pricing strategy to maximize the total expected revenue. Although the objective discussed in this study is to maximize the revenue, the general model formulation and solution can be easily adapted to any other objective, like the maximizing of the total throughput, etc.

## 1.3 Research Objectives

The objective of this research is to develop a stochastic macroscopic traffic flow model, which makes a more accurate prediction of the future traffic state than other traditional deterministic macroscopic models. The on-line traffic state and model parameter estimation algorithms for this macroscopic traffic flow model will be studied as well. This research will also investigate the application of this traffic flow model in an ITS application: the optimal dynamic congestion pricing problem which considers the travel time as well as distance traveled in the pricing.

Figure 1.1 illustrates the scheme of the research plan in this dissertation. Based

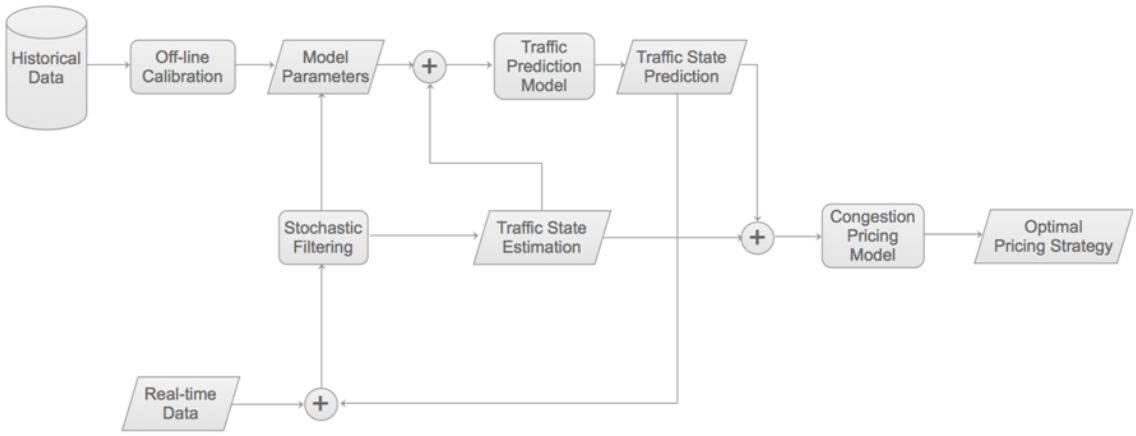



Figure 1.1: Scheme of the research objective

on Figure 1.1, the whole research objective can be decomposed into several main research tasks as described below:

1. To propose a stochastic macroscopic traffic flow model which captures the real-life traffic phenomenons.

2. To develop an off-line model calibration algorithm, which is able to estimate the model parameters using historical highway data.

3. To develop a numerical procedure for future traffic state prediction based on proposed traffic flow model and calibrated parameters.

4. To develop an on-line model calibration algorithm, which is able to update the model parameters and current traffic state in real time.

5. To formulate a mathematical model for the problem of optimal congestion pricing and to explore the optimal pricing strategy under specified objective.

## 1.4 Organization of the Dissertation

This dissertation is presented in a multiple manuscript format. It consists of three main chapters. Each of Chapters II, III and IV is written as an individual research paper focusing on a specific research topic, however these three research topics are closely connected. The organization of the dissertation follows below.

Chapter II presents a new macroscopic stochastic traffic flow model and the off-line calibration algorithm for this model. The numerical algorithm of traffic state prediction based on this model is also developed in this chapter. The model is validated by using real highway data. The empirical results show that this model outperforms, in terms of prediction accuracy, the traditional macroscopic traffic flow model.

Chapter III is devoted to the development of on-line parameter calibration and traffic state estimation algorithms for the traffic flow model proposed in Chapter II. The algorithms are tested on synthetic highway data, and the numerical results show that the algorithms are able to capture the change of model parameters and improve the prediction accuracy of the traffic flow model.

Chapter IV formulates a mathematical model for the problem of optimal distance-based dynamic congestion pricing. This mathematical model utilizes the proposed stochastic traffic flow model as the underlying dynamics of the traffic flow. It develops the optimal dynamic pricing strategy under a specified objective and a numerical case study is presented to illustrate the solution.

Finally Chapter V summarizes the conclusions and the contributions of this dissertation. Several further research topics are also proposed and discussed in this chapter.

# CHAPTER II

# Stochastic Traffic Flow Modeling

## 2.1 Introduction

Just as the normal blood circulation necessitates a healthy body, the smooth traffic flow is necessary for healthy business and community development in a city and a region. Traffic congestion haunts cities and communities from various perspectives: It inflicts uncertainties, drains resources, reduces productivity, stresses commuters, and harms environment. Due to the continuous deterioration of urban traffic conditions, increasing numbers of Intelligent Transportation System (ITS) applications have been developed and implemented to help administrators manage the highway and make the usage of the whole traffic network more efficient. For example, ramp metering system, a device usually consisting of a traffic light together with a signal controller installed on the ramp, has been implemented on many U.S. highways to control the rate of traffic entering the highway from the ramp. It was first implemented on Interstate 290 in Chicago in 1963 and has been deployed in urban highways of almost all major US cities since then. Ramp metering has proved to be effective in reducing travel time and delays, increasing highway throughput and improving safety. In 2000, a $650,000 experiment was conducted in Minneapolis-St. Paul area highways to test the effectiveness of ramp meters. In the experiment, 433 ramp meters were shut down for eight weeks. By comparing the performance statistics during these eight weeks

with earlier ones, it concluded that when the ramp meters were shut down highway capacity decreased by 9%, travel times increased by 22%, highway speeds dropped by 7% and crashes increased by 26%. This conclusion is also supported by a survey made by Federal Highway Administration (FHWA) on seven ramp metering system in the U.S. and Canada (*Arnold*, 1998). Many other ITS applications have also been implemented on highways, including congestion pricing, travel time estimation and variable speed limits. These ITS applications are intended to make the highway run smoothly and efficiently.

The success of most of the ITS applications are dependent on an accurate prediction of the future traffic state. Such accurate prediction can help the decision maker choose the right strategy and/or provide reliable travel information to the drivers. For example, in traffic flow control applications, such as ramp metering and congestion pricing, if future traffic state can be precisely predicted, the utilization of available highway capacity will be improved by letting an appropriate number of vehicles enter the highway. Travel time estimation is another example of the value of traffic state prediction. Based on the prediction of future traffic evolution, the travel time could be calculated in real time. This information can assist the drivers make right decisions on route choice and departure time. Basically, traffic state prediction has connections to ITS applications and the prediction accuracy usually has significant impact on the performance of these ITS applications. Nowadays, with the extensive use of the detection and surveillance devices on the road, a massive amount of traffic data has been collected and has become available for research purposes. In many cities in the United States, the administrators of the transportation department have built such databases and these can provide historical and real time traffic data to the public. These databases provide researchers an opportunity to build and validate models for the purpose of traffic state prediction.

In this chapter, we build a new macroscopic traffic flow model and validate its

prediction power by comparing its performance with other prediction models. There are two innovations in this traffic flow model: First, the model introduces Brownian Sheet (*Walsh*, 1986) to capture the stochastic characteristic of traffic evolution, so that the model becomes more realistic. Second, the model considers the impact of on-ramp inflow and off-ram outflow to the traffic evolution. It incorporates them into the model and can dynamically calibrate the inflow and outflow, so that the prediction can be adaptively adjusted.

The outline of this chapter is described as follows. In Section 2.2, we will first introduce the general macroscopic traffic flow models and enumerate some examples of macroscopic models such as LWR model and Payne-Whitham(PW) model. In Section 2.3, our new traffic flow model will be presented in detail, including the methodology of model calibration and the algorithms used for traffic state prediction. In Section 2.4, we present a case study using real highway traffic data to evaluate the prediction accuracy of this new traffic flow model. The performance is compared with other common prediction models. And Section 2.5 concludes this chapter.

## 2.2 General Macroscopic Traffic Flow Models

Since *Lighthill and Whitham* (1955) and *Richards* (1956) proposed the LWR model in 1950s, it has become the building block of many macroscopic traffic flow models, such as PW model (*Payne*, 1971; *Whitham*, 1974) and Zhang's model (*Zhang*, 1998, 2000). Generally in every macroscopic traffic flow model, three primary aggregated variables are used to describe the traffic: flux, density and average speed. The definitions of these variables are given in the notation description below:

- $(x, t)$ is the space and time pair for location $x$ and time $t$.
- $q(x, t)$ is the *flux* at location $x$ and time $t$, which is defined as the number of vehicles passing through the location in a unit of time.

- $\rho(x,t)$ is the *density* at location $x$ and time $t$, which is defined as the number of vehicles in a unit distance.

- $v(x,t)$ is the *average speed* of the vehicles at location $x$ and time $t$.

For simplicity, we may use $q$, $\rho$ and $v$ in subsequent discussion of this dissertation with the understanding that these quantities are dependent on $x$ and $t$.

### 2.2.1 The Fundamental Flow Relationship

The three aggregated variables are not independent. There exists a fundamental relationship, shown in equation (2.1), which connects them.

$$q = \rho \cdot v \tag{2.1}$$

The justification of the fundamental relationship is as the follows. Let $\phi(x,t,u)$ denote the density of vehicles with speed $u$ at location $x$ and time $t$, so $\int_{u_1}^{u_2} \phi(x,t,u)du$ represents the number of vehicles with speed between $u_1$ and $u_2$ per unit distance. According to the definition of $\rho$ and $v$, obviously

$$\rho(x,t) = \int_0^\infty \phi(x,t,u)du \tag{2.2}$$

and

$$v(x,t) = \frac{\int_0^\infty \phi(x,t,u)udu}{\int_0^\infty \phi(x,t,u)du} \tag{2.3}$$

The number of vehicles passing through location $x$ with speed between $u$ and $u+du$ during the next short time interval $dt$ is equal to the number of vehicles with speed between $u$ and $u+du$ in the short segment $u \cdot dt$ right before $x$. Since $\phi(x,t,u)$

is the density of vehicles with speed $u$, the total volume passing through $x$ would be:

$$q(x,t) = \frac{\int_0^\infty \phi(x,t,u)udtdu}{dt} = \int_0^\infty \phi(x,t,u)udu \tag{2.4}$$

Therefore the fundamental relationship is obtained by combining equations (2.2), (2.3) and (2.4). Given the measurement of any two among the three variables, the value of the third one can be derived according to the fundamental relationship. Since the average speed and volume are relatively easier to measure than the density, today's traffic surveillance systems measure these two. Therefore in almost all the traffic databases including the database we used for the model validation in this chapter, volume and average speed are available at the locations where the sensors are installed, and the density data is not explicitly available. However, using the fundamental relationship, density data could be inferred from the volume and average speed data.

### 2.2.2 The Conservation Law in Traffic Flow Models

In macroscopic traffic flow models, another important component is the conservation law as described in equation (2.5).

$$\rho_t + q_x = 0 \tag{2.5}$$

The justification of the conservation law in equation (2.5) is as the follows: Consider the segment between $x_1$ and $x_2$. If there is no on-ramp or off-ramp in this segment, the increment of the number of vehicles in this segment from $t_1$ to $t_2$ is simply due to the difference between the upstream inflow and downstream outflow.

Therefore:

$$\int_{x_1}^{x_2} \rho(x, t_2)dx - \int_{x_1}^{x_2} \rho(x, t_1)dx = \int_{t_1}^{t_2} q(x_1, t)dt - \int_{t_1}^{t_2} q(x_2, t)dt \tag{2.6}$$

The above equation could be rewritten (using to the fundamental theorem of Calculus) as:

$$\int_{x_1}^{x_2}\int_{t_1}^{t_2} \rho_t dtdx = -\int_{t_1}^{t_2}\int_{x_1}^{x_2} q_x dxdt \tag{2.7}$$

By moving the right hand side to the left, we can get the following equation:

$$\int_{x_1}^{x_2}\int_{t_1}^{t_2} (\rho_t + q_x)dtdx = 0 \tag{2.8}$$

The above equation must be satisfied for any choice of $x_1$, $x_2$, $t_1$ and $t_2$, which gives the conservation equation (2.5).

### 2.2.3 The Speed-Density Relationship

The speed-density relationship, together with the fundamental relationship and conservation law, form the integrated macroscopic traffic model. There are two types of speed-density relationship: One is an equilibrium model and the other one is non-equilibrium.

In equilibrium models, it is assumed that the equilibrium speed is a function of the density and the average speed can reach the equilibrium speed instantly no matter how fast the density changes. LWR model is an equilibrium model which is described in equation (2.9), where $V_*(\rho)$ is the equilibrium speed function.

$$\rho_t + q_x = 0$$

$$
\begin{aligned}
q &= \rho \cdot v \\
v &= V_*(\rho)
\end{aligned}
\tag{2.9}
$$

Traffic researchers have long been interested in functionally estimating the speed-density relationship. Several functional forms of the equilibrium speed function have been proposed in the literature. Table 2.1 summarizes the functions that have been frequently used in the literature. Greenshield suggested a linear speed-density relation for free flow and forced-flow conditions: $v = v_f(1-\rho/\rho_j)$, where $\rho_j$ is a parameter meaning the upper limit of the density. This linear model is the simplest model however not a very good fit with the field data. Greenberg proposed a logarithmic form for speed versus density: $v = v_0 \ln(\rho_j/\rho)$. Underwood used an exponential form: $v = v_f \exp(-\rho/\rho_j)$. The fourth function, $v = \min\{v_f, \alpha\rho^m\}$ $(m < 0)$, called the Piecewise model, is a new form proposed in this dissertation. In this function, the speed is constantly equal to the free flow speed for when density is low and exponentially decreasing for high density. The function will be justified by real data later in this chapter. Figure 2.1 plots the speed function and flux function for Piecewise speed-density function. It gives an idea how the speed and flux changes over the density.

Table 2.1: Speed-density functions

| Functions | $V_*(\rho)$ | $Q(\rho) = \rho \cdot V_*(\rho)$ |
|---|---|---|
| Greenshields | $v_f(1 - \rho/\rho_j)$ | $\rho v_f(1 - \rho/\rho_j)$ |
| Greenberg | $v_0 \ln(\rho_j/\rho)$ | $\rho v_0 \ln(\rho_j/\rho)$ |
| Underwood | $v_f \exp(-\rho/\rho_j)$ | $\rho v_f \exp(-\rho/\rho_j)$ |
| Piecewise | $\min\{v_f, \alpha\rho^m\}$ | $\rho \min\{v_f, \alpha\rho^m\}$ |

The Payne-Whitham (PW) model developed by *Payne* (1971) and *Whitham* (1974) is a non-equilibrium model described in equation (2.10). In non-equilibrium models, there is time decay for the average speed to reach the equilibrium speed. A partial differential equation in (2.10) describes how the average speed evolves towards to the

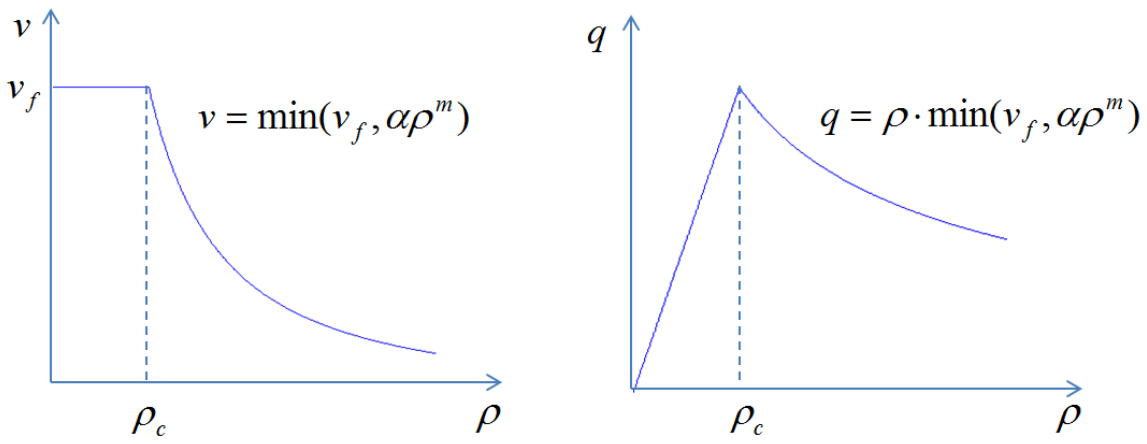



Figure 2.1: Speed and flux functions for Piecewise speed-density function

equilibrium speed over time, where $V_*(\rho)$ is the equilibrium speed function, $c_0 > 0$ and $\tau > 0$ are constant parameters. $\tau$ represents the time decay. When $\tau = 0$, PW model becomes LWR model.

$$\begin{aligned} \rho_t + q_x &= 0 \\ q &= \rho \cdot v \\ v_t + vv_x + \frac{c_0^2}{\rho}\rho_x &= \frac{V_*(\rho) - v}{\tau} \end{aligned} \tag{2.10}$$

In conclusion, fundamental flow relationship, conservation law and speed evolution are three major components of all macroscopic traffic models.

## 2.3 The Stochastic Traffic Flow Modeling

In this section, we present and justify the stochastic macroscopic traffic model, and then discuss the methodology for traffic prediction using this model.

### 2.3.1 The Model

The stochastic model is based on the LWR model. As described in equation (2.9), the LWR model is a first-order partial differential equation, which assumes that the

traffic flow reaches the equilibrium state immediately. It predicts traffic flow well in relatively heavy traffic when the effects of individual driver behavior are minimal. From table 2.1, the average velocity at a location depends only on the density so that the nearby vehicles are traveling at the same speed. As a result, the model fails to capture the phenomenon of disconnected platoons of vehicles traveling at various velocities along the highways. See *Wong and Wong* (2002) and *Daganzo* (1995) for the critics of the LWR model. There are second-order adjustments based on fluid dynamics, proposed to enhance the LWR model, however, as shown in *Daganzo* (1995), any model that smoothers the discontinuity of density may predict backward flow and negative velocity.

Traffic data collected on highways reflects the cumulative effects of random microscopic phenomenon occurring along the highway and it suggests that an effective traffic flow model must also account for the stochastic and time-varying nature of traffic flow. The traffic flow evolution depends on the time of the day, the day of the week and the locations of segments of freeway. At any time, the traffic at a location might deviate from its nominal value because of the unpredictable and uncontrollable microscopic phenomena like sudden acceleration/deceleration, lane shifts, lane surface conditions, accidents, etc. However, after some time, the effects of the microscopic phenomena die down and the traffic reverts back to its nominal value. The deterministic model is not able to capture such mean-reverting property of the traffic evolution. In addition, traditional macroscopic models, such as LWR model, do not consider the on-ramp and off-ramp traffic flow along the highway. They normally assume that there are no cars entering or leaving the highway by putting zero on the right side of the conservation law in equation (2.5). That is obviously not the case in reality, the on-ramp and off-ramp traffic flow has to be incorporated into the traffic

flow model.

$$\begin{aligned} \rho_t + q_x &= g(\rho, x, t) \\ q &= \rho \cdot v \\ v &= V_*(\rho) \\ g(\rho, x, t) \cdot dx \cdot dt &= (a(x,t) + b(x,t) \cdot \rho) \cdot dx \cdot dt + \sigma(x,t) \cdot dW(x,t) \end{aligned} \tag{2.11}$$

In order to capture these real-life traffic phenomenons, we propose a stochastic model by adding a forcing function composed of mean-reverting drift and a Brownian Sheet (*Walsh*, 1986) on the right hand side of the conservation law to make it non-homogeneous. Equation (2.11) describes our stochastic model. The mean-reverting drift term $a(x,t) + b(x,t) \cdot \rho$, where $b(x,t) < 0$, is designed to capture the effects of the means of inflow/outflow and other factors that may affect flow conservation. We expect it to be positive at locations near entrances and time when traffic is entering the highway, and negative at locations near exits and time when traffic is leaving the highway. An intuitive explanation to assume a mean-reverting drift term is as the following. When the traffic is congested, i.e. the density is very high, some drives already on the highway will choose to leave the highway instead of wasting time in the traffic jam. In addition, the vehicles planning to enter the highway would reroute their trip. Therefore the traffic density would be pulled down with more people leaving the highway and less people entering the highway. In the other hand, when the traffic density is low, the traffic demand will tend to pull up the density with more vehicles entering the highway and fewer vehicles leaving the highway.

$$\begin{aligned} E[W(x,t)] &= 0 \\ Cov[W(x,t), W(y,s)] &= \min(x,y) \cdot \min(t,s) \end{aligned} \tag{2.12}$$

The volatility term $\sigma(x,t) \cdot \dot{W}(x,t)$ is designed to capture the magnitude of the resulting disturbance volatility to the flow conservation caused by the microscopic effects along the highway. The Brownian sheet $W(x,t)$ is a Gaussian process indexed by two parameters $x$ and $t$ with mean and covariance functions following equation (2.12). And we interpret $dW(x,t)$ as defined in equation (2.13), as both $dx$ and $dt$ go to zero (or are infinitesimal). Thus a Brownian Sheet assigns a random 'white' noise to each small rectangular region of the plane.

$$dW(x,t) = W(x+dx, t+dt) - W(x, t+dt) - W(x+dx, t) + W(x,t) \tag{2.13}$$

According to the property of Brownian sheet in equation (2.12) and the definition in equation (2.13), $dW(x,t)$ is independent and identically distributed normal random variable with mean and variance functions as shown in equation (2.14)

$$\begin{aligned} E[dW(x,t)] &= 0 \\ Var[dW(x,t)] &= dx \cdot dt \end{aligned} \tag{2.14}$$

Application of Brownian sheet allows the modeling of stochastic disturbances to flow conservation that result from microscopic effects, and makes them a function of space and time. As an example, erratic driver behavior and other microscopic phenomenon occur at random points in time and space, and have a varying degree of intensity. The intensity is captured by the parameter $\sigma(x,t)$.

In conclusion, the stochastic model changes the deterministic partial differential equation in the LWR model to a stochastic partial differential equation by incorporating the forcing function. The solution to the stochastic model and the methodology using this model for traffic prediction are described in detail in the following sections.

### 2.3.2 Godunov's Scheme

The core of a traffic flow model is a stochastic partial differential equation. One specialty of this stochastic partial differential equation is that, due to the nonlinearity of the flux function, none of the partial derivatives exists. However one can rewrite it as an integral equation and show that there is a unique weak solution (*Gunnarsson*, 2006). Such solution exhibits the phenomenon of shock waves under the appropriate conditions. This means that the solution is a discontinuous function. At the backup, the density suddenly increases, forcing vehicles to slow down while approaching it. As more vehicles join the backup, this jump point moves backwards along the highway. This is the shock-wave predicted by the model.

It is difficult to obtain an analytic solution for the stochastic partial differential equation. Therefore finite difference schemes will be used to solve it numerically for approximate solutions. Since the solution is discontinuous, a straight forward finite difference scheme will fail to solve the resulting discretization of the stochastic partial differential equation. Godunov's scheme is one of the numerical schemes which can handle this difficulty and give a high-resolution solution to the stochastic partial differential equation. The following paragraphs explain Godunov's scheme in some details.

$$\rho_t + Q(\rho)_x = g(\rho, x, t) \tag{2.15}$$

A partial differential equation like equation (2.15) is called conservation law with source term, where $g(\rho, x, t)$ is the source term and $Q(\rho)$ is the flux function. Typically the flux function is nonlinear and it will cause the solution of the partial differential equation to be discontinuous. In this case, naive finite difference schemes cannot be used to solve this system, because they may work well for smooth parts of the solution, however they can give disastrous results when discontinuities are present (*LeVeque*,

1992). For the case of homogeneous conservation law when $g(\rho, x, t) \equiv 0$, many high-resolution numerical algorithms have been developed to overcome this difficulty. These numerical algorithms are able to give second or higher order spatial accuracy in smooth parts of the solution and sharp resolution near the discontinuities as well (*LeVeque*, 1998). Godunov's scheme is one of the most efficient of such schemes.

$$\rho_t + Q(\rho)_x = 0 \tag{2.16}$$

In Godunov's scheme for solving homogeneous conservation law in equation (2.16), the variable $\rho$ is considered as piecewise constant over a mesh of cells at each time step and the evolution of $\rho$ is determined by the exact solution of the Riemann problem at the boundaries between adjacent cells. The solution to Riemann problem is the key component of Godunov's scheme. So before introducing Godunov's scheme, we first discuss the Riemann problem and its solution.

#### 2.3.2.1 Riemann Problem

Riemann problem consists of the homogeneous conservation law and piecewise constant initial condition with jumps at the origin. When $Q$ is a concave function (this assumption is satisfied in most traffic flow models), equation (2.17) describes the Riemann problem.

$$\begin{aligned} &\rho_t + Q(\rho)_x = 0 \\ &\rho(x,0) = \begin{cases} \rho_l & \text{if } x < 0 \\ \rho_r & \text{if } x > 0 \end{cases} \end{aligned} \tag{2.17}$$

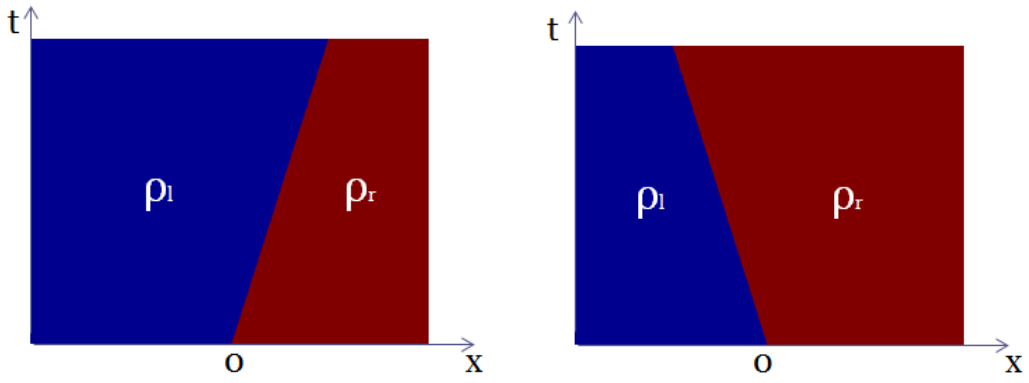
Due to the nonlinearity of the flux function, the characteristic lines will intersect at some point if the initial condition is not constant. When two characteristic lines intersect, there is a jump in the value of $\rho$, and it is discontinuous at this point.

Generally, when the flux function is nonlinear, there only exists a weak solution to the conservation law. Lax proved that there exists a unique weak solution to the Riemann problem under the so-called "Lax's entropy condition" (*Lax*, 1972). There are two types of solutions to the Riemann problem depending on the values of $\rho_l$ and $\rho_r$. When $\rho_l$ and $\rho_r$ satisfy the following Lax's entropy condition in equation (2.18), the solution to the Riemann problem is in the form of a shock wave; Otherwise, the solution is in the form of a rarefaction wave.

$$Q'(\rho_l) > \frac{Q(\rho_l) - Q(\rho_r)}{\rho_l - \rho_r} > Q'(\rho_r) \tag{2.18}$$

Since the flux function is assumed to be concave, equation (2.18) is equivalent to the condition in equation (2.19). This means that when the upstream density is less than the downstream density, a shock wave occurs.

$$\rho_l < \rho_r \tag{2.19}$$




(a) Forward shock wave (b) Backward shock wave

Figure 2.2: The shock wave

When $\rho_l < \rho_r$, the solution to the Riemann problem is given in equation (2.20). From the solution, we can see that the boundary of $\rho_l$ and $\rho_r$, where the shock wave occurs, is traveling at the speed of $s = \frac{Q(\rho_l)-Q(\rho_r)}{\rho_l-\rho_r}$. Therefore $s$ is also called the speed

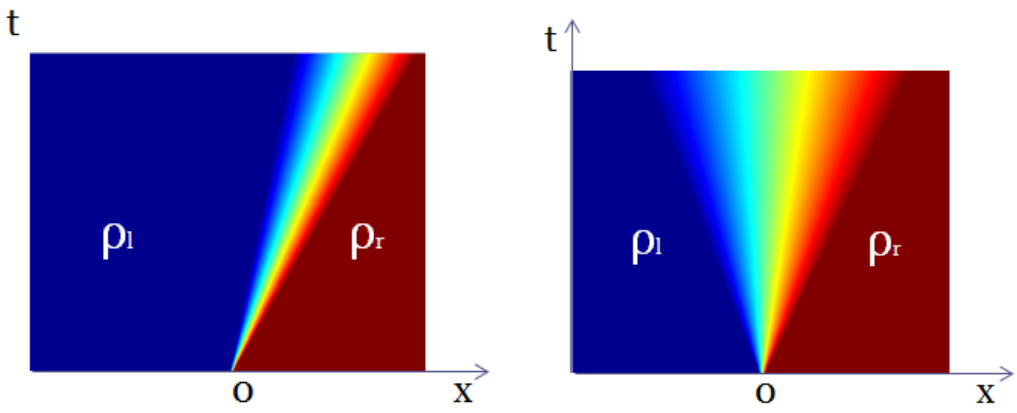
of the shock wave. The shock wave has a realistic meaning: When $Q(\rho_l) < Q(\rho_r)$, the shock wave speed is positive. Under this situation, it is like the light traffic upstream will push the more congested traffic ahead downstream, and the shock wave travels forward. Figure 2.2(a) shows the forward shock wave. When $Q(\rho_l) > Q(\rho_r)$, the shock wave speed is negative. In this situation, $\rho_r$ must be quite large, which means the downstream traffic is very congested, the congestion will stop the light traffic upstream, and the shock wave travels backward. Figure 2.2(b) shows the backward shockwave.

$$\begin{aligned} \text{When } \rho_l &< \rho_r : \\ \rho(x,t) &= \begin{cases} \rho_l & \text{if } x/t < \frac{Q(\rho_l)-Q(\rho_r)}{\rho_l-\rho_r} \\ \rho_r & \text{if } x/t > \frac{Q(\rho_l)-Q(\rho_r)}{\rho_l-\rho_r} \end{cases} \end{aligned} \tag{2.20}$$

When the entropy condition in equation (2.19) is not satisfied, in other words, when $\rho_l > \rho_r$, the solution to the Riemann problem is in the form of rarefaction wave given in equation (2.21). When $0 < Q'(\rho_l) < Q'(\rho_r)$, the solution is a forward rarefaction wave as shown in Figure 2.3(a). When $Q'(\rho_l) < 0 < Q'(\rho_r)$, the solution is a centered rarefaction wave as shown in Figure 2.3(b). When $Q'(\rho_l) < Q'(\rho_r) < 0$, the solution is a backward rarefaction wave as shown in Figure 2.3(c).

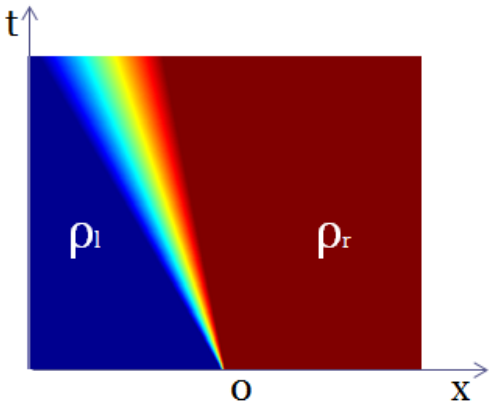
$$\begin{aligned} \text{When } \rho_l &> \rho_r : \\ \rho(x,t) &= \begin{cases} \rho_l & \text{if } x/t < Q'(\rho_l) \\ Q'^{-1}(x/t) & \text{if } Q'(\rho_l) < x/t < Q'(\rho_r) \\ \rho_r & \text{if } x/t > Q'(\rho_r) \end{cases} \end{aligned} \tag{2.21}$$

The value of $\rho$ at the origin is important for Godunov's scheme. By combining the shock wave solution (2.20) and rarefaction solution (2.21), the density at the origin




(a) Forward rarefaction wave

(b) Centered rarefaction wave




(c) Backward rarefaction wave

Figure 2.3: The rarefaction wave

is given in equation (2.22)

$$\rho(0,t) = \Psi(\rho_l, \rho_r) = \begin{cases} \rho_l & \text{if } \rho_l < \rho_r, s > 0 \text{ or } \rho_r < \rho_l < \rho_c \\ \rho_r & \text{if } \rho_l < \rho_r, s < 0 \text{ or } \rho_c < \rho_r < \rho_l \\ \rho_c & \text{if } \rho_r < \rho_c < \rho_l \end{cases} \tag{2.22}$$

where

$$\begin{aligned} s &= \frac{Q(\rho_l) - Q(\rho_r)}{\rho_l - \rho_r} \\ \rho_c &= Q'^{-1}(0) \end{aligned}$$

$\Psi(\rho_l, \rho_r)$ is the solver of the Riemann problem, it will be used later in Godunov's scheme to calculate the flow between two adjacent discretized cells. $\rho_c$ is the critical density, which maximizes the flux function. The details will be discussed in the next section.

#### 2.3.2.2 Godunov's Scheme for Homogeneous Conservation Law

In the numerical scheme for homogeneous conservation law of equation (2.16), the space region $[0, L]$, where $L$ is the total length of the highway, is divided into $N$ small cells with $x_0 = 0$, $x_1$, ..., $x_{N-1}$, $x_N = L$, where $x_i$ $(i = 0, ..., N)$ are the boundaries of the cells. Let $\triangle x_i (i = 1, ..., N)$ represent the length of cell $i$ with $\triangle x_i = x_i - x_{i-1}$. The time dimension is equally discretized into $M$ time intervals with $t_0$, $t_1$, ..., $t_M$, and the time interval is $\triangle t$. Godunov's scheme is a conservative numerical technique, suggested by S. K. Godunov (*Godunov*, 1959) for solving partial differential equations. In this method, the conservative variables are considered as piecewise constant over each cell at each time step, and the time evolution is determined by the exact solution of the Riemann problem at the inter-cell boundaries (*Hirsch*, 1990). The steps of Godunov's scheme for homogeneous conservation law are described in Algorithm 1.

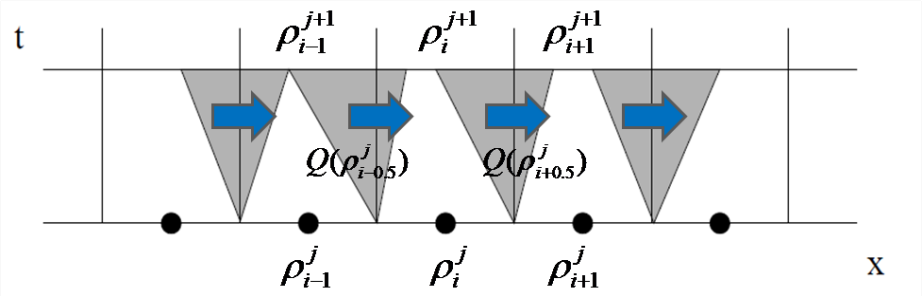



Figure 2.4: Godunov's scheme for homogeneous conservation law

In order to guarantee that the waves do not interact with each other, $\triangle t$ should satisfy Courant-Friedrichs-Lewy(CFL) condition given in equation (2.26), where $|a_{max}|$

**Algorithm 1** Godunov's scheme for homogeneous conservation law.

**Step 1: Initialization** Let $\rho_i^j$ denote the average density over cell $i$ at time $t_j$. When $j = 0$, obtain the initial condition for $\rho_i^0$ according to equation (2.23):

$$\rho_i^0 = \int_{x_{i-1}}^{x_i} \rho(x, 0)dx \qquad \forall i = 1, 2, ..., N \tag{2.23}$$

**Step 2: Riemann Solver** Let $\rho_{i-0.5}^j$ and $\rho_{i+0.5}^j$ denote the average value of $\rho$ over time interval $[t_j, t_{j+1}]$ at the left boundary and the right boundary of cell $i$ respectively. The value of $\rho_{i-0.5}^j$ and $\rho_{i+0.5}^j$ is obtained according to the solution of the Riemann problem in equation (2.22):

$$\begin{aligned} \rho_{i-0.5}^j &= \Psi(\rho_{i-1}^j, \rho_i^j) \\ \rho_{i+0.5}^j &= \Psi(\rho_i^j, \rho_{i+1}^j) \end{aligned} \tag{2.24}$$

**Step 3: Update** Given the boundary density in equation (2.24), the boundary traffic flow is derived based on speed-density function. According to conservation law, the density in the next time interval is updated following equation (2.25). The state variables in the next time interval are also averaged over each cell and become a piecewise constant approximation. After the update, $j = j + 1$, and go to step 2. The scheme is illustrated in Figure 2.4.

$$\rho_i^{j+1} = \rho_i^j + \frac{\triangle t}{\triangle x_i}[Q(\rho_{i-0.5}^j) - Q(\rho_{i+0.5}^j)] \tag{2.25}$$

is the maximum wave speed.

$$|a_{max}|\triangle t < \triangle x_i \quad \forall i = 1, 2, \cdots, N \tag{2.26}$$

In conclusion, by following Godunov's scheme, the evolution of the traffic density for the LWR model can be obtained. For the nonhomogeneous conservation law stochastic partial differential equation, Godunov's scheme must be modified to accommodate the forcing function. The details are introduced in the following section.

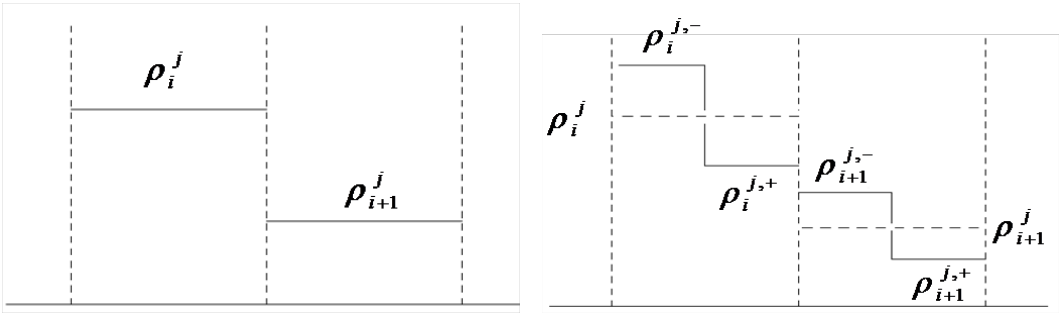
#### 2.3.2.3 Modified Godunov's Scheme for Nonhomogeneous Conservation Law

For conservation equation with source such as equation (2.15), the simplest approach is to use fractional step splitting method, which solves the nonhomogeneous conservation equation in two steps for every time interval. In the first step, the homogeneous conservation law (2.16) is solved by using the Godonuv's scheme introduced above. Then in the second step, an ordinary differential equation is solved based on the solution obtained from the first step:

$$\rho_t = g(\rho, x, t)$$

The simple fractional step splitting method is successful for many problems, however will fail in some types of problems, particularly when $\rho_t$ is relative small to $q_x$ and $g(\rho, x, t)$, as there may exist a steady state solution in which the flux and source term are exactly balanced. However in the two steps of the simple fractional step splitting method, the solution of homogeneous conservation law obtained in the first step will deviate far away from the true solution because $g(\rho, x, t)$ is relatively larger than $\rho_t$. The second step then pulls it back around the true solution. It is unlikely that the second step will exactly cancel the change caused by the first step because different numerical methods are used in the two steps. Even if the change and counter-change are canceled exactly, a small perturbation in the source term may result in a big noise in the state variable $\rho$.

*LeVeque* (1998) developed an algorithm in which the source term could be incorporated into the Godunov's method without fractional steps. The idea of the algorithm is to introduce a Riemann problem in the center of each grid cell whose flux difference can exactly cancel the source term. In this way, the source term will not cause any wave within the cell. The steps of the modified Godunov's method for




(a) Godunov's scheme (b) Modified Godunov's scheme

Figure 2.5: Godunov decomposition

nonhomogeneous conservation law are described in Algorithm 2.

Compared with Algorithm 1, Algorithm 2 adds a Godunov decomposition step as illustrated in Figure 2.5. Equation (2.28) in the step of Godunov decomposition means: 1. the average density in this cell is unchanged after the decomposition. 2. The discontinuity of the traffic flow between the two decomposed sub segments can exactly cancel the source term, so there is no shock wave caused by the source term. Modified Godunov's scheme gives stable and accurate solution to the nonhomogeneous conservation law.

### 2.3.3 Simulation-Based Traffic Prediction

Since equation (2.11) usually does not have an analytic solution, simulation is a commonly used approach to solve complicated stochastic systems. Section 2.3.2.3 gives the numerical scheme to solve the nonhoogeneous conservation law. In order to predict the traffic state in the future, given the initial condition $\rho(x, 0)$ at $t = 0$, we project the density evolution according to the numerical scheme in Algorithm 2. Since the model we proposed is stochastic, Monte Carlo simulation is applied to generate a number of paths for the density projection. The average density over all simulated paths is considered as the prediction of the future density. The simulation scheme for the traffic prediction is described in Algorithm 3:

In conclusion, given the initial density condition on the road, the simulation is

**Algorithm 2** Godunov's scheme for nonhomogeneous conservation law.

**Step 1: Initialization** Let $\rho_i^j$ denote the average density over cell $i$ at time $t_j$. When $j = 0$, obtain the initial condition for $\rho_i^0$ according to equation (2.27):

$$\rho_i^0 = \int_{x_{i-1}}^{x_i} \rho(x, 0)dx \qquad \forall i = 1, 2, ..., N \tag{2.27}$$

**Step 2: Godunov Decomposition** In order to exactly cancel the source term within each cell, each cell is decomposed into two parts. The difference of the traffic flow between these two parts is compensated by the source term so that there are no shock waves within the cell. Let $\rho_i^{j,-}$ and $\rho_i^{j,+}$ denote the average value of $\rho$ in the upstream part and the downstream part of cell $i$ respectively. $\rho_i^{j,-}$ and $\rho_i^{j,+}$ should satisfy the following conditions:

$$\begin{aligned} \rho_i^{j,-} + \rho_i^{j,+} &= 2\rho_i^j \\ Q(\rho_i^{j,+}) - Q(\rho_i^{j,-}) &= g(\rho_i^j, x_i, t_j)\triangle x_i \end{aligned} \tag{2.28}$$

**Step3: Riemann Solver** Let $\rho_{i-0.5}^j$ and $\rho_{i+0.5}^j$ denote the average value of $\rho$ over time interval $[t_j, t_{j+1}]$ at the left boundary and the right boundary of cell $i$ respectively. The value of $\rho_{i-0.5}^j$ and $\rho_{i+0.5}^j$ is obtained according to the solution of the Riemann problem in equation (2.22):

$$\begin{aligned} \rho_{i-0.5}^j &= \Psi(\rho_{i-1}^{j,+}, \rho_i^{j,-}) \\ \rho_{i+0.5}^j &= \Psi(\rho_i^{j,+}, \rho_{i+1}^{j,-}) \end{aligned} \tag{2.29}$$

**Step 4: Update** Given the boundary density in equation (2.29), the boundary traffic flow is derived based on speed-density function. According to conservation law, the density in the next time interval is updated following equation (2.30). The state variables in the next time interval are also averaged over each cell and become a piecewise constant approximation. After the update, $j = j + 1$, and go to step 2.

$$\rho_i^{j+1} = \rho_i^j + \frac{\triangle t}{\triangle x_i}[Q(\rho_{i-0.5}^j) - Q(\rho_{i+0.5}^j) + g(\rho_i^j, x_i, t_j)\triangle x_i] \tag{2.30}$$

applied to generate a number of projections for traffic state evolution. Based on these projections, the distribution of the traffic state at any location in the future is obtained. Different from the deterministic traffic model, which gives a specific value as a future prediction for the traffic state, this stochastic model gives a dis-

**Algorithm 3** Simulation-based traffic prediction.

**Step 1: Scenario Generation** Generate $K$ different paths of Brownian sheets $W^k(x,t)$, $k = 1, 2, ..., K$ where $W^k(x,t)$ represent the simulated Brownian sheet in scenario $k$.

**Step 2: Numerical Simulation** In scenario $k$, given the Brownian sheet $W^k(x,t)$ and initial condition $\rho(x,0)$, solve the stochastic partial differential equation using modified Godunov's scheme for nonhomogeneous conservation law described in Algorithm 2. Obtain the evolution of the density under this scenario: $\rho^k(x,t)$.

**Step 3 Report Prediction** Take the average of the $\rho^k(x,t)$ over $K$ scenarios as the prediction made by the stochastic traffic flow model.

tribution of the future traffic state as the output. This distribution can be used to estimate the average as the prediction and the standard deviation as the reliability of the prediction. This information is very important and useful for many intelligent transportation applications.

## 2.4 Case Study

In this section, an empirical case study using real highway data is investigated and it will provide insights on the prediction performance of the stochastic traffic flow model. The data used in this case study is obtained from the Virginia Department of Transportation and it records the traffic flow information on one segment of highway I95 Northbound heading to Washington DC. The raw data consists of one-minute aggregated traffic flow and average speed information at 23 sensor stations along one segment of the highway on everyday from April 1st 2009 to June 1st 2009. The total length of the highway segment is 15.67 miles and the average distance between adjacent sensor stations is 0.78 miles. The highway segment is marked as red in the left map of Figure 2.6.

Although there are 23 sensor stations in the raw data, after preprocessing, we found 8 of them do not have valid data because of malfunction or other unknown

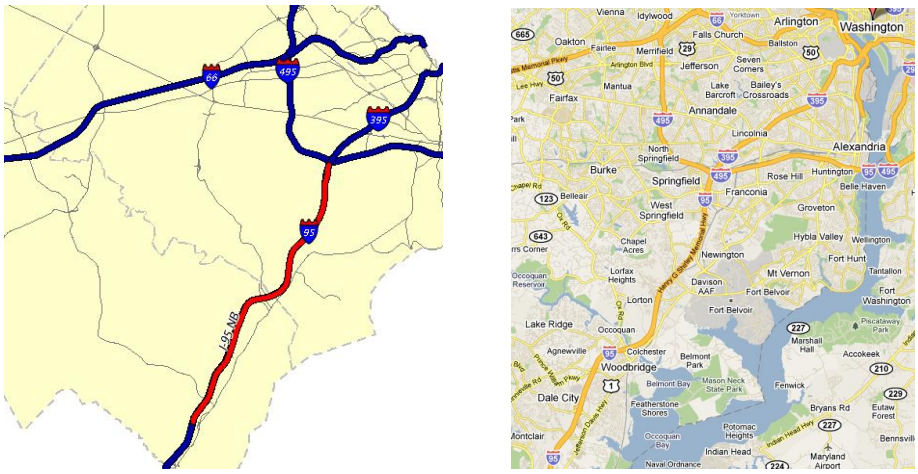



Figure 2.6: The map of the highway

reasons. Therefore we exclude these 8 sensor stations from this case study and only use the data from the rest 15 sensor stations.

### 2.4.1 Speed-Density Function Calibration

As explained in Section 2.2.3, the speed-density function is a significant component in macroscopic traffic flow modeling. The function also serves as a bridge between the model and other applications: Sometimes the traffic density is the target of model prediction. However many ITS applications, such as real-time travel time estimation, need the prediction of future speed. In this case, the speed-density function can be used to transform the predicted density into speed for the ITS application. The speed-density function is important for the performance of the traffic flow model and the calibration of this function is usually the first step in macroscopic traffic modeling.

$$\text{RMSE}(\hat{x}) = \sqrt{\frac{\sum_{i=1}^{n}(\hat{x}_i - x_i)^2}{n}} \tag{2.31}$$

As described in Table 2.1, there are four potential speed-density functions. We will test these on the field data and the one that fits the data best will be selected as the prediction model. Since the parameters of sensor stations are different, the

speed-density function is calibrated at each sensor station. The general steps of this calibration are as follow: The data of a sensor station is divided into training and testing sets. The data between April 1st and May 15th is training data set and the data between May 16th and June 1st is the testing data set. Then, for each of the four functions, the parameters are calibrated using the training data set, and the Root Mean Square Error (RMSE) of speed for each calibrated function on the training data set is evaluated as the training error. The RMSE is defined in equation (2.31), where $\hat{x}$ is estimated value of the variable from the model and $x$ is observed value; In the last, the RMSE for each calibrated function is evaluated on the testing data set as the testing error. By comparing the four speed-density functions in terms of both training error and testing error, the function which fits the data best will be selected.

$$\ln(v) = \min\{\ln(v_f), \ln(\alpha) + m\ln(\rho)\} \tag{2.32}$$

For the speed-density function calibration, Greenshields, Greenberg and Underwood functions can be calibrated by using regular linear regression after variable transformation. For the last function, Piecewise function, the natural logarithm transformation is taken on both sides of the equation, and then $\ln(v)$ is a piecewise linear function of $\ln(\rho)$ as shown in equation 2.32. The piecewise linear regression is applied to calibrate the parameters using ordinary least square.

Table 2.2 summarizes the training and testing errors of all four functions on every sensor station. It shows that the Piecewise function has the smallest training and testing errors. Therefore Piecewise function is chosen as the speed-density function in this case study and will be used in the following sections. Figure 2.7 shows the regression results of Piecewise function at all 15 stations. The dots are observed data points and the solid line is fitted Piecewise function. From the figures, we can see that Piecewise function fits the data very well.

Table 2.2: Fitting errors of speed-density functions

| RMSE | Greenshields | | Greenberg | | Underwood | | Piecewise | |
|---|---|---|---|---|---|---|---|---|
| Stations | Training | Testing | Training | Testing | Training | Testing | Training | Testing |
| 1 | 8.9 | 9.06 | 13.88 | 14.09 | 8.8 | 9.04 | 4.48 | 4.24 |
| 2 | 7.46 | 6.99 | 10.6 | 10.94 | 6.52 | 6.03 | 4.35 | 3.04 |
| 3 | 12.66 | 12.95 | 12.03 | 12.31 | 11 | 11.25 | 5.85 | 5.94 |
| 4 | 10.98 | 10.39 | 11.68 | 11.44 | 9.47 | 9.07 | 3.52 | 2.74 |
| 5 | 14.84 | 14.29 | 12.69 | 12.43 | 12.9 | 12.56 | 5.07 | 5.45 |
| 6 | 19.37 | 17.46 | 14.31 | 12.92 | 17.11 | 15.53 | 8.22 | 8.42 |
| 7 | 14.67 | 12.9 | 12.51 | 11.05 | 13.24 | 11.81 | 7.19 | 8.77 |
| 8 | 11.51 | 9.9 | 11 | 9.85 | 11.33 | 10.1 | 6.75 | 6.9 |
| 9 | 10.84 | 9.44 | 11.1 | 9.79 | 8.91 | 7.67 | 6.64 | 7.31 |
| 10 | 10.8 | 9.68 | 10.52 | 9.5 | 10.05 | 9.01 | 5.06 | 5.15 |
| 11 | 11.08 | 9.94 | 9.79 | 9.31 | 9.33 | 8.72 | 6.84 | 5.5 |
| 12 | 5.78 | 5.03 | 9.17 | 8.68 | 6.94 | 6.36 | 4.11 | 2.86 |
| 13 | 5.82 | 5.16 | 10.64 | 10.28 | 6.86 | 6.38 | 3.66 | 2.75 |
| 14 | 5.16 | 4.29 | 11.88 | 11.39 | 7.41 | 6.67 | 4.6 | 4.93 |
| 15 | 4.49 | 3.88 | 10.57 | 10.33 | 6.04 | 5.6 | 4.47 | 4.47 |
| Average | 10.29 | 9.42 | 11.49 | 10.95 | 9.73 | 9.05 | 5.39 | 5.23 |

Figure 2.8 presents the calibrated parameters of the Piecewise function in all 15 stations. $\rho_c$ is the critical density starting from which the speed is decreasing from the free flow speed $v_f$. Based on the Piecewise function, $\rho_c = (\frac{v_f}{\alpha})^{1/m}$. A reason why the critical densities differ for stations is that the number of lanes change along the highway, and the density is aggregated over all the lanes.

### 2.4.2 Forcing Function Calibration

This section will describe the procedure to calibrate the parameters of the forcing function $g$ in the stochastic traffic flow model equation (2.11). The objective of forcing function calibration is to determine the value of $a(x,t)$, $b(x,t)$ and $\sigma(x,t)$ at any location $x$ and any time $t$. Because the parameters are different for the sensor stations, the forcing function will be calibrated for each sensor station separately (as was done for the speed-density function). We aggregate the parameters in 5-minute

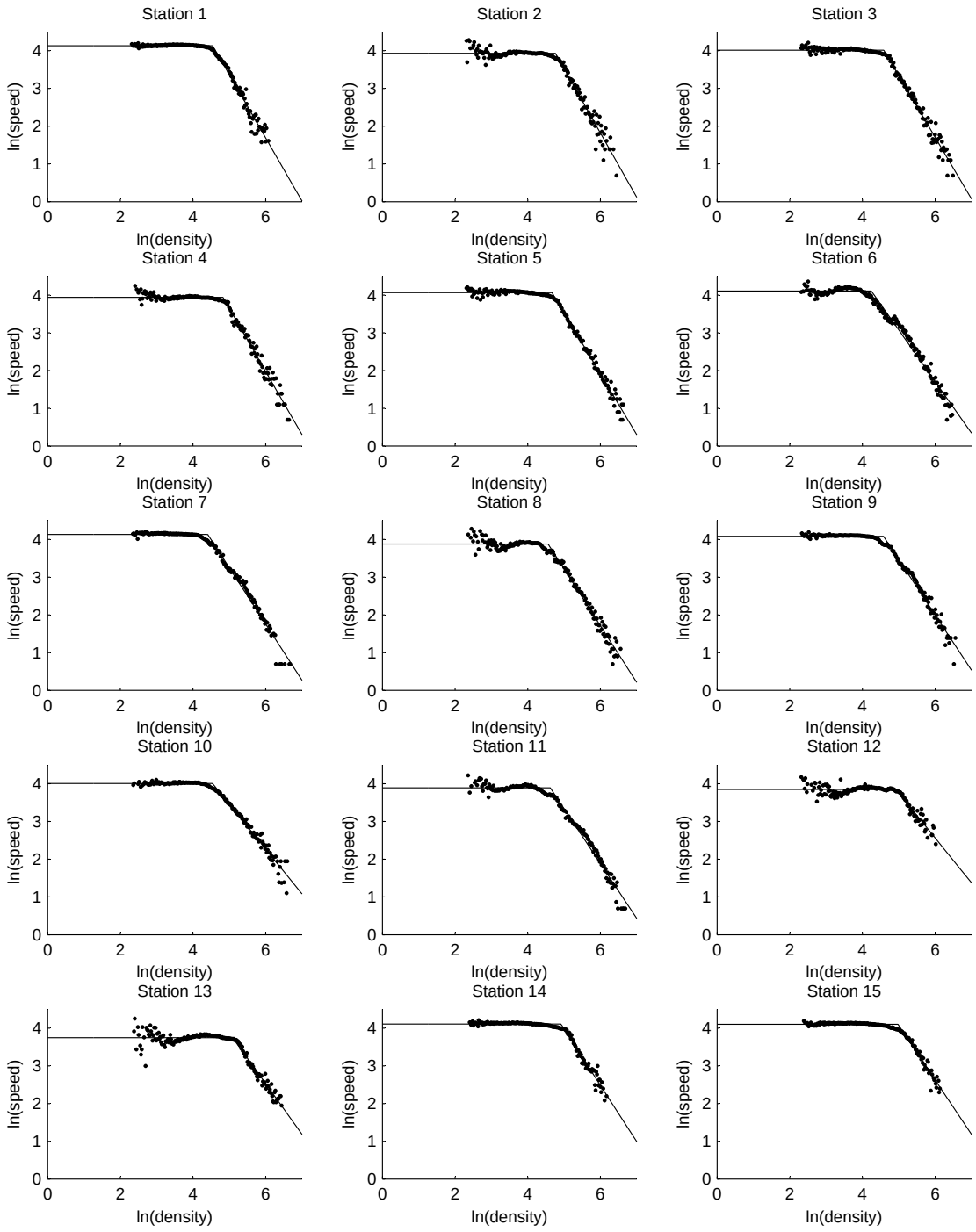



Figure 2.7: Regression results of the Piecewise speed-density function

time interval and assume that the parameters of every sensor station are constant in the same time period for different days. For example, the parameters at sensor station 3 are the same for the time period between 7:15am and 7:19am on any day. Therefore

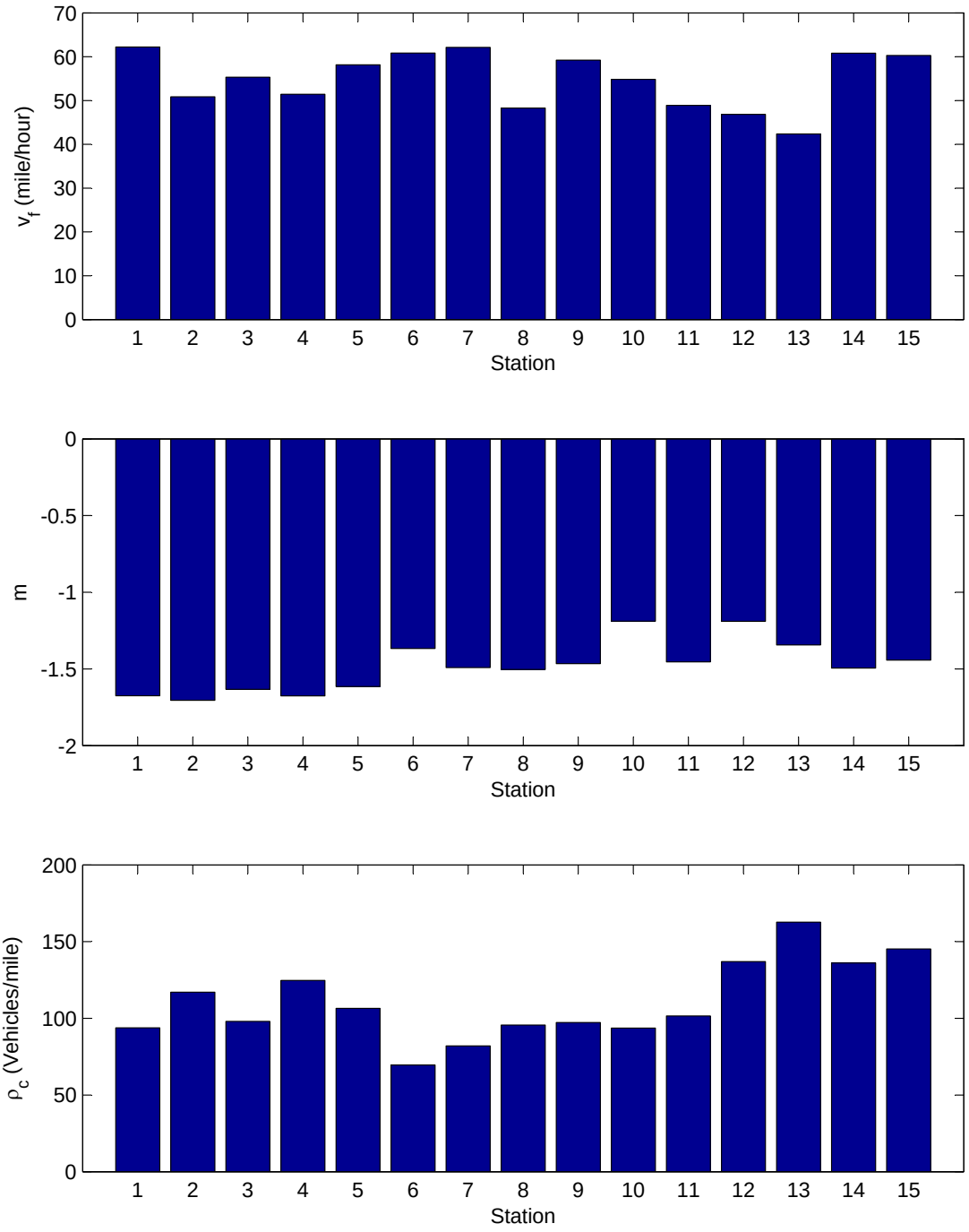



Figure 2.8: Calibrated parameters of the Piecewise speed-density function

in order to calibrate the parameters during this time period, we can aggregate all the

data between 7:15am and 7:19am.

$$g(x_i, t_j) = \frac{\rho(x_i, t_{j+1}) - \rho(x_i, t_j)}{t_{j+1} - t_j} + \frac{q(x_{i+1}, t_j) - q(x_{i-1}, t_j)}{x_{i+1} - x_{i-1}} \tag{2.33}$$

In order to calibrate the parameters for station $i$ at time $t_j$, we first apply finite difference method to calculate the value of the source term at $x_i$ and $t_j$, $g(x_i, t_j)$, according to equation (2.33). In equation (2.33), $\rho_t$ is calculated using forward difference and $q_x$ is calculated using central difference.

$$\begin{aligned} g(x_i, t_j) &= a(x_i, t_j) + b(x_i, t_j)\rho(x_i, t_j) + \sigma(x_i, t_j)\frac{\Delta W(x_{i+1} - x_{i-1}, t_{j+1} - t_j)}{(x_{i+1} - x_{i-1})(t_{j+1} - t_j)} \\ Var[\Delta W(x_{i+1} - x_{i-1}, t_{j+1} - t_j)] &= (x_{i+1} - x_{i-1})(t_{j+1} - t_j) \end{aligned} \tag{2.34}$$

Since the source term $g(x_i, t_j)$ is a linear function of the density $\rho(x_i, t_j)$, as shown in equation (2.34), linear regression between $g(x_i, t_j)$ and $\rho(x_i, t_j)$ can be applied to estimate the parameters $a(x_i, t_j)$ and $b(x_i, t_j)$. $\sigma(x_i, t_j)$ is calculated by equation (2.35), where $\epsilon$ is the residual of the linear regression. The justification for equation (2.35) is that the residual $\epsilon$ is independent and identically normally distributed with distribution $N(0, \frac{\sigma(x_i,t_j)}{(x_{i+1}-x_{i-1})(t_{j+1}-t_j)})$ according to equation (2.34).

$$\sigma(x_i, t_j) = \sqrt{E[\epsilon^2] \cdot (x_{i+1} - x_{i-1})(t_{j+1} - t_j)} \tag{2.35}$$

Figure 2.9 shows an example of the linear regression between $g$ and $\rho$ at sensor station 3 using aggregated data between 7:15am and 7:19am on different days. The $R^2$ of linear regression is 0.85, which means the linear fit is nice. The estimated parameters are: $\hat{a} = 1621.3$, $\hat{b} = -21.9$, $\hat{\sigma} = 10.9$ at sensor location 3 from 7:15am to 7:19am. The p-value of the linear model is also checked and the 95% confidence interval is $[1513.1, 1729.5]$ for $\hat{a}$ and $[-23.5, -20.4]$ for $\hat{b}$. 0 is not included in the confidence interval for neither $\hat{a}$ or $\hat{b}$, which means $\rho$ is significant for $g$ in the linear

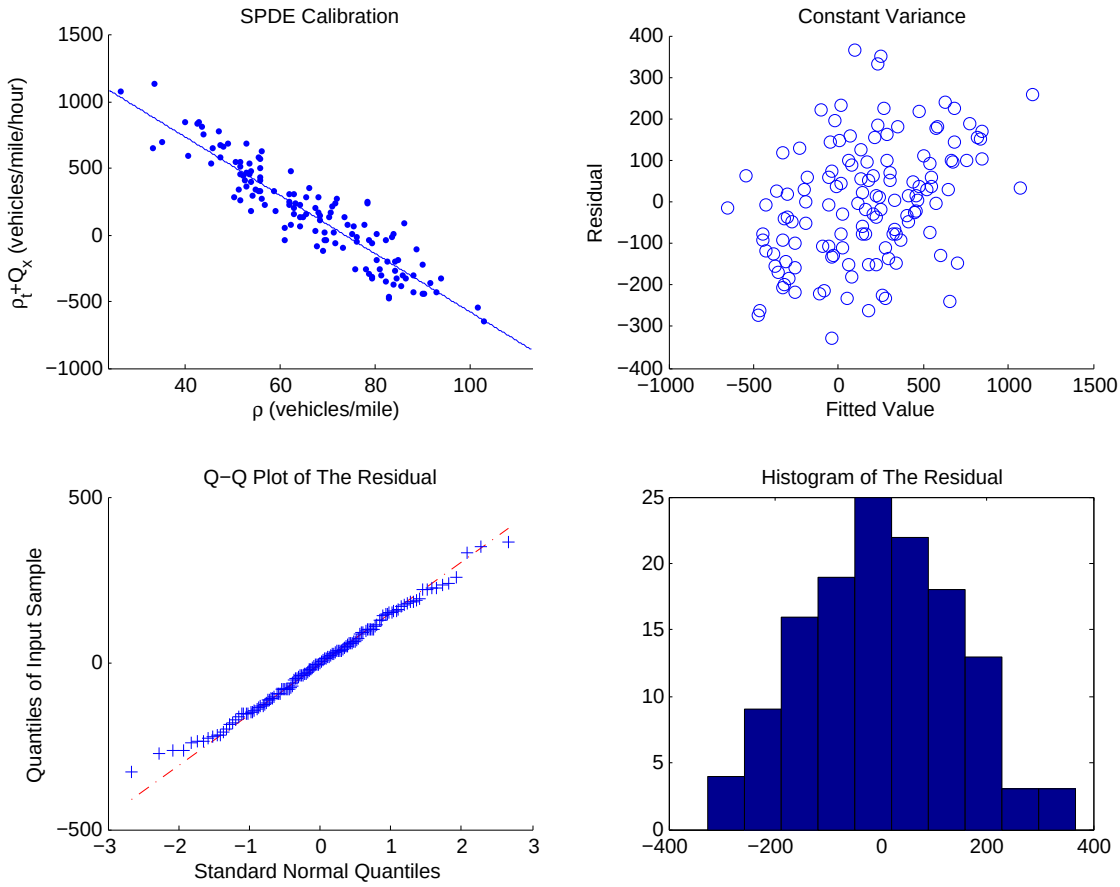



Figure 2.9: Regression results of the forcing function at station 3 between 7:15am and 7:19am

regression. In addition, nether the Q-Q plot nor the histogram of the residual can reject the hypothesis that the residual follows a normal distribution. Therefore Figure 2.9 provides support to the underlying assumption of the linear regression. By following the above procedure, the parameters can be estimated at any location and any time.

Figure 2.10 plots the evolution of estimated parameters over time on sensor station 2 as an example. These estimated parameters on all 15 stations will be plugged into the model for evaluating the prediction performance in the next section.

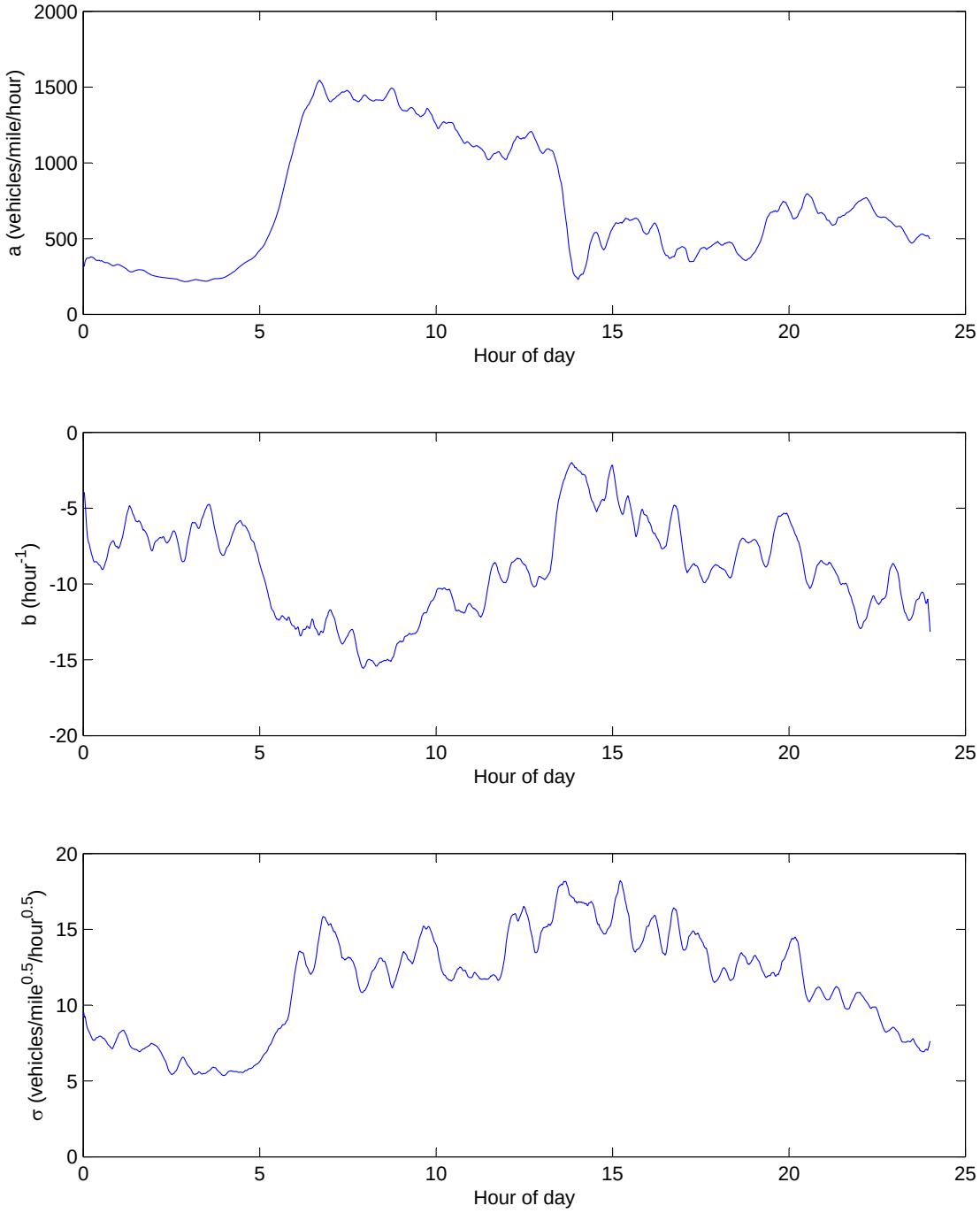



Figure 2.10: Evolution of the forcing function parameters over the day at station 2

### 2.4.3 Comparison of the Stochastic Model with the Deterministic Model

After the calibration of speed-density function and forcing function, the model is validated by evaluating its prediction performance. This step is necessary to assure

that the model has a robust predictive capability. In this section, we will investigate the improvement of the prediction accuracy of the stochastic model over the deterministic LWR model.

Section 2.3.3 gives a detailed description of the procedure for prediction using the stochastic traffic flow model. We use the calibrated parameters and initial traffic state to obtain the prediction. In the deterministic LWR model, the source term is equal to 0. We apply the same procedure in Section 2.3.3 by setting $g \equiv 0$. Figure 2.11 illustrates and compares the RMSE of stochastic and deterministic models over all 15 sensor stations. It clearly shows the advantage of the stochastic model over the deterministic model. From the figure, we can see that, on average, the stochastic model decreases the RMSE about 44% compared with the deterministic model.

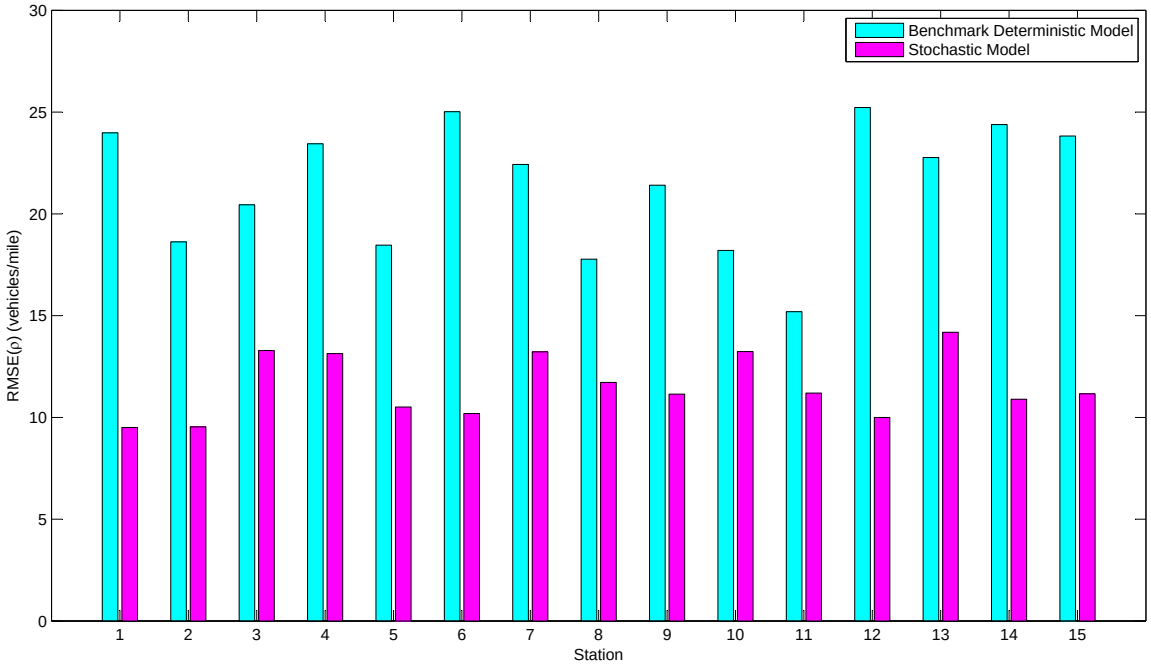



Figure 2.11: Comparison of RMSE between the deterministic model and the stochastic model at all stations

In conclusion, here we present the result of speed-density function and forcing function calibration using real traffic data. Then, the comparison of the stochastic model with deterministic model shows that the stochastic model can provide a more

accurate prediction. This is is the reason we prefer a stochastic model for the traffic flow prediction.

## 2.5 Conclusion

In this chapter, we propose a stochastic traffic flow model, which is an extension of the classical LWR model. The stochastic model fixes some deficiencies of the deterministic LWR model and introduces the effects of real stochastic factors impacting the traffic flow evolution. The solution of the stochastic partial differential equation in the model is non trivial, and simple numerical algorithms will fail to solve it. A numerical scheme is introduced to solve this stochastic model which is based on the classical Godunov's scheme. This numerical scheme enables the prediction of the density. We conclude with a case study using practical traffic data and show how this model could be applied. The case study also shows that the stochastic traffic model increases the prediction accuracy of the future traffic density by an average of 44% when compared with the deterministic LWR model. The stochastic model provides a foundation for many other ITS applications.

# CHAPTER III

# Stochastic Filtering For Real Time Traffic Prediction

## 3.1 Introduction

Real time traffic state estimation is the problem of estimating the current value of the traffic state variables (such as traffic density, traffic model parameters etc.) over the highway, in real time and as precisely as possible given limited measurements of traffic conditions at specified locations along the highway. These measurements are made by detection and surveillance devices installed on the highway collecting data needed to perform traffic management functions. There are mainly two kinds of such measurements. One measurement is called Eulerian measurement. This refers to sensors installed at fixed locations along the highway to collect the measurements. These sensors include inductive loop detectors, RFID transponders, magnetometers, video image processors, microwave radars and acoustic sensors, and they usually provide measurements of volume, speed and occupancy at their locations. The other type of measurement is called Lagrangian measurement. This measurement refers to data gathered from sensors which move along a trajectory in the field being sensed (*Work et al.*, 2009). Lagrangian sensing utilizes mobile communication devices, such as GPS, smart phones etc. to collect data. Such mobile devices can provide very

accurate information of speed, acceleration, position etc. about individual vehicles along the highway.

Like traffic state prediction discussed in Chapter II, traffic state estimation is also a very important component in many applications of intelligent transportation system. For example, in the ITS application of trip advisory, the more accurate the estimation of current traffic state, the more reliable the travel information provided to the drivers. On-line traffic state estimation can also be used for quick anomaly detection. If one lane is forced to close because of an accident, the change in the parameter 'number of lanes' can be quickly detected by the traffic state estimation algorithm and reported to the highway administer. From the point of view of impact, estimation is more fundamental than prediction since the accuracy of estimation not only directly affects the performance of ITS applications but also significantly affects the prediction accuracy (in most prediction models). In most traffic flow models, prediction is based on the projection of traffic state evolution starting from an initial state condition, which is usually obtained from a traffic state estimator. Therefore the estimation accuracy will significantly affect the prediction performance.

Although surveillance devices are measuring traffic along the highway, there are still many challenges in estimating the 'state' of the traffic. First due to the cost of surveillance equipment and the resulting data collection, it is not possible to have all the locations on the highway covered by sensors. There are segments where the sensors are sparsely installed. At these segments, the traffic state cannot be accurately estimated by the measurement information alone. In addition, the inductive loop detectors are not very reliable and the probability of malfunction is high. When the sensors are not working, even at the location covered by the sensors, traffic state cannot be estimated without algorithms. Moreover there is unavoidable noise in the measurement obtained by the sensors. Also, some state variables, like traffic density, are difficult to measure and have to be estimated from the measurement of traffic flow

and speed. If we simply use the fundamental relationship to estimate traffic density from volume and speed, the measurement error in both volume and speed will amplify the error in estimated density. Thus algorithms must be developed and applied to filter the measurement noise and thus to better estimate the traffic state. Lastly, for some traffic flow models, the parameters in the model change over time and cannot be measured by any kind of sensors, and must be estimated by specific methods and algorithms.

In this chapter, we focus on developing a traffic state estimation algorithm for the specific stochastic traffic flow model developed in Chapter II. We show that the discretized version of the stochastic traffic flow model can be reformulated as a state space model and various nonlinear Kalman filter algorithms can be applied. Since this traffic flow model is different from all the other models in the literature, we will evaluate the performance of various nonlinear Kalman filter algorithms. In addition, by comparing the prediction performance of the traffic flow model with and without the on-line estimation algorithm, we demonstrate that on-line estimation is able to improve the prediction accuracy of the future traffic state.

The outline of this chapter is described as follows: In Section 3.2, we present and formulate the research problem explicitly. In Section 3.3, various nonlinear Kalman filter algorithms are presented and discussed for general state space models. In Section 3.4, the specific state space model for the traffic state and parameters are defined to fit the Godunov scheme described in chapter II. The state space model is the basis for various Kalman filter algorithms. In Section 3.5, two schemes for estimating the traffic state and parameters, including joint estimation and dual estimation, are explained in detail. Section 3.6 validates the performance of the on-line estimation algorithms by using synthetic data. Section 3.7 concludes this chapter.

## 3.2 Problem Formulation

The objective of this chapter is to find an algorithm to recursively update the current traffic state and model parameters under the specific traffic flow model developed in Chapter II by using low frequency Eulerian measurements. Suppose the highway starts at 0 and ends at $L$. It is divided into $N$ small cells with $x_0 = 0$, $x_1$, ..., $x_{N-1}$, $x_N = L$, where $x_i$ $(i = 0, ..., N)$ is the boundary of the cell $i$. There are $S$ fixed inductive loop detectors spread on the highway, where sensor $s$ is located within cell $n_s$ $(n_s \in [1, N])$ to measure the speed and volume at this cell. These detectors can measure aggregated traffic volume and average speed at discrete time intervals, for example the measurement is available for every one minute. The speed and volume measured at these predefined locations consist of the Eulerian measurements for traffic state estimation. However the time interval of receiving measurement data is normally larger than the maximum possible time discretization interval required by equation (2.26) in Godunov's scheme. In other words, the measurement information comes in with relatively low frequency compared with what the state update equation requires in the numerical scheme.

The estimation targets include the traffic density in $N$ cells and the parameters in the traffic flow model including $a_i$ and $b_i$ $(i = 1, 2, ..., N)$. As discussed in chapter II, for the stochastic traffic flow model developed in that chapter, these parameters reflect the traffic volume entering and leaving the highway. The travel pattern is dynamic. In other words, at the locations close to residential area, in the morning there is more traffic volume entering than leaving the highway, while in the evening there is more traffic volume leaving than entering the highway. Therefore the parameters $a_i$ and $b_i$ are changing dynamically in time and must be estimated on-line. There are other parameters in the speed-density function. These parameters do not change much unless the structure of that highway section is modified, such as changing the speed limit or closing some lanes. However these events do not happen very frequently, so

we can assume that the parameters in the speed-density function are constant and they need not be estimated on-line. The parameters in the speed-density function can be estimated off-line by using the method described in Section 2.4.1.

In chapter II, we discussed an off-line calibration procedure to estimate the value of the parameters $a_i$ and $b_i$ $(i = 1, 2, ..., N)$ in each time period of the day. However there are several disadvantages of this procedure which makes it unsuitable for on-line calibration: 1. The off-line procedure incurs a huge computation cost which would take long time to tune the parameters. That is not suitable for on-line calibration due to the speed requirement of the on-line calibration. 2. The off-line procedure is not adaptive. It assumes that the parameters during one time period are constant over different days. This assumption is reasonable when there are no dramatic changes in the travel pattern. However it may fail if the situation today differs a lot from that in previous days. For example, if there is a big snow today, we cannot use the parameters estimated from previous days, when there is no snow, for the prediction of today's traffic state evolution. We need an on-line algorithm to capture the change of parameters quickly and adaptively. 3. The off-line procedure requires the storage of all the historical data for calibration purpose. However normally on-line calibration only requires the storage of current estimator, and keeps updating it while receiving new measurement information. This could avoid a lot of issues, such as reading/saving data into the database and so on.

In conclusion, the research objective of this chapter is to estimate the traffic density and the model parameters all over the highway on-line based on low frequency noisy measurement data from the fixed loop detectors. The on-line estimation is able to capture the changes of the model parameters more quickly and adaptively. The following sections will present the methods in detail.

## 3.3 State Space Models and Filtering

### 3.3.1 State Space Representation

The general representation of a state space model is described in equation (3.1) and (3.2). Equation (3.1) is called the state transition equation of the state space model, where $x_k$ is the state variable of the system at $k$th time instant and $\omega_k$ is the transition disturbance making the state evolution stochastic. From the state transition equation, we can see that the state variable in the current period is purely dependent on the previous state and the disturbance. Equation (3.2) is called observation equation, where $z_k$ is observation variable and $\epsilon_k$ is the measurement noise. In state estimation, the transition equation and observation are both given. The observation variable $z_k$ is measured at every time instant, the objective is to obtain the best estimation on $x_k$ based on historical observations of $z_k$.

$$x_k = f_k(x_{k-1}, \omega_k) \tag{3.1}$$

$$z_k = h_k(x_k, \epsilon_k) \tag{3.2}$$

### 3.3.2 Kalman Filter

When $f_k$, $h_k$ are both linear functions, for example as shown in equation (3.3) and (3.4), and $\omega_k$, $\epsilon_k$ are independent and identically distributed with normal distribution: $\omega_k \sim N(0, Q_k)$ and $\epsilon_k \sim N(0, R_k)$, the state space model is called linear and it is known that Kalman filter (*Kalman et al.*, 1960) will give the best estimation on $x_k$ under this situation.

$$x_k = F_k x_{k-1} + \omega_k \tag{3.3}$$

$$z_k = H_k x_k + \epsilon_k \tag{3.4}$$

Let $\hat{x}_{k|k}$ denote the posteriori estimation of state $x_k$ at time $k$ given observations up to and including at time $k$, and $P_{k|k}$ denote the posteriori error covariance matrix of the state estimation at time $k$; Let $\hat{x}_{k|k-1}$ denote the priori estimation of state $x_k$ at time $k-1$ given observations up to and including at time $k-1$, and $P_{k|k-1}$ denote the priori error covariance matrix of the state estimation at time $k-1$. Kalman filter is described in Algorithm 4. For detailed derivation, please refer to *Kalman et al.* (1960).

**Algorithm 4** Kalman filter

**Step 1: Initialization** Initialize with equation (3.5) and then set $k = 1$.

$$\begin{aligned} \hat{x}_{0|0} &= \mathrm{E}[x_0] \\ P_{0|0} &= \mathrm{E}[(x_0 - \hat{x}_{0|0})(x_0 - \hat{x}_{0|0})^T] \end{aligned} \tag{3.5}$$

**Step 2: Predict** Obtain the priori distribution of the state in the next period before receiving the observation information according to equation (3.6).

$$\begin{aligned} \hat{x}_{k|k-1} &= F_k \hat{x}_{k-1|k-1} \\ P_{k|k-1} &= F_k P_{k-1|k-1} F_k^T + Q_k \end{aligned} \tag{3.6}$$

**Step 3: Update** Update the posteriori distribution of the state after receiving the observation information according to equation (3.7)

$$\begin{aligned} y_k &= z_k - H_k \hat{x}_{k|k-1} \\ S_k &= H_k P_{k|k-1} H_k^T + R_k \\ K_k &= P_{k|k-1} H_k^T S_k^{-1} \\ \hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k y_k \\ P_{k|k} &= (I - K_k H_k) P_{k|k-1} \end{aligned} \tag{3.7}$$

**Step 4: Recursion** $k = k + 1$, go to Step 2.

Kalman filter gives the optimal estimation for linear state space model, however for the stochastic traffic flow model presented in Chapter II, the complexity of the model makes both $F_k$ and $H_k$ nonlinear and even implicit because of modified Godunov's scheme. Therefore nonlinear Kalman filter has to be applied. In the following sections,

we want to discuss several nonlinear filtering algorithms which will be subsequently used for on-line estimation.

### 3.3.3 Extended Kalman Filter

Extended Kalman filter (EKF) is a popular nonlinear filtering algorithm for nonlinear models like in equation (3.1) and (3.2) when both $f_k$ and $h_k$ are nonlinear functions. It has been widely used in practice because of its simplicity. The basic idea of EKF is linearizing the transition equation and measurement equation around the current estimated state, and then apply Kalman filter to estimate the approximated linear model. Compared with the procedure of Kalman filter in Algorithm 4, EKF is very similar except that there are slight differences in the steps: predict and update.

In EKF, suppose $\omega_k \sim N(0, Q_k)$ and $\epsilon_k \sim N(0, R_k)$, the predict equation is given in equation (3.8), where $F_k$ is the partial derivative of the transition equation with respect to state variable around $x = \hat{x}_{k-1|k-1}$ and $\omega = 0$, and $A_k$ is the partial derivative of the transition equation with respect to transition disturbance around the current state estimation.

$$\begin{aligned} F_k &= \frac{\partial f_k}{\partial x}(\hat{x}_{k-1|k-1}, 0) \\ A_k &= \frac{\partial f_k}{\partial \omega}(\hat{x}_{k-1|k-1}, 0) \\ \hat{x}_{k|k-1} &= f_k(\hat{x}_{k-1|k-1}, 0) \\ P_{k|k-1} &= F_k P_{k-1|k-1} F_k^T + A_k Q_k A_k^T \end{aligned} \tag{3.8}$$

For the update step, the measurement equation is linearized around $x = \hat{x}_{k|k-1}$ and $\epsilon = 0$. The update equation is given in equation (3.9), where $H_k$ is the partial derivative of the measurement equation with respect to state variable around $x = \hat{x}_{k|k-1}$ and $\epsilon = 0$, and $B_k$ is the partial derivative of the transition equation with

respect to measurement noise around the current state estimation.

$$
\begin{aligned}
H_k &= \frac{\partial h_k}{\partial x}(\hat{x}_{k|k-1}, 0) \\
B_k &= \frac{\partial h_k}{\partial \epsilon}(\hat{x}_{k|k-1}, 0) \\
y_k &= z_k - H_k \hat{x}_{k|k-1} \\
S_k &= H_k P_{k|k-1} H_k^T + B_k R_k B_k^T \\
K_k &= P_{k|k-1} H_k^T S_k^{-1} \\
\hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k y_k \\
P_{k|k} &= (I - K_k H_k) P_{k|k-1}
\end{aligned}
\tag{3.9}
$$

Extended Kalman filter is obtained by replacing the predict equation (3.6) in Algorithm 4 with equation (3.8), and the update equation (3.7) with equation (3.9). For detailed derivation of extended Kalman filter, please refer to *Ribeiro* (2004).

### 3.3.4 Unscented Kalman Filter

Unscented Kalman filter (UKF) is a nonlinear Kalman filter which has better performance than EKF when the transition equation is highly nonlinear. In EKF, the priori distribution of the state variable is approximated by a Gaussian distribution. This results in a poor approximation when the transition equation is highly non-linear. Moreover UKF does not calculate the derivative, making it more efficient than EKF in the case where the state variable is high dimensional. With those two advantages, UKF might be a better choice for the on-line traffic state estimation. We will discuss the selection of filters later.

The fundamental idea of UKF is unscented transform. Unscented transform is a mathematical sampling technique used to estimate the mean and covariance matrix of the output variable from a non-linear transformation of the input variable, given the mean and covariance matrix of the input variable. UKF first generates a set

of sample points (called sigma points), where each sample point is associated with a corresponding weight, from the mean and covariance matrix of the input random variable by using unscented transform. Each sigma point is considered as a realized input of the non-linear function. Then the set of sigma points is propagated through the non-linear function to generate a set of sample outputs, so that the mean and covariance matrix of the output variable could be estimated from these sample outputs.

The procedure to generate a set of sigma points from mean $\mu$ and covariance matrix $P$ is described as below: A Gaussian distribution with mean $\mu$ and covariance matrix $P$ is represented by a set of $2L+1$ sigma points $\chi^i$ $(i = 0, 1, ..., 2L)$, where $L$ is equal to the dimension of the state variable. $\sqrt{(L+\lambda)P}$ represents the Cholesky factor of $(L+\lambda)P$. In other words, the Cholesky factor $B$ of matrix $(L+\lambda)P$ should satisfy: $(L+\lambda)P = BB^T$. And $(\sqrt{(L+\lambda)P})_i$ represents the $i$th column of its Cholesky factor $B$. Then the set of $2L+1$ sigma points are generated according to equation (3.10), where $\chi^i$ represents the $i$th sigma point.

$$\begin{aligned} \chi^0 &= \mu \\ \chi^i &= \mu + (\sqrt{(L+\lambda)P})_i \quad i = 1, 2, ..., L \\ \chi^i &= \mu - (\sqrt{(L+\lambda)P})_{i-L} \quad i = L+1, L+2, ..., 2L \end{aligned} \tag{3.10}$$

$\lambda$ is a parameter determined by $\alpha$ and $\kappa$ as shown in equation (3.11), where $\alpha$ and $\kappa$ control the spread of sigma points. Normally $\alpha = 10^{-3}$ and $\kappa = 1$.

$$\lambda = \alpha^2(L+\kappa) - L \tag{3.11}$$

The weights associated with each sigma point are shown in equation (3.12), where $W_s^i$ is the weight associated with sigma point $i$ for mean estimation and $W_c^i$ is the

weight associated with sigma point $i$ for covariance estimation.

$$\begin{aligned} W_s^0 &= \frac{\lambda}{L+\lambda} \\ W_c^0 &= \frac{\lambda}{L+\lambda} + (1-\alpha^2+\beta) \\ W_s^i &= W_c^i = \frac{1}{2(L+\lambda)} \qquad i = 1, 2, ..., 2L \end{aligned} \tag{3.12}$$

The procedure of UKF is described in Algorithm 5. For detailed derivation of UKF, please refer to *Julier and Uhlmann* (2004).

### 3.3.5 Particle Filter

Particle filter (PF), which is also known as sequential Monte Carlo method, is another filter suitable for highly non-linear and complicated transition and measurement equations (*Doucet et al.*, 2001). Similar to UKF, PF approximates the posterior probability distribution of the state variable by a set of particles (sample points): Initially in the first step, a set of particles with equal weights are drawn from the prior probability distribution of the state. Then the particles evolve via Monte Carlo simulation according to the transition equation. Each particle represents one trajectory of the state variable and the weight represents the probability of the occurrence of this trajectory. At each time step when new information arrives, PF updates the weights of all the particles. When the number of particles is large enough, the updated discrete distribution of the particles approximates the posterior probability distribution of the state. A resampling procedure is used at some points of time in order to keep variety among the particles. The procedure of PF is described in Algorithm 6. For detailed derivation of particle filter, please refer to *Arulampalam et al.* (2002).

**Algorithm 5** Unscented Kalman filter

**Step 1: Initialization** Initialize with equation (3.13) and set $k = 1$.

$$\hat{x}_{0|0} = \mathrm{E}[x_0]; \quad P_{0|0} = \mathrm{E}[(x_0 - \hat{x}_{0|0})(x_0 - \hat{x}_{0|0})^T] \tag{3.13}$$

**Step 2: Predict** Define augmented variable $x^p = [x^T \ w^T]^T$ as the input of transition equation $F$. The mean and covariance are given in equation (3.14).

$$\hat{x}^p = [\hat{x}^T_{k-1|k-1} \ E[w_k]^T]^T; \quad P^p = \begin{bmatrix} P_{k-1|k-1} & 0 \\ 0 & Q_k \end{bmatrix} \tag{3.14}$$

Generate a set of $2L + 1$ sigma points $\chi^i_p$ from $\hat{x}^p$ and $P^p$ by using unscented transform. Derive the priori mean and covariance from sample output $\mathcal{X}^i_{k|k-1}$.

$$\mathcal{X}^i_{k|k-1} = f_k(\chi^{i,x}_p, \chi^{i,w}_p) \qquad i = 0, 1, ..., 2L \tag{3.15}$$

$$\hat{x}_{k|k-1} = \sum_{i=0}^{2L} W^i_s \mathcal{X}^i_{k|k-1}; \quad P_{k|k-1} = \sum_{i=0}^{2L} W^i_c [\mathcal{X}^i_{k|k-1} - \hat{x}_{k|k-1}][\mathcal{X}^i_{k|k-1} - \hat{x}_{k|k-1}]^T$$

**Step 3: Update** Define augmented variable $x^u = [x^T \ v^T]^T$ as the input of measurement equation $h_k$. The mean and covariance are given in equation (3.16).

$$\hat{x}^u = [\hat{x}^T_{k|k-1} \ E[v_k]^T]^T; \quad P^u = \begin{bmatrix} P_{k|k-1} & 0 \\ 0 & R_k \end{bmatrix} \tag{3.16}$$

Generate $2\tilde{L}+1$ sigma points $\chi^i_u$ from $\hat{x}^u$ and $P^u$ by using unscented transform. Derive the posteriori mean and covariance from sample output $\mathcal{Y}^i_{k|k-1}$. And update the posteriori mean and covariance matrix.

$$\mathcal{Y}^i_{k|k-1} = h_k(\chi^{i,x}_u, \chi^{i,v}_u) \qquad i = 0, 1, ..., 2\tilde{L} \tag{3.17}$$

$$\hat{y}_{k|k-1} = \sum_{i=0}^{2\tilde{L}} W^i_s \mathcal{Y}^i_{k|k-1}; \quad P_{y_k y_k} = \sum_{i=0}^{2\tilde{L}} W^i_c [\mathcal{Y}^i_{k|k-1} - \hat{y}_{k|k-1}][\mathcal{Y}^i_{k|k-1} - \hat{y}_{k|k-1}]^T$$

$$P_{x_k y_k} = \sum_{i=0}^{2\tilde{L}} W^i_c [\chi^{i,x}_u - \hat{x}_{k|k-1}][\mathcal{Y}^i_{k|k-1} - \hat{y}_{k|k-1}]^T; \quad K_k = P_{x_k y_k} P^{-1}_{y_k y_k}$$

$$\begin{aligned} \hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k(y_k - \hat{y}_{k|k-1}) \\ P_{k|k} &= P_{k|k-1} - K_k P_{y_k y_k} K^T_k \end{aligned} \tag{3.18}$$

**Step 4: Recursion** $k = k + 1$, go to Step 2.

**Algorithm 6** Particle filter

**Step 1: Initialization** Draw a set of $N$ particles $x_0^i$ $(i = 1, \cdots, N)$ from the prior initial distribution $p(x_0)$. The weights of all particle are equal as shown in equation (3.19). Then set $k = 1$.

$$w_0^i = \frac{1}{N} \tag{3.19}$$

**Step 2: Predict** Simulate the evolution of every particle and obtain the set of particles in the next time period according to equation (3.20).

$$x_k^i \sim p(x_k|x_{k-1}^i) \quad i = 1, 2, \cdots, N \tag{3.20}$$

**Step 3: Update** Update the weights of all the particles according to equation (3.21), and then normalize the weights according to equation (3.22). The optimal estimator is given as: $\hat{x}_k = \sum_{i=1}^{N} w_k^i x_k^i$.

$$\tilde{w}_k^i = w_{k-1}^i p(z_k|x_k^i) \tag{3.21}$$

$$w_k^i = \frac{\tilde{w}_k^i}{\sum_{i=1}^{N} \tilde{w}_k^i} \tag{3.22}$$

**Step 4: Resampling** Resampling is used when $N_{eff} < N_{thr}$ to avoid the problem of degeneracy: Resample $N$ particles $x_k^i$ $(i = 1, 2, ..., N)$ from current particle set with probabilities proportional to their weights. Replace the particle set with the new one with equal weights: $w_k^1 = w_k^2 = ... = w_k^N = 1/N$.

$$N_{eff} = \frac{1}{\sum_{i=1}^{N} [(w_k^i)]^2} \tag{3.23}$$

**Step 5: Recursion** $k = k + 1$, go to Step 2.

## 3.4 State Space Formalization of the Traffic Flow Model

### 3.4.1 State Transition Equation

As discussed in Chapter II, for the continuous-time stochastic traffic flow model given in equation (3.24), Algorithm 2 shown in Section 2.3.2.3 describes the numerical procedure of simulating the evolution of the traffic density. In this numerical procedure, the highway is divided into $N$ cells and the time is also discretized by $\triangle t$. At any time, the traffic density on the highway could be represented by the density vector of the $N$ cells. The numerical procedure of Algorithm 2 updates the traffic density at every time instant of $\triangle t$. As required by CFL condition shown in equation (2.26), $\triangle t$ must be small enough to avoid shock wave within the cell. Suppose the measurement information is received at every time instant of $T$. Normally $T$ is greater than or equal to one minute, which is normally larger than $\triangle t$. However if we want to formulate a state space model of the traffic flow model for filtering purpose, the time interval of the state space model should be equal to $T$ not $\triangle t$, and the numerical procedure in Algorithm 2 cannot be directly used as the state transition equation of the state space model.

$$\begin{aligned} \rho_t + q_x &= g(\rho, x, t) \\ g(\rho, x, t) \cdot dx \cdot dt &= (a(x,t) + b(x,t) \cdot \rho) \cdot dx \cdot dt + \sigma(x,t) \cdot dW(x,t) \end{aligned} \tag{3.24}$$

Algorithm 2 can be used to derive the state transition equation. Since each application of Algorithm 2 gives the traffic density $\triangle t$ units of time later, running it $T/\triangle t$ times will give the traffic density $T$ units of time later, which is the state transition equation.

Since the time interval of the state space model is $T$, let $\rho_k^i$ denote the traffic density of cell $i$ at time $kT$, and $\rho_k$ denote the density vector at time $kT$ as shown in

equation (3.25).

$$\rho_k = [\rho_k^1,\ \rho_k^2,\ \cdots,\ \rho_k^N]^T \tag{3.25}$$

Like the traffic density vector, the parameter vector can also be discretized. Let $\pi_k$ denote the parameter vector of the system between time $(k-1)T$ and $kT$. As discussed in Section 3.2, the parameter vector $\pi_k$ includes $a_k^i$ and $b_k^i$ $(i = 1, 2, \cdots, N)$ as shown in equation (3.26), where $a_k^i$ and $b_k$ represent the inflow/outflow parameters of cell $i$ between $(k-1)T$ and $kT$.

$$\pi_k = [\ a_k^1,\ b_k^1, \cdots,\ a_k^N,\ b_k^N]^T \tag{3.26}$$

Given $\rho_{k-1}$ and $\pi_k$, by running the numerical procedure in Algorithm 2, $T/\triangle t$ times, we can obtain $\rho_k$ as presented in equation (3.27), where $\omega_k$ is the disturbance noise. $\omega_k$ consists of $N\frac{T}{\triangle t}$ i.i.d standard normal random variables because each step has $N$ i.i.d disturbance random variables and there are $\frac{T}{\triangle t}$ steps in total. $F$ is the state transition equation of $\rho_k$. It has no closed form because of the complexity of the numerical procedure. But this transition equation exists and Algorithm 2 explains how it works.

$$\rho_k = F(\rho_{k-1}, \pi_k, \omega_k) \tag{3.27}$$

As discussed in Section 3.2, the parameters also appear to be unknown and stochastic. The estimation of the parameters will be as important as estimating the state density, because the evolution of the traffic density depends on the parameters. If the estimated parameters deviate from their true values, the priori distributions of both the traffic density and observed variable will have large errors. Then the Bayesian based filtering algorithms will not work correctly, since an accurate priori

distribution is required to make accurate estimations.

In order to estimate the parameters as well, an assumption about the dynamic evolution of the parameters has to be made. Normally, the parameters $\pi_k$ are assumed to follow a random walk as shown in equation (3.28), where $\xi_k$ follows normal distribution.

$$\pi_k = \pi_{k-1} + \xi_k \tag{3.28}$$

In conclusion, equation (3.27) and (3.28) describe the transition equation of traffic density and parameters respectively. They constitute the state transition equation of the state space model.

### 3.4.2 Measurement Equation

Measurements can be obtained from many different sources. For example, the loop detectors can measure the volume and speed of the cells. Let $v_k^i$ and $q_k^i$ denote the observed speed and volume in cell $i$ at time $kT$ respectively. Since the measurement is noisy, the measured speed and volume will deviate from their true values. Let $\epsilon_k^{v,i}$ denote the noise for the measurement of $v_i^k$ and $\epsilon_k^{q,i}$ denote the noise for the measurement of $q_i^k$. So the measurement equation for fixed loop detectors is given in equation (3.29), where $V$ is the speed-density function and $Q$ is the volume-density function.

$$\begin{aligned} v_k^i &= V(\rho_k^i) + \epsilon_k^{v,i} \\ q_k^i &= Q(\rho_k^i) + \epsilon_k^{q,i} \end{aligned} \tag{3.29}$$

The loop detectors are not placed in all cells. Let $\Omega$ denote the set of cells which

have detectors. Define $z_k$ as the observation variable shown in equation (3.30).

$$z_k = [v_k^{i_1}, q_k^{i_1}, v_k^{i_2}, q_k^{i_2}, \cdots, v_k^{i_{|\Omega|}}, q_k^{i_{|\Omega|}}] \quad i_1, i_2, \cdots, i_{|\Omega|} \in \Omega \tag{3.30}$$

Then we get the observation equation (3.31), where $\epsilon_k$ represents the measurement noise defined as $\epsilon_k = [\epsilon_k^{v,i_1}, \epsilon_k^{q,i_1}, \epsilon_k^{v,i_2}, \epsilon_k^{q,i_2}, \cdots, \epsilon_k^{v,i_{|\Omega|}}, \epsilon_k^{q,i_{|\Omega|}}, ]$.

$$z_k = H(\rho_k, \epsilon_k) \tag{3.31}$$

## 3.5 Online Calibration of the Traffic Flow Model

By combining the transition equation (3.27), (3.28) and measurement equation (3.31), a complete state space model for the stochastic traffic flow model can be formulated as equation (3.32)

$$\begin{aligned} \rho_k &= F(\rho_{k-1}, \pi_k, \omega_k) \\ \pi_k &= \pi_{k-1} + \xi_k \\ z_k &= H(\rho_k, \epsilon_k) \end{aligned} \tag{3.32}$$

Given the model in equation (3.32), there are two approaches to combine the estimation of traffic density and the model parameters. The first approach is called joint estimation. In this approach, the model parameters and traffic density are combined together as the augmented state variable. Then the traffic density and the model parameters can be estimated jointly by estimating the augmented state variable at every time step. The other approach is called dual estimation. In this approach, the traffic density and the parameters are estimated sequentially at each time step. In other words, each step has two phases; In phase one, only the model parameters are updated. In phase two, the traffic density is updated by using the newly up-

dated model parameters. The details of the algorithms for these two approaches are introduced in the following sections.

### 3.5.1 Joint Estimation

The joint estimation of the traffic density and model parameters is as the same as normal nonlinear Kalman filter except that the augmented state variable is a combination of the traffic density and the model parameters. The detail is given in Algorithm 7.

---

**Algorithm 7** Joint estimation for traffic density and model parameters

---

**Step 1: Initialization** Define an augmented state variable $x$ to be the combination of the traffic density and the parameters:

$$x_k = [\rho_k^T \ \pi_k^T]^T$$

Initialize with equation (3.33) and set $k = 1$.

$$\begin{aligned} &\hat{\rho}_{0|0} = \mathrm{E}[\rho_0]; \quad P^{\rho}_{0|0} = \mathrm{E}[(\rho_0 - \hat{\rho}_{0|0})(\rho_0 - \hat{\rho}_{0|0})^T] \\ &\hat{\pi}_{0|0} = \mathrm{E}[\pi_0]; \quad P^{\pi}_{0|0} = \mathrm{E}[(\pi_0 - \hat{\pi}_{0|0})(\pi_0 - \hat{\pi}_{0|0})^T] \\ &\hat{x}_{0|0} = [\hat{\rho}^T_{0|0} \ \hat{\pi}^T_{0|0}]^T; \quad P_{0|0} = \begin{bmatrix} P^{\rho}_{0|0} & 0 \\ 0 & P^{\pi}_{0|0} \end{bmatrix} \end{aligned} \tag{3.33}$$

**Step 2: Update the augmented state** Choose a filter applying to the state space model below to update $\hat{x}_{k|k}$ and $P_{k|k}$:

$$\begin{aligned} x_k &= \begin{bmatrix} \rho_k \\ \pi_k \end{bmatrix} = \begin{bmatrix} F(\rho_{k-1}, \pi_{k-1} + \xi_k, \omega_k) \\ \pi_{k-1} + \xi_k \end{bmatrix} \\ z_k &= H(x_k, \epsilon_k) \end{aligned}$$

**Step 3:** $k = k + 1$, go to Step 2.

---

### 3.5.2 Dual Estimation

**Algorithm 8** Dual estimation for traffic density and model parameters

**Step 1: Initialization** Initialize with equation (3.34) and set $k = 1$.

$$\begin{aligned} \hat{\rho}_{0|0} &= \mathrm{E}[\rho_0]; \quad P^{\rho}_{0|0} = \mathrm{E}[(\rho_0 - \hat{\rho}_{0|0})(\rho_0 - \hat{\rho}_{0|0})^T] \\ \hat{\pi}_{0|0} &= \mathrm{E}[\pi_0]; \quad P^{\pi}_{0|0} = \mathrm{E}[(\pi_0 - \hat{\pi}_{0|0})(\pi_0 - \hat{\pi}_{0|0})^T] \end{aligned} \tag{3.34}$$

**Step 2: Update the parameters** Plug $\hat{\rho}_{k-1|k-1}$ as a constant into the state space model below, and choose a filter to update $\hat{\pi}_{k|k}$ and $P^{\pi}_{k|k}$ by considering $\pi_k$ as the state variable.

$$\begin{aligned} \pi_k &= \pi_{k-1} + \xi_k \\ z_k &= H(F(\hat{\rho}_{k-1|k-1}, \pi_k, \omega_k), \epsilon_k) \end{aligned}$$

**Step 3: Update the state** Plug $\hat{\pi}_{k|k}$ as a constant into the state space model below, and choose a filter to update $\hat{\rho}_{k|k}$ and $P^{\rho}_{k|k}$ by considering $\rho_k$ as the state variable.

$$\begin{aligned} \rho_k &= F(\rho_{k-1}, \hat{\pi}_{k|k}, \omega_k) \\ z_k &= H(\rho_k, \epsilon_k) \end{aligned}$$

**Step 4:** $k = k + 1$, go to Step 2.

The dual estimation is different from the joint estimation. It separates the estimation of the traffic density and model parameters. Therefore two state space models are formulated, one for traffic density and the other for the model parameters. And each is used independently to estimate the traffic density or model parameters. A

property of the state space model in (3.32) is that the state transition of $\rho_k$ depends on $\pi_k$, whose estimation is unknown at $k-1$ time instant. Therefore we first estimate $\pi_k$ by using the previous estimation of traffic density: $\hat{\rho}_{k-1|k-1}$. After that, $\rho_k$ can be estimated by using the updated estimation of the model parameters: $\hat{\pi}_{k|k}$. The detail is given in Algorithm 8.

## 3.6 Numerical Results

In this section, the numerical experiments are mainly focused on answering three questions: First, can the filtering algorithms improve the accuracy of the traffic state estimation in real-time if the model parameters are known? Second, can the filtering algorithms improve the estimation accuracy of both the traffic state and model parameters in real-time if the model parameters are unknown and stochastic? Third, can the filtering algorithms help the stochastic traffic flow model improve the accuracy of the future traffic state prediction if the model parameters are unknown and stochastic? The following three subsections will answer these questions respectively.

### 3.6.1 Traffic State Estimation

This subsection will answer the first question. We use synthetic traffic flow data generated by simulation to evaluate the performance of UKF and PF in terms of their abilities to estimate the true traffic state in all the cells over the highway from noisy measurement data in a subset of the cells. It is assumed that the model parameters are known in this numerical experiment.

The synthetic traffic flow data is generated according to the steps in Algorithm 2 in Section 2.3.2.3. The model parameters used in the simulation are shown in table 3.1: The piecewise function in Section 2.2.3 is chosen as the speed-density function for this simulation. $v_f$, $\rho_c$ and $m$ are the three parameters in the speed-density function, where $\rho_c = (v_f/\alpha)^{1/m}$. The length of the highway is 6 miles and it is equally divided

| Parameters | Values |
|---|---|
| Free flow speed | $v_f = 65$ miles/hour |
| Critical density | $\rho_c = 60.87$ vehicles/mile |
| Speed-density function parameter | $m = -1.2$ |
| Number of cells | $N = 10$ |
| Measurement cell set | [1 4 7 10] |
| Cell length | $\triangle x = 0.6$ miles |
| Time interval for simulation | $\triangle t = 30$ s |
| Time interval for measurement | $T = 60$ s |
| Speed measurement noise | $Var[\epsilon_k^{v,i}] = 4^2$ (miles/hour)$^2$ |
| Volume measurement noise | $Var[\epsilon_k^{q,i}] = 200^2$ (vehicles/mile)$^2$ |

Table 3.1: Parameters in the experiment for traffic state estimation.

into 10 cells with the length of each cell equal to 0.6 miles. In the simulation of the stochastic traffic flow model, the time is discretized by the time interval of 30 seconds, while the measurement is taken for every 60 seconds. The measurement data consists of the speed and volume measurement only in those cells included in the measurement cell set. $\epsilon_k^{v,i}$ is the noise of every speed measurement and $\epsilon_k^{q,i}$ is the noise of every volume measurement.

The total length of the simulation will be 3 hours. The forcing function parameters $a$ and $b$ in the stochastic traffic flow model in the following cells: 1, 3, 6 and 9, are assumed to follow the evolution as shown in Figure 3.1. For the other cells, $a$ and $b$ are constantly equal to 0. $\sigma$ is assumed to follow the evolution shown in Figure 3.1 in all the cells.

In the simulation, for every 60 seconds, the volume and speed of the cells included in measurement cell set will be measured with noise, and the noisy measurement will be plugged into UKF and PF to update the estimation of current traffic state over all the cells. Figure 3.2 compares the true state with the estimation obtained by UKF and PF as well as the simple estimation without any filtering in those cells included in the measurement cell set. The simple estimation of the traffic density without any filtering is directly calculated by dividing the measured volume with measured speed

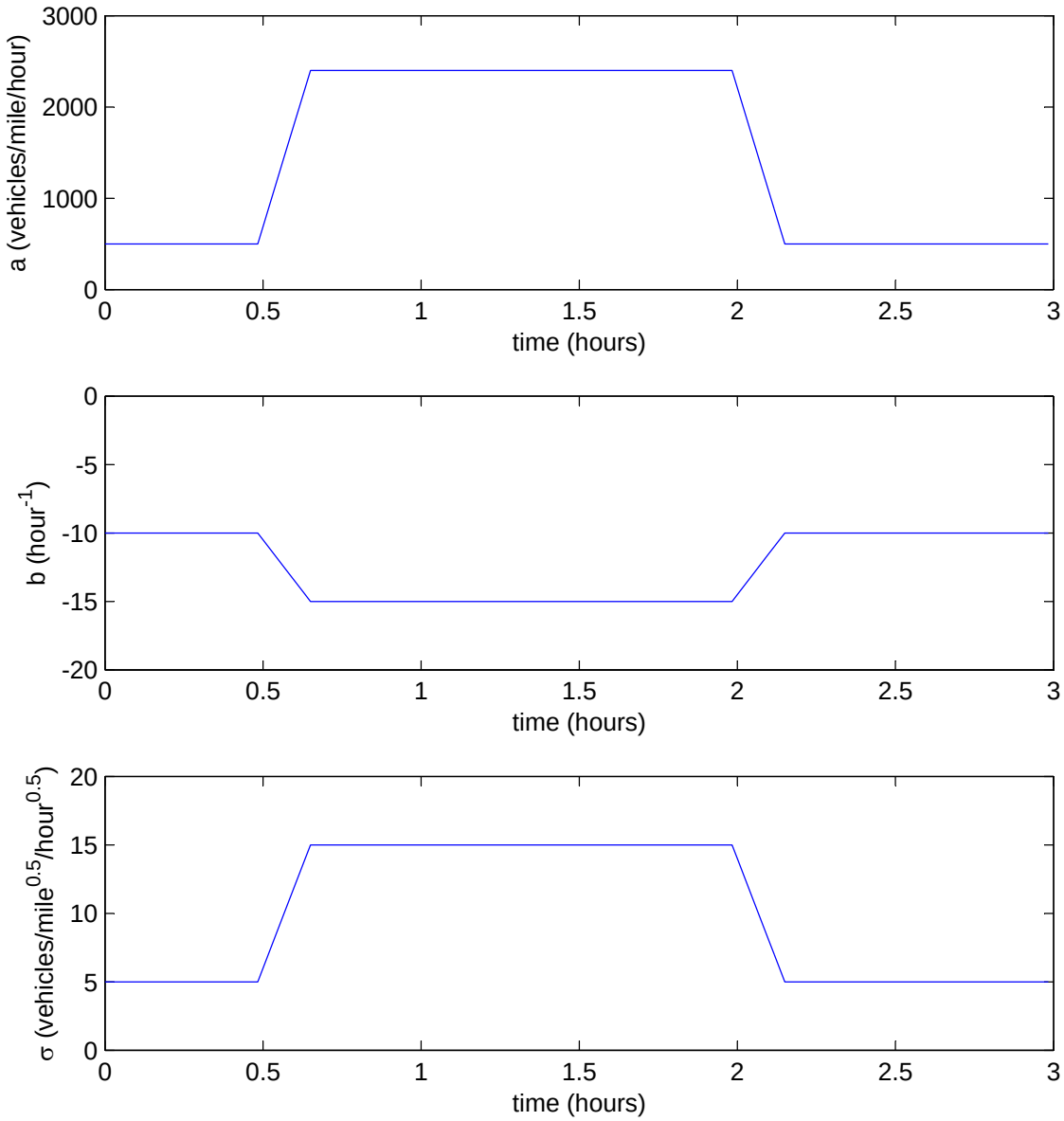



Figure 3.1: Evolution of the forcing function parameters

according to the fundamental relationship in equation (2.1). From the figure, we can see that both PF and UKF track with the true state very closely in all the cells, while the simple estimation without any filtering has comparatively larger error than PF and UKF in cell 1, 4 and 7 and similar performance with UKF and PF in cell 10. The reason why the performance of simple estimation is much worse in cell 1, 4 and 7 than in cell 10 is because the traffic density in cell 10 never gets congested. When the

speed is low due to the high density in that location, the measurement noise in the speed will have much higher impact to the estimated density than the situation when the speed is high. In addition, the measurement noise of speed will also be increasing as the density increases due to the fact that the density is not a continuous variable. Therefore, when traffic gets congested, the simple estimation will not work properly and filtering algorithms, such as UKF and PF, can preserve the estimation error.

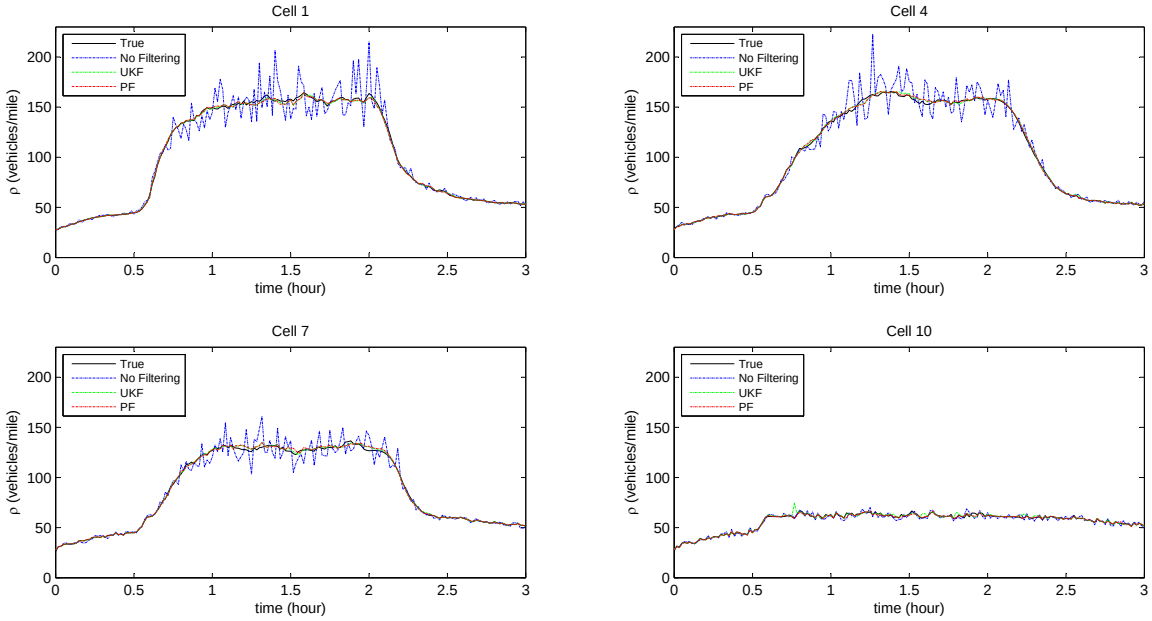



Figure 3.2: The result of traffic state estimation given known model parameters

Figure 3.3 compares the three estimation algorithms quantitatively using the RMSE between estimated density and the true density. From the figure, we can see that the performance of UKF and PF are pretty stable in all locations. However the performance of the simple estimation depends on the traffic congestion level. When the traffic is heavily congested, the error of simple estimation is much larger than that of UKF and PF. When the traffic is under free flow condition, the error of simple estimation is similar to that of UKF and PF.

In conclusion, UKF and PF can improve the estimation accuracy of the traffic state given the model parameters are known especially when the traffic is congested.

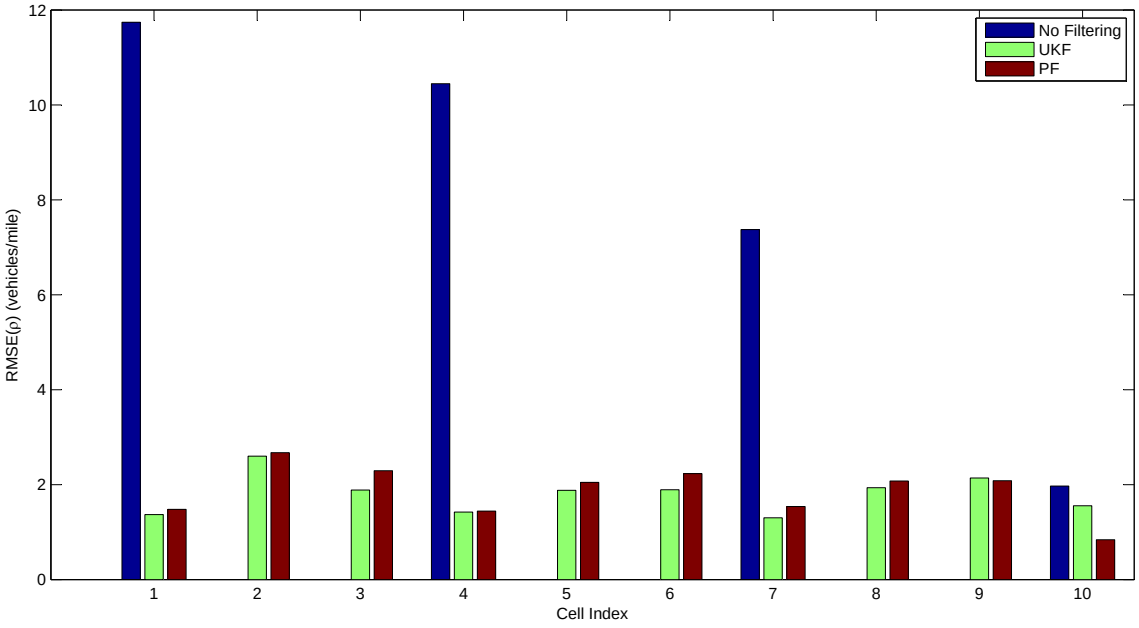



Figure 3.3: RMSE comparison of the density estimation given known model parameters

### 3.6.2 Traffic State and Model Parameters Estimation

In Section 3.6.1, we have proved the advantage of UKF and PF on estimating the traffic state given that the model parameters are known. In this experiment, it will be assumed that the model parameters $a$ and $b$ evolve stochastically, and the performance of UKF and PF on simultaneously estimating the traffic state and model parameters will be investigated. As described in Section 3.5, there are two types of schemes to estimate traffic state and model parameters simultaneously: joint estimation and dual estimation. Combining with UKF and PF, there are 4 filtering algorithms: Joint UKF, Joint PF, Dual UKF and Dual PF. These 4 filtering algorithms will be discussed and compared with simple estimation without any filtering.

The traffic state and model parameters are following the state space model in (3.32). Similar to the steps in Section 3.6.1, the traffic state is simulated according to Algorithm 2 in Section 2.3.2.3. The parameters used in this numerical experiment are summarized in Table 3.2.

| Parameters | Values |
|---|---|
| Free flow speed | $v_f = 65$ miles/hour |
| Critical density | $\rho_c = 60.87$ vehicles/mile |
| Speed-density function parameter | $m = -1.2$ |
| Number of cells | $N = 10$ |
| Measurement cell set | [1, 2, 3, 4, 5, 6 7, 8, 9, 10] |
| Cell length | $\triangle x = 0.6$ miles |
| Time interval for simulation | $\triangle t = 30$ s |
| Time interval for measurement | $T = 60$ s |
| Transition disturbance for $a$ | $Var[\xi_k^{a,i}] = 400^2$ (vehicles/mile/hour)$^2$ |
| Transition disturbance for $b$ | $Var[\xi_k^{b,i}] = 0.05^2$ (hour$^{-1}$)$^2$ |
| Speed measurement noise | $Var[\epsilon_k^{v,i}] = 4^2$ (miles/hour)$^2$ |
| Volume measurement noise | $Var[\epsilon_k^{q,i}] = 200^2$ (vehicles/mile)$^2$ |

Table 3.2: Parameters in the experiment for traffic state and model parameters estimation.

Figure 3.4 shows the evolution of true traffic state and the estimated traffic state obtained by different estimation algorithms in all the cells. Figure 3.5 quantitatively evaluates the performance of these estimation algorithms using the RMSE between the true traffic state and estimated traffic state. From the figures, we can see that all of those 4 filtering algorithms: Joint UKF, Dual UKF, Joint PF and Dual PF, track the true state closely in every location. While the simple estimation algorithm only give acceptable performance in the cells where the traffic is not congested. That also supports the conclusion in Section 3.6.1.

Figure 3.6 shows the performance of 4 different algorithms on tracking the change of model parameters. In this experiment, it is assumed that parameter $a$ is constantly equal to 0 except for cell 1, 5 and 9. For cell 1, 5 and 9, parameter $a$ jumps up at around 0.5 hour and jumps down at 2 hours. Parameter $b$ is constantly equal to 0 except for cell 3 and 7. For cell 3 and 7, parameter $b$ jumps down at around 0.5 hour and jumps up at 2 hours. Parameter $\sigma$ is constantly equal to 10 for all the cells. The experiment will investigate whether the filtering algorithms could capture the jumps of the model parameters. From the result shown in Figure 3.6, we can see

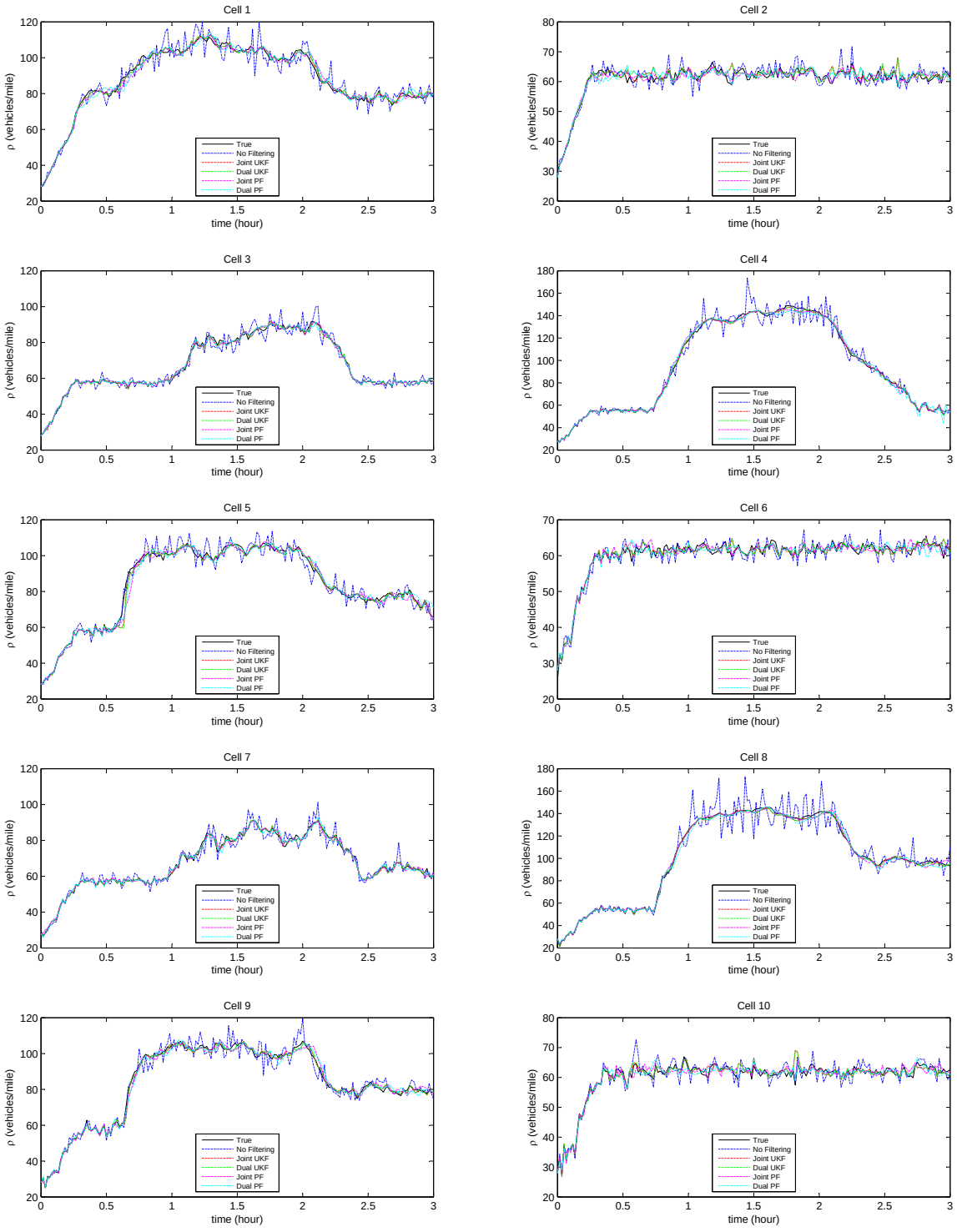



Figure 3.4: Evolution of true and estimated traffic state in cells

that the filtering algorithms do capture the jumps in model parameters. Therefore in conclusion, filtering algorithms can improve the estimation accuracy of both the

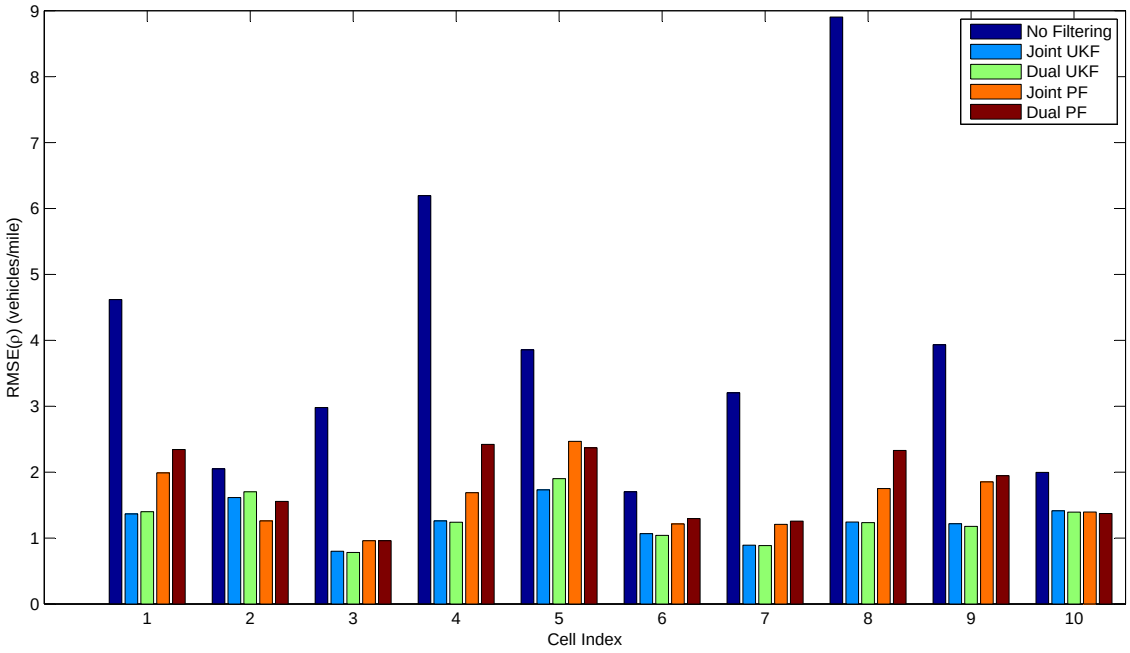



Figure 3.5: RMSE comparison of the density estimation given unknown model parameters

traffic state and model parameters when the model parameters are unknown and stochastic.

### 3.6.3 Filtering Based Traffic Prediction

This section is an extension to Section 3.6.2. In this section, we will investigate whether the filtering algorithms can help improve the accuracy of future traffic state prediction. The basic idea is plugging the estimated traffic state and model parameters obtained by the filtering algorithms in the prediction algorithm in Section 2.3.3 to output the prediction of the traffic state 5 minutes later.

Figure 3.7 presents the evolution of true traffic state and predicted traffic state under different filtering algorithms and simple estimation. In simple estimation, the traffic state is estimated according to the fundamental relationship while the model parameters are estimated using the initial condition, since the simple estimation does not have the ability to update the model parameters on-line. Figure 3.8 quantita-

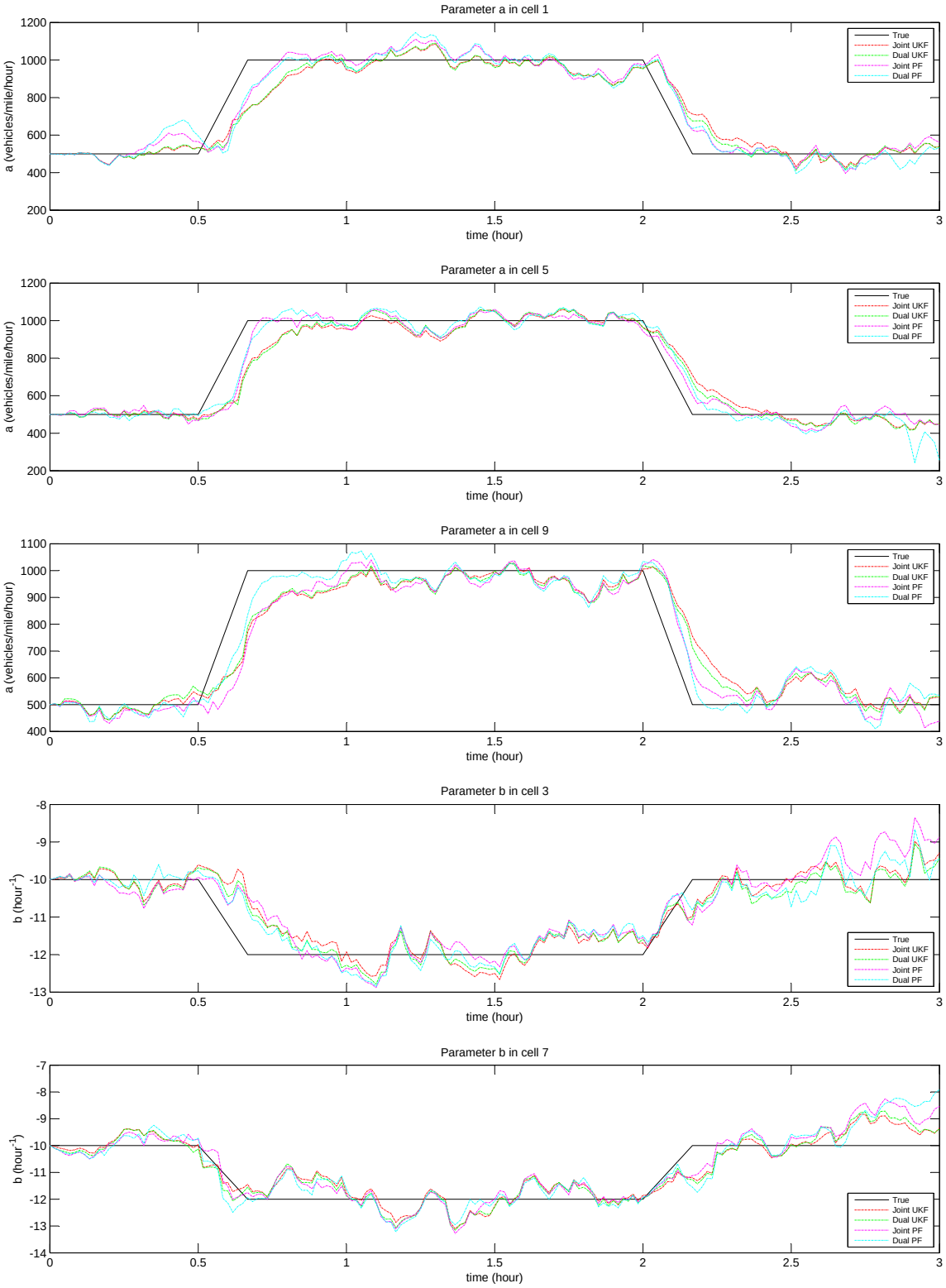



Figure 3.6: Evolution of true and estimated model parameters

tively compares the prediction accuracy of different estimation algorithms by using the RMSE between the true traffic state and predicted traffic state. From the figures, we can see that the performance of the four filtering algorithms outperforms simple estimation algorithm significantly. Therefore in conclusion, the filtering algorithms can improve the prediction accuracy of the future traffic state when the model parameters are unknown and stochastic.

## 3.7 Conclusion

In this chapter, we first introduce state space modeling and various filtering algorithms for on-line state estimation, such as unscented Kalman filter and particle filter. Based on that, we developed the on-line calibration algorithm for the stochastic traffic flow model developed in Chapter II. The on-line calibration algorithms can estimate the traffic density and capture the change of model parameters in real-time. We present some numerical results to show the performance of the on-line calibration algorithms. From the result, we can see that the on-line calibration algorithms are able to track the traffic density and model parameters simultaneously. In addition, the prediction accuracy can also be improved by using the estimated traffic density and model parameters by the on-line calibration algorithm.

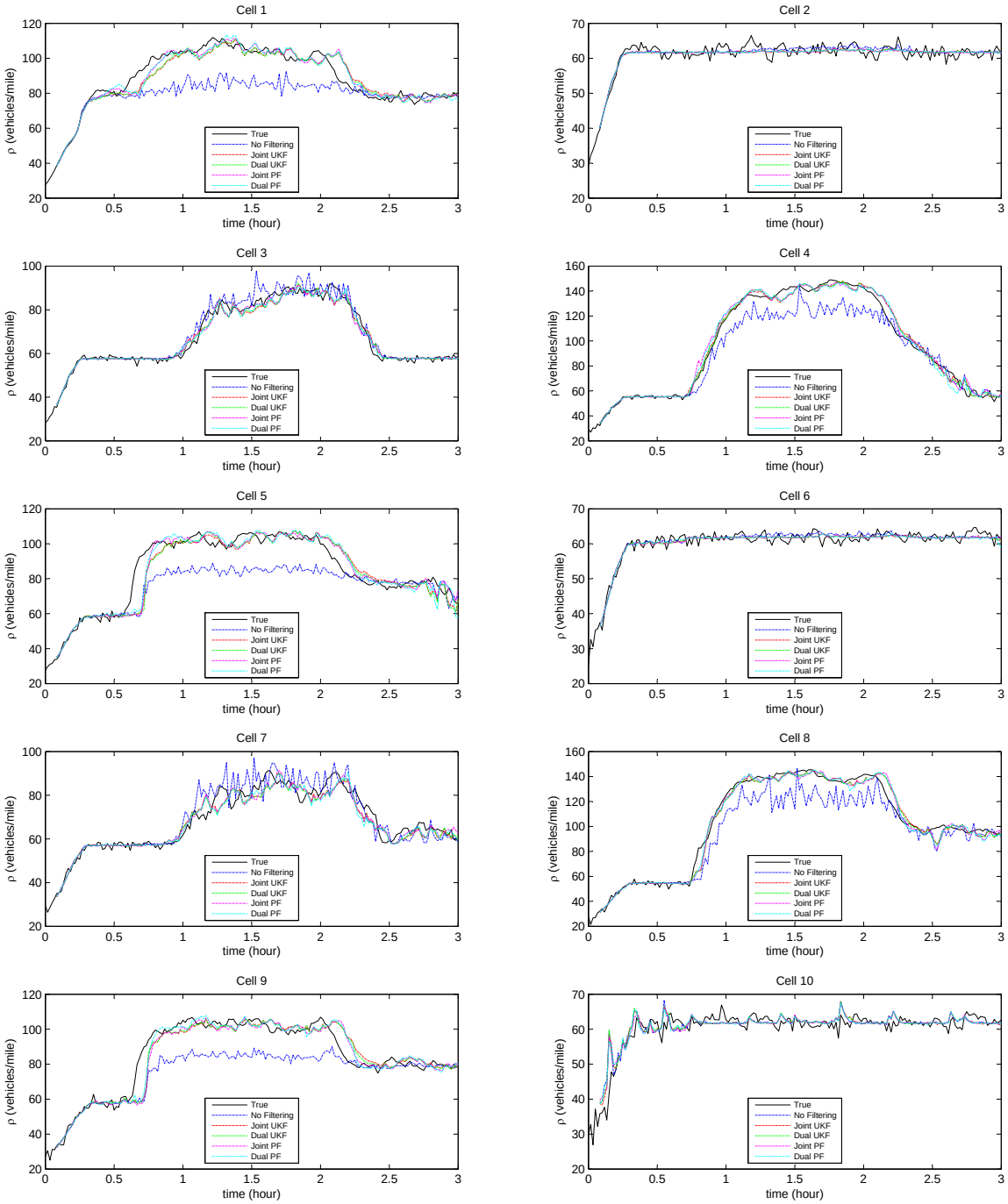



Figure 3.7: Evolution of true and predicted traffic state in cells

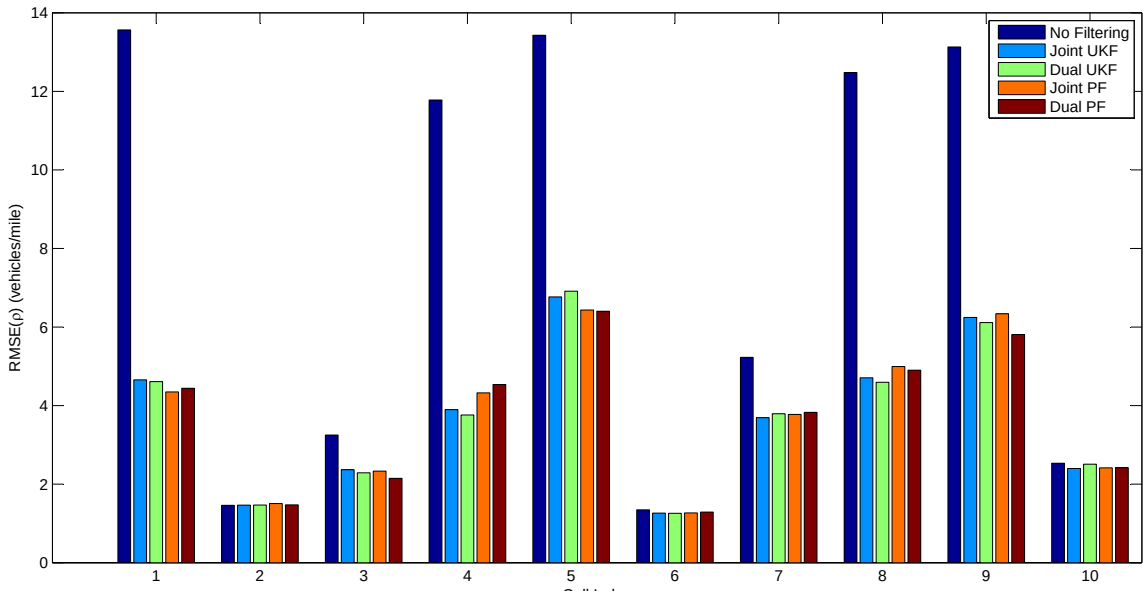



Figure 3.8: RMSE comparison of traffic state prediction for different filtering algorithms

# CHAPTER IV

# Dynamic Pricing For Managed Toll Lanes

## 4.1 Introduction

Managed lanes are defined as the highway lanes that are operated under fixed or real-time dynamic strategies to achieve a variety of objectives. The objectives include but not limited to improving facility utilization efficiency, controlling congestion levels and increasing the return on investment. Strategies for managing these kinds of lanes can be roughly classified into three categories: eligibility control, access control and pricing. Eligibility control strategies set restrictions (such as high occupancy of vehicle) on vehicles to enter the managed lanes. Different from eligibility control strategies, access control strategies allow all vehicles to enter but the access rate is controlled. For example, continuous monitoring of the traffic on the ramp is one type of access control strategies. On the other hand, pricing strategy charges the drivers a toll to enter the managed lane. A managed lane implementing pricing strategy is also called a managed toll lane. In the United States, increasing investments are being made in managed toll lanes because of their obvious effectiveness on maintaining the level of service on the road and their revenue generating ability, which helps pay off the investors or finance the highway construction. A managed toll lane provides an alternative for the drivers to avoid congestion on the general purpose lane. A driver pays an amount of money to enter the managed toll lane, and the toll lane guarantees

that the driver will travel under free flow speed. This not only saves the paying driver time but also decreases the load on the congested road and helps clear the congestion more effectively.

The strategy to price tolls on managed toll lanes is an important problem. There are different pricing strategies for setting the toll and they are highly dependent on the objective of the toll lane administers, i.e., if the toll lane is owned by a private company, the administrator of the company might want to maximize the revenue from the toll lane in order to earn as much profit as possible; Or, if the toll lane is built by the government, the objective might be to minimize the total travel time or to maximize the total throughput. In recent years, research in congestion pricing strategy has received much more attention from both the academic professionals and the practitioners.

The toll pricing scheme can be classified according to the toll collection base and rate patterns. There are three types of toll collections: pass-based, per use-based and distance-based. Also there are three types of rate patterns: flat rate, time-of-day rate and dynamic rate. Table 4.1 reproduced from *Chung and Recker* (2011) summarizes the existing HOT facilities and their pricing strategies in the United States. Pass-based toll collection scheme issues a toll lane pass to the drivers, and the vehicles with pass can enter the toll lane at any time. It can be either flat rate or dynamic rate. In flat rate, the price of the pass is constant while in a dynamic rate, the price of the pass is adjusted in each month. This is the simplest toll collection scheme. However it is not a good traffic control strategy since once the pass is issued, there is no restriction on the entry to the toll lane for the vehicles with a pass even when the toll lane is congested. Therefore pass-based toll collection is not adaptive and is thus not suitable for real-time traffic control.

Table 4.1: Overview of the existing HOT facilities in the United States

| Location | HOT Configuration | Toll Policy | Toll Pattern and Range |
|---|---|---|---|
| I-15 Salt Lake City UT | 45.6-mile 2 lanes. Midway accessible. 8 general lanes | HOV2+:free. SOV: tolled | Monthly pass: $50 per month. |
| US-290 Houston TX | 14-mile reversible 1 lane. Midway inaccessible. 8 general lanes. | Peak (HOT lane)<br>HOV3+: free<br>HOV2: tolled<br>SOV: prohibited<br>Off peak (HOV2+ lane)<br>HOV2+: free<br>SOV: prohibited | A flat rate: $2 per use. |
| SR-91 Orange County CA | 10-mile 4 lanes. Midway inaccessible. 8 general lanes. | Peak<br>HOV3+: 50% toll off<br>HOV2, SOV: fully tolled<br>Off peak<br>HOV3+: free<br>HOV2, SOV: fully tolled | Time-of-day tolls: $1.25-9.55 per use. |
| I-25 Denver CO | 6.6-mile reversible 2 lanes. Midway inaccessible. 8 general lanes. | HOV2+: free<br>SOV: tolled | Time-of-day tolls: $0.5-3.5 per use. |
| I-10W Houston TX | 12-mile 4 lanes. Midway accessible. 10 general lanes. | Bus: free<br>HOV2+: peak free; off peak tolled<br>SOV: tolled | Time-of-day tolls by distance: $0.3-1.6 per toll section. |
| I-15S San Diego CA | 8-mile reversible 2 lanes. Midway inaccessible. 10 general lanes. | HOV2+: free<br>SOV: tolled | Dynamic tolls: $0.5-8 per use. |
| I-95 Miami FL | 7.75-mile 4 lanes. Midway inaccessible. 8 general lanes. | HOV3+: free<br>HOV2, SOV:tolled | Dynamic tolls: $0.25-7.25 per use. |
| SR-167 Seattle WA | 9-mile 2 lanes. Midway accessible. 4 general lanes. | HOV2+: free<br>SOV: tolled | Dynamic tolls:$0.5-9 per use. |
| I-15N San Diego CA | 12-mile reversible 4 lanes. Midway accessible. 8 general lanes | HOV2+: free<br>SOV: tolled | Dynamic tolls by distance: $0.5-8. |
| I-394 Minneapolis MN | 7-mile 2 lanes and 3.3-mile reversible 2 lanes. 2 tolling sections. Midway accessible. 4 general lanes. | HOV2+: free<br>SOV: tolled | Dynamic tolls by distance: $0.25-8. |

The other two toll collection schemes, per use-based and distance-based, are more adaptive than pass-based toll collection scheme. Per use-based scheme charges the same price even though drivers may travel different distances on the toll lane. Distance-based toll scheme charges the drivers a toll based on the distance travelled, and is more reasonable than per use-based scheme. However it is also more complicated than per use-based scheme and finding the appropriate pricing strategy is difficult. Both per use-based and distance-based toll collection schemes can be applied through a flat rate, time-of-day rate and dynamic rate patterns.

In this chapter, we formulate and solve a very general optimal distance-based dynamic pricing model based on the stochastic macroscopic traffic flow model developed in Chapter II. An advantage of using macroscopic model (over a microscopic simulation model) is that it can be calibrated on-line by applying the traffic state estimation algorithm developed in Chapter III. Following the formulation, we develop a methodology to obtain an optimal pricing strategy to maximize the total expected revenue. Although the objective discussed in this study is to maximize the revenue, the general model formulation and solution can be easily adaptable to any other objective, like the maximizing of the total throughput, etc. We then propose a (simulation-based) numerical algorithm to obtain the optimal prices efficiently in real time. An important fact about this pricing scheme is that it is applicable to the toll lane with any number of toll entrances and exits. In addition, the general pricing model developed in this chapter is not limited to specific traffic flow model. It is readily adapted to other macroscopic traffic models, such as the classical LWR model, and other traffic flow models. The objective of this chapter is to present the formulation and solution of a general dynamic toll pricing scheme with an underlying traffic flow model (developed in Chapter II).

The outline of this chapter is described as follows: Section 4.2 describes the model

formulation in detail. In section 4.3, we analyze the mathematical optimization model and present the method to solve this problem. A numerical example is solved in Section 4.4 to validate the performance of proposed dynamic pricing strategy; and Section 4.5 concludes this chapter.

## 4.2 Distance-Based Dynamic Pricing Model

### 4.2.1 Infrastructure

In this model, there are two different types of lanes running in parallel. One is a general lane where drivers can enter for free. The other is a managed toll lane where drivers pay a toll to enter. During peak hours, the general lane is normally congested and the toll lane is maintained at free flow speed by adjusting the toll prices. This is very useful for drivers who can pay a toll to avoid congestion as well as in emergency situations.

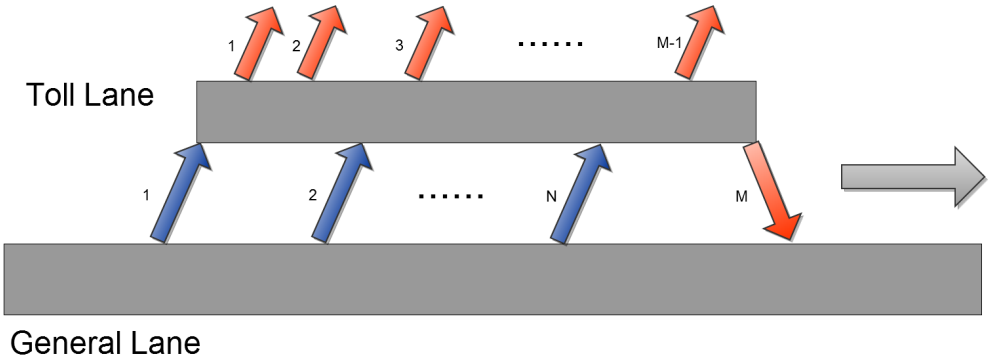



Figure 4.1: General lane and managed toll lane

Figure 4.1 illustrates the structure of the managed toll lane and the general lane in this model. Suppose the managed toll lane starts at mile post 0 and ends at mile post $L$. There are $N$ toll entrances located at $l_1$, $l_2$,...,$l_N$. Normally the first toll entrance is located at the start of the managed toll lane, therefore $l_1 = 0$. There are also $M$ toll exits located at mile posts $\tilde{l}_1$, $\tilde{l}_2$,...,$\tilde{l}_M$. Normally the last toll exit is

located at the end of the toll lane, therefore $\tilde{l}_M = L$. Vehicles can only enter the toll lane through the $N$ toll entrances and leave the toll lane through the $M$ toll exits.

For distance-based pricing strategy, the vehicles are charged according to the distance they travel on the toll lane. However since the distance information is only available when the vehicles leave the toll lane, the toll should be collected at the toll exits. In order to record the travel information of vehicles on the toll lane, all the vehicles must be equipped with an electronic device which can communicate with the sensors installed at toll entrances and exits. The driver of a vehicle approaching a toll entrance can see the pricing information and the corresponding estimated travel time on both lanes to every subsequent toll exit. Based on the information, the driver then decides whether to switch to the toll lane by evaluating the trade-off between the price and time saving. If the driver decides to enter the toll lane, when the vehicle crosses the toll entrance, the sensor at the toll entrance will record the time, location and toll price table inside this vehicle's device. When the vehicle leaves the toll lane, the sensor at the exit will read this information from the vehicle's device and charge the appropriate toll to the driver.

### 4.2.2 The Underlying Traffic Flow Model on the General Lane

In this section, we will describe the underlying traffic flow model on the general lane. When making decisions, the drivers compare the travel time from the toll entrance to the toll exit on both lanes. Since the traffic on the toll lane is maintained under free flow condition with a hard constraint, the travel time from one location to another on the toll lane is constant. Let $v_f$ denote the free flow speed on the toll lane and $\tau_n^m$ denote the travel time from toll entrance $n$ to toll exit $m$ on the toll lane. The travel time is given in equation (4.1).

$$\tau_n^m(t) \equiv \frac{\tilde{l}_m - l_n}{v_f} \quad n \in [1, N] \quad m \in [1, M] \quad \tilde{l}_m > l_n \tag{4.1}$$

Unlike the situation on the toll lane, the general lane is normally congested especially in peak hours. Thus the travel time on the general lane is not constant. The prediction of the travel time on the general lane is needed as it is displayed at the toll entrances. Here we make the following assumption about the traffic flow model on the general lane, presented in Assumption IV.1.

**Assumption IV.1.** *The traffic flow switching to the toll lane has no impact on the traffic density evolution on the general lane, and the density on the general lane is a stochastic process following the stochastic partial differential equation model in equation* (4.2) *shown in Chapter II.*

$$\begin{aligned} \hat{\rho}_t + \hat{q}_x &= g(\hat{\rho}, x, t) \\ \hat{q} &= \hat{\rho} \cdot V_*(\hat{\rho}) \\ g(\hat{\rho}, x, t) \cdot dx \cdot dt &= (a(x,t) + b(x,t) \cdot \hat{\rho}) \cdot dx \cdot dt + \sigma(x,t) \cdot dW(x,t) \end{aligned} \tag{4.2}$$

*where*

- $(x, t)$ *is the space and time pair,* $x \in [0, L]$ *and* $t \in [0, T]$*;*
- $\hat{q}(x, t)$ *is the volume (i.e, number of vehicles passing through per unit time) at location* $x$ *and time* $t$ *on the general lane;*
- $\hat{\rho}(x, t)$ *is the density (i.e, number of vehicles per unit distance) at location* $x$ *and time* $t$ *on the general lane;*
- $V_*(\cdot)$ *is the speed-density function.*
- $W(x, t)$ *is the Brownian Sheet, a Gaussian process indexed by two parameters.*

The stochastic traffic flow model in (4.2) has been described in Chapter II, therefore we omit the justification of this model here. There are two reasons why we assume that the traffic flow evolution on the general lane is not affected by the flow

switching to the toll lane. First, under this assumption, the travel time between any two locations on the general lane is not affected by the toll prices. Therefore in the stochastic control model of the later sections for solving the optimal pricing problem, a subset of the state variables becomes independent of the decision variables. That makes it possible to solve the complicated multidimensional stochastic control problem explicitly: It can be decomposed into several easier and smaller problems, as explained by Theorem IV.4. Without this assumption, the optimal pricing problem is unsolvable (due to the curse of dimensionality) when the number of toll stations is large. The solution of optimal pricing without this assumption is under study. Another reason for this assumption is that the volume capacity in the general lane is normally much larger than that of the toll lane since the number of general lanes is usually larger than the number of toll lanes. Thus the traffic switching to the toll lane is a small proportion of the traffic flow in the general lane. The impact of the switching traffic flow on the general lane density evolution is thus small, and can be neglected.

Let $\hat{\tau}_n^m(t)$ denote the travel time from toll entrance $n$ to toll exit $m$ on the general lane at time $t$, and $\hat{\rho}(\cdot, t)$ denote the density vector of the general lane at time $t$. Then under the traffic flow model in (4.2), $\hat{\tau}_n^m(t)$ has the following property as shown in Theorem IV.2.

**Theorem IV.2.** *Under Assumption IV.1, the travel time between any two locations on the general lane at time $t$ is a function of the traffic density vector on the general lane at that time. The relationship is given in equation* (4.3)*, where $f_n^m(\cdot)$ does not have a closed-form solution and can only be derived by simulation.*

$$\hat{\tau}_n^m(t) = f_n^m(\hat{\rho}(\cdot, t)) \quad n \in [1, N] \quad m \in [1, M] \quad \tilde{l}_m > l_n \tag{4.3}$$

*Proof of Theorem IV.2.* Due to the shock wave caused by the nonlinearity of the flux

function, there is no closed-form solution for the stochastic model in (4.2). We apply a numeric scheme, Godunov's scheme, to solve the stochastic model. The detailed steps of Godunov's scheme are explained in Algorithm 2. By using Godunov's scheme, we can simulate a number of scenarios for the density evolution in the future given the density vector $\hat{\rho}(\cdot, t)$ at time $t$. Under each scenario, the travel time can be calculated via the speed-density function. The steps of obtaining simulated scenarios of density evolution are described in Section 2.3.3 in detail. Algorithm 9 describes the steps to obtain $\hat{\tau}_n^m(t)$, where $\hat{\rho}^k(x,t)$ represents the simulated density evolution under scenario $k$. □

**Algorithm 9** Travel time prediction.

**Step 1** Suppose there is a vehicle driving from location $l_n$ to $\hat{l}_m$ along the general lane. At time $t$, this vehicle is at location $l_n$. Let $T(x)$ denote the time when the vehicle arrives at location $x$ on the general lane. Obviously:

$$T(l_n) = t$$

**Step 2** When the vehicle arrives at location $x$ at time $T(x)$, the density at the location is $\hat{\rho}^k(x, T(x))$. According to the speed-density function, the speed of the vehicle should be equal to $V_*(\hat{\rho}^k(x, T(x)))$, therefore $T(x)$ should satisfy the following differential equation:

$$dT(x) = \frac{dx}{V_*(\hat{\rho}^k(x, T(x)))}$$

**Step 3** Solve the above differential equation numerically by discretizing the time and space to obtain $T^k(\hat{l}_m)$, which is the time when this vehicle arrives at location $\hat{l}_m$ under scenario $k$. Then the travel time from $l_n$ to $\hat{l}_m$ under this scenario is:

$$\hat{\tau}^k = T^k(\hat{l}_m) - t$$

**Step 4** Take the average over all scenarios as the prediction of the travel time.

$$\hat{\tau}_n^m(t) = \frac{1}{K} \sum_{k=1}^{K} \hat{\tau}^k$$

---

Based on the explanation above, function $f_n^m(\cdot)$ is very complicated and can only

be obtained by simulation. However the complication of the travel time function will not affect the solution of the pricing model. Since the travel time will be independent of the decision variables, the predicted travel time can be calculated in advance before they are used in the stochastic control model. An important characteristic of this optimal pricing model is that it is not limited to any specific underlying traffic flow model: For any traffic flow model, as long as the travel time can be expressed as an function of the current density vector, the pricing model and our methodology will apply.

### 4.2.3 Demand Function

For every toll entrance, the traffic demand can be classified by the destination. For example, for toll entrance $n$, Let $Z_n$ denote the index of the first exit following toll entrance $n$. Let $\Phi_n$ denote the set of indexes of all toll exits behind toll entrance $n$, so $\Phi_n = \{Z_n, Z_n+1, \cdots, M\}$. Let $D_n(t)$ denote the traffic volume entering the toll lane through toll entrance $n$ at time $t$, it consists of $D_n^m(t)$ $(m \in \Phi_n)$, where $D_n^m(t)$ represents the traffic volume entering at toll entrance $n$ at time $t$ and and destined to leave at toll exit $m$. Equation (4.4) describes this relationship.

$$D_n(t) \quad = \quad \sum_{m \in \Phi_n} D_n^m(t) \tag{4.4}$$

At time $t$, the toll entrance $n$ will publish the prices for every downstream exit as well as the corresponding estimated travel time on both general lane and the toll lane. The drivers make decisions whether to stay on the general lane or switch to the toll lane based on this information. The drivers who plan to leave at toll exit $m$, observe $\tau_n^m(t)$, $\hat{\tau}_n^m(t)$ and $p_n^m(t)$, where $p_n^m(t)$ is the toll for entering at entrance $n$ at time $t$ and leaving at exit $m$, and decide to/not to enter. Therefore the demand $D_n^m(t)$ is a function of $\tau_n^m(t)$, $\hat{\tau}_n^m(t)$ and $p_n^m(t)$, and let $d_n^m$ denote the function, which

is given in equation (4.5). The demand function should be decreasing in $p_n^m(t)$, $\tau_n^m(t)$ and increasing in $\hat{\tau}_n^m(t)$, since a higher price and longer travel time on the toll lane will cause less people to enter the toll lane. Similarly, more drivers will opt for the toll lane with a longer travel time on the general lane.

$$D_n^m(t) = d_n^m(p_n^m(t), \hat{\tau}_n^m(t), \tau_n^m(t)) \tag{4.5}$$

Multinomial logit model is well established in marketing research literature to model the demand when multiple alternatives are available. In recent years, it has been widely used in traffic demand analysis. For examle, *Lou et al.* (2011) applied multinomial logit model to approximate the demand function for solving the optimal pricing problem. The basic idea of multinomial logit model in traffic demand analysis is described as below. When $I$ drivers are making decisions choosing one route from $J$ alternatives. The individual utility $U_{ij}^\star$ of individual $i$ choosing route $j$ is:

$$U_{ij}^\star = U_j + \epsilon_{ij} \tag{4.6}$$

where $U_j$ is the expected utility of route $j$ and $\epsilon_{ij}$ is the individual noise. Multinomial logit model assumes that $\epsilon_{ij}$ are independently and identically distributed with an extreme value distribution. The cumulative distribution function of $\epsilon_{ij}$ is:

$$P(\epsilon_{ij} < x) = e^{-e^{-x}} \tag{4.7}$$

Each individual will choose the route which has the highest utility for himself/herself. Given the distribution of $\epsilon_{ij}$ in equation (4.7), we can derive that for any individual $i$, the probability this individual chooses route $j$ will be:

$$P_{ij} \quad = \quad P(U_{ij}^\star \geq U_{ik}^\star \quad \forall k = 1, 2, ..., J)$$

$$= \frac{\exp(U_j)}{\sum_{k=1}^{J} \exp(U_k)} \tag{4.8}$$

In our model, there are two alternatives: general lane and managed toll lane. *Lou et al.* (2011) assumes that the utility function for either lane is a linear function of its toll price and travel time as shown in equation (4.9), where $\tau$ and $\hat{\tau}$ are travel time on the toll lane and general lane from the driver's origin to destination; $p$ is the toll price for the driver's traveling distance. $\alpha$, $\eta$ and $\gamma^h$, $\gamma^g$ are parameters. $\frac{\alpha}{\eta}$ represents the driver's trade-off between the time saved and money spent. $\gamma^h$ and $\gamma^g$ represents the driver's fixed preference between the two lanes. When $\tau = \hat{\tau}$ and $p = 0$, drivers may have different preference because of the road condition and lane capacity. The difference of $\gamma^h$ and $\gamma^g$ represents this preference. It should be noted that $\alpha$, $\eta > 0$ because the utility function is decreasing in the travel time and toll price.

$$\begin{aligned} U^h &= -\alpha\tau - \eta p + \gamma^h \\ U^g &= -\alpha\hat{\tau} + \gamma^g \end{aligned} \tag{4.9}$$

Given the utility functions in equation (4.9), we can obtain the demand function in equation (4.10). $D$ is the realized traffic demand for the toll lane, $A$ is the potential total demand. Thus the traffic volume entering the toll lane is a function of the time saving $\hat{\tau} - \tau$ and toll price $p$.

$$\begin{aligned} D &= A\frac{\exp(U^h)}{\exp(U^h) + \exp(U^g)} \\ &= A\frac{1}{1 + \exp(\alpha(\tau - \hat{\tau}) + \eta p + \gamma)} \\ \gamma &= \gamma^g - \gamma^h \end{aligned} \tag{4.10}$$

The traffic demand for each origin-destination pair is assumed to follow the multinomial logit model in equation (4.10). For each pair of toll entrance $n$ and toll exit

$m$, there is a set of parameters associated with it: $A_n^m$, $\alpha_n^m$, $\eta_n^m$ and $\gamma_n^m$. Out of the four parameters, only $A_n^m$ is dynamically changing over time, while the other three are constant, because $\alpha_n^m$, $\eta_n^m$ and $\gamma_n^m$ are only affected by the value of time of the population living in this area, it will only change slightly over a long time. However $A_n^m$ is affected by the travel pattern, it is changing quickly during the day. Therefore generally the demand function of $D_n^m$ at time $t$ is given in equation (4.11). *Lou et al.* (2011) applied reactive self-learning algorithm to estimate these parameters.

$$D_n^m(t) = A_n^m(t)\frac{1}{1+\exp(\alpha_n^m(\tau_n^m(t)-\hat{\tau}_n^m(t))+\eta_n^m p_n^m(t)+\gamma_n^m)} \tag{4.11}$$

### 4.2.4 Flow Conservation on the Toll Lane

Unlike the density evolution on the general lane, which is an uncontrolled Markov process described in equation (4.2), the flow on the toll lane is a controlled Markov process, because the traffic volume entering the toll lane is controlled by the decision variables, the toll prices. Let $q^m(x,t)$ denote the traffic flow, whose destination is toll exit $m$, on the toll lane at location $x$ and time $t$. Then the evolution of $q^m(x,t)$ should follow Theorem IV.3.

**Theorem IV.3.** *$q^m(x,t)$ follows the following transition equation for $\forall m \in [1,M]$, where $\delta(\cdot)$ is Dirac delta function.*

$$\begin{aligned} \frac{1}{v_f}\frac{\partial}{\partial t}q^m(x,t)+\frac{\partial}{\partial x}q^m(x,t) &= \sum_{n:m\in\Phi_n} D_n^m(t)\delta(x-l_n) \\ q^m(0,t) &= 0 \qquad \forall t \\ q^m(x,t) &= 0 \qquad \forall t, \forall x > \tilde{l}_m \end{aligned} \tag{4.12}$$

$$\delta(x) = \begin{cases} +\infty & x = 0 \\ 0 & x \neq 0 \end{cases} \tag{4.13}$$

$$\int_{-\infty}^{+\infty} \delta(x)dx = 1 \tag{4.14}$$

*Proof of Theorem IV.3.* Since the traffic on the toll lane is under free flow speed $v_f$, then at any $x$ and $t$, we have:

$$q^m(x + v_f dt, t + dt) - q^m(x, t) = \sum_{n:m\in\Phi_n} D_n^m(t)\mathbf{1}_{\{x=l_n\}} \tag{4.15}$$

$\mathbf{1}_{\{x=l_n\}}$ is an indicator function, it is equal to 1 if $x = l_n$ and 0 otherwise. $D_n^m(t)$ is the entering toll traffic volume through toll entrance $n$ with destination at toll exit $m$. Therefore, along the characteristic line $x = x_0 + v_f t$, $q^m$ only have jumps at those toll entrances and is constant elsewhere. That is why the right hand side of the equation (4.15) involves the indicator function.

Using Taylor's expansion on the left hand side of equation (4.15), gives equation (4.16). Along the characteristic line $x = x_0 + v_f t$, $dx = v_f dt$, the right hand side of equation (4.16) can be further simplified, and the transition equation for $q^m(x, t)$ in equation (4.12) is proved. The boundary condition $q^m(0, t) = 0$ is based on the assumption that the first toll entrance is located at the beginning of the toll lane. $q^m(x, t) = 0(\forall x > \tilde{l}_m)$ because all the vehicles with destination at toll exit $m$ will leave the toll lane at $\tilde{l}_m$.

$$\frac{\partial}{\partial t}q^m(x, t) + v_f \frac{\partial}{\partial x}q^m(x, t) \quad = \quad \sum_{n:m\in\Phi_n} D_n^m(t)\frac{\mathbf{1}_{\{x=l_n\}}}{dt}$$

$$= \sum_{n:m\in\Phi_n} D_n^m(t) v_f \frac{\mathbf{1}_{\{x=l_n\}}}{dx} \tag{4.16}$$
$$= \sum_{n:m\in\Phi_n} D_n^m(t) v_f \delta(x - l_n)$$

□

### 4.2.5 The Constraint

Maintaining free flow speed on the toll lane is a hard constraint, and the toll prices will be adjusted dynamically to make sure this constraint is not violated. However in practice, this constraint might be violated because of sudden spike in demand. Once this situation happens, the pricing strategy will be switched to another mode under which all the toll prices will be set to the maximum price or even the toll entrances will be temporally closed until the managed toll lane recovers to free flow speed condition again. The maximum price is usually predetermined by an agreement between the public and highway administrators, which serves as an upper threshold for the toll prices.

In this study, we mainly focus on the pricing strategy when this constraint is not violated due to sudden spike in demand. So we consider keeping the toll lane congestion free as a hard constraint. In other words, the traffic on the toll lane must travel at the free flow speed. In order to satisfy this constraint, the traffic flow entering the toll lane must not impact the existing traffic on the toll lane. Every lane has a traffic flow capacity. We assume that if the flow does not exceed this capacity then free flow speed is attained. Obviously $q^m(x,t)$ is not continuous at the locations of toll entrances because of the jump of its value caused by the entering traffic volume. We use $q^m(l_n^-,t)$ to denote the traffic volume right before toll entrance $n$ and $q^m(l_n^+,t)$ to denote the traffic volume right after toll entrance $n$. $q^m(l_n^+,t)$ should satisfy the

constraint in equation (4.17)

$$\sum_{m\in\Phi_n} q^m(l_n^+,t) \le C_n \quad \forall n \in [1,N] \tag{4.17}$$

### 4.2.6 Complete Model Formulation

By summarizing the above sections, we can see that optimal pricing problem can actually be formulated as a stochastic control model. The state variable of the system at time $t$ is defined by the density vector on the general lane $\hat{\rho}(x,t)$ and the flow vector on the toll lane $q^m(x,t)(m \in [1,M])$. The control variable is defined by $p_n^m(t)$ $(n \in [1,N]), m \in \Phi_n)$. Equation (4.2) and equation (4.12) are the transition equations of the state variables in the stochastic control model. From the transition equation, we can see that $q^m(x,t)$ is controlled by the control variable $p_n^m(t)$ and $\hat{\rho}(x,t)$ is uncontrolled. Besides the transition equation, there is a constraint on the state variable $q^m(x,t)$ as shown in equation (4.17). The objective of the stochastic control problem is to maximize the expected total revenue from current time $t$ to a specified time horizon $T$. The complete stochastic control model of optimal pricing problem is given in equation (4.18).

$$\begin{aligned}
&\max E[(\int_t^T \sum_{n=1}^{N} \sum_{m\in\Phi_n} D_n^m(s)p_n^m(s)ds)] \\
&D_n^m(t) = A_n^m(t)\frac{1}{1+\exp(\alpha_n^m(\tau_n^m(t)-\hat{\tau}_n^m(t))+\eta_n^m p_n^m(t)+\gamma_n^m)} \quad \forall n \in [1,N], \forall m \in \Phi_n \\
&\tau_n^m(t) = \frac{\tilde{l}_m - l_n}{v_f} \quad n \in [1,N], \forall m \in \Phi_n \\
&\hat{\tau}_n^m(t) = f_n^m(\hat{\rho}(\cdot,t)) \quad n \in [1,N], \forall m \in \Phi_n \\
&\frac{\partial}{\partial t}\hat{\rho}(x,t) + \frac{\partial}{\partial x}[\hat{\rho}(x,t)\cdot v(\hat{\rho}(x,t))] = g(\hat{\rho}(x,t),x,t) \\
&\frac{1}{v_f}\frac{\partial}{\partial t}q^m(x,t) + \frac{\partial}{\partial x}q^m(x,t) = \sum_{n:m\in\Phi_n} D_n^m(t)\delta(x-l_n) \quad \forall m \in [1,M] \\
&\sum_{m\in\Phi_n} q^m(l_n^+,t) \le C_n \quad \forall n \in [1,N]
\end{aligned} \tag{4.18}$$

In the objective function in equation (4.18), $D_n^m(s)p_n^m(s)$ is the revenue rate collected from the vehicles entering the toll lane at entrance $n$ and time $s$ and leaving the toll lane at exit $m$. The summation over all the pairs of entrance-exit is the total revenue. From the model in (4.18), the stochastic control problem is very complicated: One state variable $\hat{\rho}(x,t)$ is driven by a stochastic partial differential equation, while another state variable $q^m(x,t)$ is controlled by the intermediate variable $D_n^m(t)$, which is a function of the decision variable $p_n^m(t)$ and state variable $\hat{\rho}(x,t)$. In addition, the decision variable $p_n^m(t)$ will be constrained so that the state variable $q^m(x,t)$ could satisfy the constraint in equation (4.17). The following sections will continue to discuss the solution of this stochastic control model.

## 4.3 Analysis of the Stochastic Control Model

### 4.3.1 Complexity Analysis

There are different kinds of approaches to solve a stochastic control problem. An classical approach is Bellman equation. It writes the optimal value at a certain point in time in terms of the immediate reward incurred by the optimal decision plus the optimal value of the remaining control problem that results from the optimal decision. In the stochastic control model in equation (4.18), let $J(t, q^1(\cdot,t), \cdots, q^M(\cdot,t), \hat{\rho}(\cdot,t))$ represent the maximum expected total revenue from time $t$ to $T$ given that the state of the system at $t$ is $q^1(\cdot,t)$, $\cdots$, $q^M(\cdot,t)$ and $\hat{\rho}(\cdot,t)$. $\mathbb{P}$ represents the space of all possible prices satisfying the constraint in equation (4.17). The Bellman equation, which is also called Hamilton-Jacobi-Bellman(HJB) equation in continuous-time stochastic control problem, is given in equation (4.19).

$$\begin{aligned} J(t, q^1(\cdot,t), \cdots, q^M(\cdot,t), \hat{\rho}(\cdot,t)) = \sup_{p_n^m(t)\in\mathbb{P}} \{ & R(p_n^m(t), q^1(\cdot,t), \cdots, q^M(\cdot,t), \hat{\rho}(\cdot,t))dt \\ & + E[J(t+dt, q^1(\cdot,t+dt), \cdots, q^M(\cdot,t+dt), \hat{\rho}(\cdot,t+dt))]\} \end{aligned}$$

$$J(T, q^1(\cdot, T), \cdots, q^M(\cdot, T), \hat{\rho}(\cdot, T)) \equiv 0 \tag{4.19}$$

The reward function in the HJB equation in (4.19) is defined in equation (4.20). It is the total revenue rate at time $t$.

$$\begin{aligned} R(p_n^m(t), q^1(\cdot, t), \cdots, q^M(\cdot, t), \hat{\rho}(\cdot, t)) &= \sum_{n=1}^{N} \sum_{m \in \Phi_n} p_n^m(t) D_n^m(t) \\ &= \sum_{n=1}^{N} \sum_{m \in \Phi_n} \frac{p_n^m(t) A_n^m(t)}{1 + \exp(\alpha_n^m(\tau_n^m(t) - \hat{\tau}_n^m(t)) + \eta_n^m p_n^m(t) + \gamma_n^m)} \end{aligned} \tag{4.20}$$

Theoretically, the optimal value function of the stochastic control problem in equation (4.19) can be obtained analytically by solving a partial differential equation. However in practice, it is usually extremely difficult to get an analytic solution for the HJB equation, especially when the state variables $\hat{\rho}(\cdot, t)$ and $q^m(\cdot, t)$ $m \in [1, M]$ are infinite dimensional. A close form solution for HJB equation only exists in a few special cases.

Often numerical solution is also impossible due to the curse of dimensionality. In order to solve the stochastic control problem in this model, both the time and state variables are discretized. After discretization, the model becomes a discrete-time Markov decision process with time intervals $\triangle t$. The highway is also discretized into small cells with length $\triangle x$, and the states of the system, $\hat{\rho}(\cdot, t)$ and $q^m(\cdot, t)$ $m \in [1, M]$, are represented by vectors instead of functions. When discretizing time and space, in order to guarantee that the waves do not interact with each other, $\triangle t$ and $\triangle x$ should satisfy Courant-Friedrichs-Lewy(CFL) condition(**?**) given in equation (4.21).

$$v_f \triangle t <= \triangle x \tag{4.21}$$

If $\triangle x$ is too large, the approximate solution obtained by the numerical algorithm is not accurate. If $\triangle x$ is too small, then $\triangle t$ has to be small too, which will increase

the computational cost. In most literature it is suggested that an acceptable value of $\triangle x$ is about 0.5 miles, and we use this in our numerical algorithm as well.

There are several difficulties encountered when solving the stochastic control problem in (4.19). First, the dimension of the state space is very high, making it impossible to solve due to the curse of dimensionality. For example, if the toll lane is 10 miles long and each cell is 0.5 miles, there would be 20 cells, which means $\hat{\rho}(\cdot, t)$, $q^1(\cdot, t)$,$\cdots$, $q^M(\cdot, t)$ all would be a vector of dimension of 20. In total, the dimension of the state variable is $20(M + 1)$. This high dimension leads to a huge computational cost and makes the problem almost unsolvable. Another difficulty is related to the dimensionality of the decision variable. The decision variable $p_n^m(t)$ is the price vector covering all the pairs of the toll entrance and its downstream toll exits. When the number of toll entrances and exits is large, the dimension of $p_n^m$ may also blow up the computational time for the solution. The high dimensionality of both the state space and control variables makes the stochastic control problem almost impossible to solve by classical approaches.

### 4.3.2 Solution

The analysis in previous section shows that it is almost impossible to solve the stochastic control problem by using traditional numerical approach. In this section, we will present a feasible solution to the stochastic control problem. At first, we will decompose the high dimensional stochastic control problem into several separate low dimensional stochastic optimization problems. Then the sub stochastic optimization problem is transformed to a deterministic convex optimization problem by using simulated scenarios to approximate the stochastic process. Since the deterministic convex optimization problem can be solved quickly and efficiently by convex optimization software packages, this strategy gives a good numerical scheme for the original stochastic control problem.

#### 4.3.2.1 Decomposition

An important property of the transition equation in the stochastic control model (4.18) is that the state variable $\hat{\rho}(x,t)$ is independent of the control variable $p_n^m(t)$. This property results in Theorem IV.4. Based on Theorem IV.4, the stochastic control problem can be decomposed into small sub problems.

**Theorem IV.4.** *At time $t$, the price vector of toll entrance $n$: $p_n^m(t)$, $(\forall m \in \Phi_n)$ only affects the price vector of downstream toll entrance $n+1$ at time $t+\kappa_n$: $p_{n+1}^{\bar{m}}(t+\kappa_n)$, $(\forall \bar{m} \in \Phi_{n+1})$, where $\kappa_n = \frac{l_{n+1}-l_n}{v_f}$. In other words, the decisions at any toll entrance only affects the decisions of downstream toll entrances along the characteristic line, which has slope $\frac{1}{v_f}$ as shown in figure 4.2. It does not have any impact on the decisions that are not on the characteristic line.*

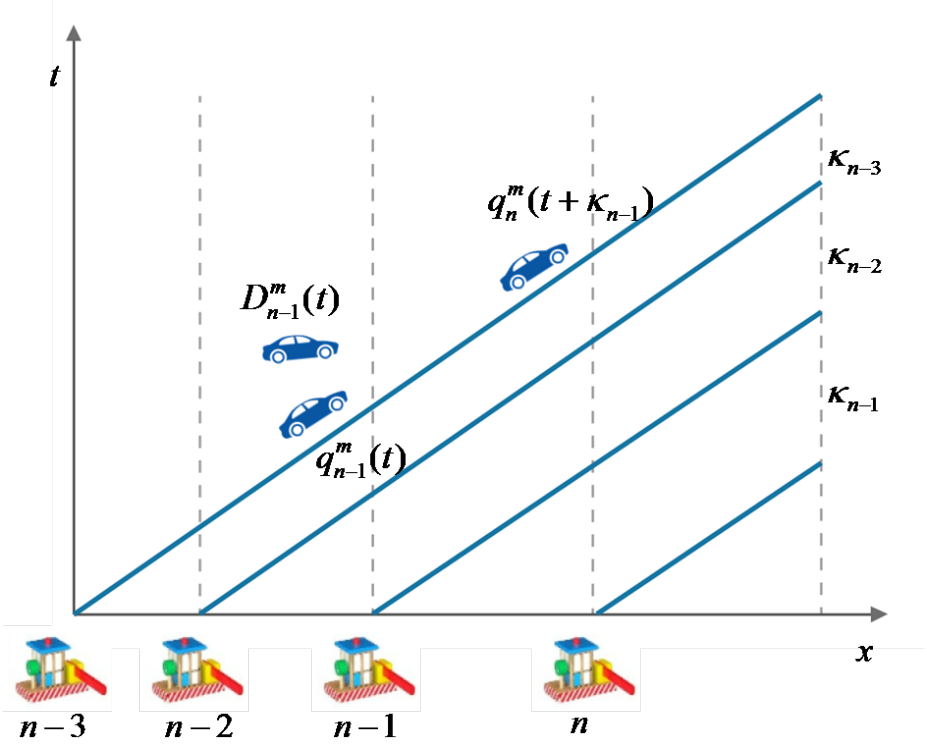



Figure 4.2: Decomposition of dynamic toll pricing problem

*Proof of Theorem IV.4.* First, let us prove that $p_n^m(t)$, $(\forall m \in \Phi_n)$ affects the price

vector of downstream toll entrance $n+1$ at time $t+\kappa_n$, which is $p_{n+1}^{\bar{m}}(t+\kappa_n)\ \forall \bar{m} \in \Phi_{n+1}$. At time $t$ in toll entrance $n$, $p_n^m(t)$ will determine $D_n^m(t)$ and it must satisfy the constraint so that the total traffic flow right after toll entrance $n$ does not exceed the lane capacity. According to the flow conservation in (4.12), the traffic will travel along the line with slope $\frac{1}{v_f}$ on the toll lane as shown in figure 4.2. The entering traffic flow $D_n^m(t)$ will reach toll station $n+1$ at time $t+\kappa_n$. Because of the capacity constraint, it will set another constraint for the price vector of toll entrance $n+1$ at time $t+\kappa_n$. For example, at time $t+\kappa_n$, the upcoming traffic flow at toll station $n+1$: $q^{\bar{m}}(l_{n+1}^-, t+\kappa_n) = q^{\bar{m}}(l_n^-, t) + D_n^{\bar{m}}(t)$. Since $q^{\bar{m}}(l_{n+1}^-, t+\kappa_n)$ would set a constraint for the decision $p_{n+1}^{\bar{m}}(t+\kappa_n)$ as illustrated in equation (4.17), obviously $p_{n+1}^{\bar{m}}(t+\kappa_n)$ will be affected by $p_n^m(t)$. For the same reason, $p_{n+2}(t+\kappa_n+\kappa_{n+1})$ is affected by $p_{n+1}(t+\kappa_n)$ and so on. Therefore, $p_n(t)$ will affect the decisions of all its downstream toll entrances on the characteristic line.

Secondly, $p_n^m(t)$ does not have any impact on the decisions of any toll entrance outside the characteristic line. For example, for toll station $\acute{n}$ at time $s$, suppose $(l_{\acute{n}}, s)$ is not on the same characteristic line with $(l_n, t)$, in other words, the line connecting $(l_{\acute{n}}, s)$ and $(l_n, t)$ does not have slope $\frac{1}{v_f}$. $p_{\acute{n}}^{\bar{m}}(s)$ is determined by $q^{\bar{m}}(l_{\acute{n}}^-, s)\ m \in \Phi_{\acute{n}}$ and $\hat{\rho}(\cdot, s)$. Because $q^{\bar{m}}(l_{\acute{n}}^-, s)\ m \in \Phi_{\acute{n}}$ is only affected by the upstream decisions on the characteristic line of $(l_{\acute{n}}, s)$, and $\hat{\rho}(\cdot, s)$ is independent of all the decision variables. So $p_{\acute{n}}^{\bar{m}}(s)$ is not affected by $p_n^m(t)$ if $(l_{\acute{n}}, s)$ and $(l_n, t)$ are not on the same characteristic line. Therefore the proof is complete. □

Based on Theorem IV.4, the problem could be decomposed along the characteristic line shown in Figure 4.2. Since the decision variables in one characteristic line are independent of those on any others, $p_n^m(t)$ only affects the total revenue rate collected on its characteristic line and it does not have any impact on other lines. Therefore at time $t$, the price decisions of every toll entrance are independent of each other, and problem (4.18) can be decomposed into $N$ subproblems, where each subproblem

obtains the optimal prices for one toll entrance to maximize the expected total revenue rate along the characteristic line passing this toll station at time $t$. In other words, $p_n^m(t)$ is determined to maximize the expected total revenue rate of all its downstream toll entrances along the characteristic line.

For toll entrance $\underline{n}$ at time $t$, the stochastic control problem for this toll entrance is formulated as equation (4.22).

$$\begin{aligned}
\max \quad & E[\sum_{n=\underline{n}}^{\overline{n}}(\sum_{m\in\Phi_n} D_n^m p_n^m)] \\
& q_n^m = q_{n-1}^m + D_{n-1}^m \quad \forall n \in [\underline{n}+1, \overline{n}], m \in \Phi_n \\
& D_n^m = A_n^m \frac{1}{1+\exp(\alpha_n^m(\tau_n^m - \hat{\tau}_n^m) + \eta_n^m p_n^m + \gamma_n^m)} \\
& \sum_{m\in\Phi_n} D_n^m + \sum_{m\in\Phi_n} q_n^m \le C_n \qquad \forall n \in [\underline{n}+1, \overline{n}] \\
& \tau_n^m = \frac{\tilde{l}_m - l_n}{v_f} \quad n \in [1, N] \quad m \in \Phi_n \\
& \hat{\tau}_n^m = f_n^m(\hat{\rho}(\cdot, t + \sum_{i=\underline{n}}^{n-1} \kappa_i)) \\
& \kappa_n = \frac{l_{n+1} - l_n}{v_f} \quad n \in [1, N]
\end{aligned} \tag{4.22}$$

In equation (4.22), $D_n^m$, $p_n^m$, $q_n^m$, $\tau_n^m$, $\hat{\tau}_n^m$ and $A_n^m$ are actually simplified notations as illustrated below:

$$\begin{aligned}
D_n^m &= D_n^m(t + \sum_{i=\underline{n}}^{n-1} \kappa_i) \\
p_n^m &= p_n^m(t + \sum_{i=\underline{n}}^{n-1} \kappa_i) \\
q_n^m &= q^m(l_n^-, t + \sum_{i=\underline{n}}^{n-1} \kappa_i) \\
\tau_n^m &= \tau_n^m(t + \sum_{i=\underline{n}}^{n-1} \kappa_i)
\end{aligned} \tag{4.23}$$

$$\hat{\tau}_n^m = \hat{\tau}_n^m(t + \sum_{i=\underline{n}}^{n-1} \kappa_i)$$
$$A_n^m = A_n^m(t + \sum_{i=\underline{n}}^{n-1} \kappa_i)$$

Since the toll lane is under free flow speed $v_f$, $\kappa_n$ represents the travel time between toll entrance $n$ and toll entrance $n+1$ on the toll lane. Therefore $t + \sum_{i=\underline{n}}^{n-1} \kappa_i$ is the time when traffic arrives at toll entrance $n$ given that at $t$ the traffic is at $\underline{n}$. In other words, $t+\sum_{i=\underline{n}}^{n-1} \kappa_i$ is the corresponding time of toll entrance $n$ along the characteristic line. From equation (4.23), we can see that in the sub problem (4.22), we only focus on the revenue and capacity constraint along the characteristic line: $q_n^m$ is the incoming traffic, which destined at toll exit $m$, on the toll lane right before toll entrance $n$ on the characteristic line. The constrain in equation (4.22) guarantees that the capacity could not be exceeded along the characteristic line. $p_n^m D_n^m$ is the revenue rate of toll entrance $n$ at the point of time on the characteristic line. Since the revenue rate of all the characteristic lines are independent of each other, the optimal solution must also maximize the revenue rate along this characteristic line.

$\overline{n}$ represents the index of the farthest toll entrance on the characteristic line the traffic could reach before time $T$. It is given in equation (4.24). $\hat{\rho}(x,t)$ follows the stochastic partial differential equation model (4.2) given the initial condition at time $t$. $q_{\underline{n}}^m(t)$ are observed and known at time $t$. Given the information of $\hat{\rho}(\cdot,t)$ and $q_{\underline{n}}^m(t)$, solving problem (4.22) gives the optimal price vector for toll entrance $\underline{n}$ at time $t$.

$$\overline{n} = \max\{k \in [\underline{n}, N] : \frac{l_k - l_{\underline{n}}}{v_f} \leq T - t\} \tag{4.24}$$

In the stochastic control problem (4.22), the solution for the optimal prices at toll entrance $n$ at time $t$ should only depend on $\hat{\rho}(\cdot,t)$ and $q_n^1, \cdots, q_n^{|\Phi_n|}$. The Bellman equation could be formulated as shown in equation (4.25): $J_n(\hat{\rho}(\cdot,t), q_n^1, \cdots, q_n^{|\Phi_n|})$

represents the optimal value function given the state of the system. $\sum_{m\in\Phi_n} D_n^m p_n^m$ is the immediate revenue rate collected at toll entrance $n$ at time $t$. $J_{n+1}(\hat{\rho}(\cdot, t+\kappa_n), q_{n+1}^1, \cdots, q_{n+1}^{|\Phi_{n+1}|})$ is the maximum expected total revenue rate starting from toll entrance $n+1$ along the characteristic line. $\mathbb{P}$ is the set of possible prices satisfying the constraint in equation (4.22).

$$\begin{aligned} J_n(\hat{\rho}(\cdot,t), q_n^1, \cdots, q_n^{|\Phi_n|}) &= \sup_{D_n^m \in \mathbb{D}} \{ \sum_{m\in\Phi_n} D_n^m p_n^m + E[J_{n+1}(\hat{\rho}(\cdot, t+\kappa_n), q_{n+1}^1, \cdots, q_{n+1}^{|\Phi_{n+1}|})]\} \\ q_{n+1}^m &= q_n^m + D_n^m \qquad \forall m \in \Phi_{n+1} \qquad (4.25) \\ \sum_{m\in\Phi_n} D_n^m + \sum_{m\in\Phi_n} q_n^m &\leq C_n \\ J_{\bar{n}+1}(\hat{\rho}, \cdot) &= 0 \qquad \forall \hat{\rho} \end{aligned}$$

#### 4.3.2.2 Uniqueness of the Solution

The stochastic control problem in (4.25) is much simpler than the original stochastic control problem in (4.19). $\hat{\rho}(\cdot, t)$ is a state variable in the sub stochastic control problem, the dimension of another state variable, $q^m(\cdot, t)$, reduces from a function to a constant variable $q_n^m$. This represents a significant simplification and helps in generating an efficient numerical solution. In addition, the dimension of decision variables in the sub stochastic control problem also decreases significantly. The decision variables in the original stochastic control problem include the price vector at all toll entrances, while the sub stochastic control problem only covers the price vector at one toll entrance. In conclusion, the dimension reduction brought by the separation makes the pricing problem solvable. This section will show that the stochastic control problem in (4.25) has a unique optimal solution. For simplification purpose, we simplify the notations as:

$$\begin{aligned} q_n &= [q_n^1, q_n^2, \cdots, q_n^{|\Phi_n|}] \\ D_n &= [D_n^1, D_n^2, \cdots, D_n^{|\Phi_n|}] \end{aligned}$$

$$p_n \quad = \quad [p_n^1, p_n^2, \cdots, p_n^{|\Phi_n|}]$$

We change the decision variables from $p_n$ to $D_n$ because the constraint on $D_n$ is linear, so changing the decision variables will make the constraint become linear.

**Theorem IV.5.** *In the sub stochastic control problem* (4.25)*, when $D \cdot p(D)$ is a concave function on $D$, then for any $n$, $J_n(\hat{\rho}, q_n)$ is a concave function on $q_n$ given $\hat{\rho}$ is fixed. And there exists a unique optimal solution $D_n^\star$ solving the stochastic control problem in* (4.25)*.*

*Proof of Theorem IV.5.* First, we will apply induction method to prove that $J_n(\hat{\rho}, q_n)$ is a concave function on $q_n$ when $\hat{\rho}$ is fixed for any $n$:

From the terminal condition: $J_{\overline{n}+1}(\hat{\rho}, \cdot) \equiv 0$, therefore $J_n(\hat{\rho}, q_n)$ is a concave function on $q_n$ when $n = \overline{n} + 1$. Now suppose that $J_n(\hat{\rho}, q_n)$ is a concave function on $q_n$, we are going to prove that $J_{n-1}(\hat{\rho}, q_{n-1})$ is also a concave function on $q_{n-1}$.

Let $q_{n-1}^1$ and $q_{n-1}^2$ denote any two realizations of $q_{n-1}$. Let $D_{n-1}^{*,1}$ and $D_{n-1}^{*,2}$ represent the optimal solution of $D_n$ for $q_{n-1}^1$ and $q_{n-1}^2$ respectively. For fixed $\hat{\rho}$, Therefore

$$\begin{aligned} J_{n-1}(\hat{\rho}, q_{n-1}^1) &= f(D_{n-1}^{*,1}) + E[J_n(\hat{\rho}(t+\kappa_{n-1}), q_{n-1}^1 + D_{n-1}^{*,1})] \\ J_{n-1}(\hat{\rho}, q_{n-1}^2) &= f(D_{n-1}^{*,2}) + E[J_n(\hat{\rho}(t+\kappa_{n-1}), q_{n-1}^2 + D_{n-1}^{*,2})] \end{aligned} \tag{4.26}$$

Because the constraint on $D_n$ is linear, for any $\lambda \in [0,1]$, $\lambda D_n^{*,1} + (1-\lambda) D_n^{*,2}$ must also be a feasible solution when $q_{n-1} = \lambda q_{n-1}^1 + (1-\lambda) q_{n-1}^2$. Therefore:

$$\begin{aligned} & J_{n-1}(\hat{\rho}, \lambda q_{n-1}^1 + (1-\lambda) q_{n-1}^2) \\ \geq \; & f(\lambda D_n^{*,1} + (1-\lambda) D_n^{*,2}) + E[J_n(\hat{\rho}(t+\kappa_{n-1}), \lambda(q_{n-1}^1 + D_n^{*,1}) + (1-\lambda)(q_{n-1}^2 + D_n^{*,2})] \\ \geq \; & \lambda f(D_n^{*,1}) + (1-\lambda) f(D_n^{*,2}) \\ & + E[\lambda J_n(\hat{\rho}(t+\kappa_{n-1}), q_{n-1}^1 + D_n^{*,1}) + (1-\lambda) J_n(\hat{\rho}(t+\kappa_{n-1}), q_{n-1}^2 + D_n^{*,2})] \\ = \; & \lambda(f(D_{n-1}^{*,1}) + E[J_n(\hat{\rho}(t+\kappa_{n-1}), q_{n-1}^1 + D_{n-1}^{*,1})]) \end{aligned} \tag{4.27}$$

$$
\begin{aligned}
&+(1-\lambda)(f(D_{n-1}^{*,2}) + E[J_n(\hat{\rho}(t+\kappa_{n-1}), q_{n-1}^2 + D_{n-1}^{*,2})]) \\
= \quad & \lambda J_{n-1}(\hat{\rho}, q_{n-1}^1) + (1-\lambda) J_{n-1}(\hat{\rho}, q_{n-1}^2)
\end{aligned}
$$

From the equations above, we proved the concavity of $J_{n-1}(\hat{\rho}, q_{n-1})$ by definition. According to induction, $J_n(\hat{\rho}, q_n)$ is concave on $q_n$ for any $n$.

Next we are going to prove that there exists a unique solution $D_n^\star$ for any $n$. Since $J_{n+1}(\hat{\rho}(t+\kappa_n), q_{n+1})$ is concave on $q_{n+1}$ for any $\hat{\rho}(t+\kappa_n)$. It must also be concave on $D_n$ because $q_{n+1}$ is a linear function of $D_n$. Then the objective function in (4.25) is a strictly concave function on $D_n$. The constraint is linear, therefore there exists a unique optimal solution for the stochastic control problem in (4.25). □

#### 4.3.2.3 Simulation-Based Numerical Algorithm

Section 4.3.2.2 shows that the stochastic control problem (4.25) has a unique solution. However it is not easy to obtain a closed form solution since $\hat{\rho}(\cdot, t)$ is a vector. Under Assumption IV.1 the evolution of $\hat{\rho}(\cdot, t)$ is independent of the decision variables. With this condition, we propose a simulation-based numerical algorithm to obtain an approximate optimal solution.

In the simulation-based numerical algorithm, we first simulate a number of scenarios, say $K$, for the evolution of $\hat{\rho}(\cdot, t)$ according to the underlying traffic flow model (4.2). Let $\hat{\rho}^k(\cdot, t)$ represents the evolution of $\hat{\rho}(\cdot, t)$ under scenario $k$. These $K$ scenarios approximate the stochastic process, and all scenarios are assumed to be equally likely with probability $1/K$. Then we develop an optimal pre-determined strategy, under which the expected total revenue over these $K$ scenarios is maximized with the capacity constraint being satisfied for each scenario. The model to obtain the optimal pre-determined pricing strategy for toll entrance $\underline{n}$ at time $t$ is formulated in equation (4.28). $\hat{\rho}^k(\cdot, t + \sum_{i=\underline{n}}^{n-1} \kappa_i)$ represents the density vector of the general lane at time $t + \sum_{i=\underline{n}}^{n-1} \kappa_i$ under scenario $k$. Similar to the notation simplification in equation

(4.22), $D_n^{m,k}$, $p_n^{m,k}$, $q_n^{m,k}$ and $\hat{\tau}_n^{m,k}$ are actually simplified notation of $D_n^m(t+\sum_{i=\underline{n}}^{n-1}\kappa_i)$, $p_n^m(t+\sum_{i=\underline{n}}^{n-1}\kappa_i)$, $q_n^m(t+\sum_{i=\underline{n}}^{n-1}\kappa_i)$ and $\hat{\tau}_n^m(t+\sum_{i=\underline{n}}^{n-1}\kappa_i)$ under scenario $k$. In the mathematical programming problem in (4.28), $q_{\underline{n}}^m$ ($m \in \Phi_{\underline{n}}$) is the input of the programming problem, since they are observed at time $t$. By changing the decision variables from $p_n^m$ to $D_n^m$, the mathematical programming problem becomes a convex programming problem and it can be solved efficiently.

$$
\begin{aligned}
&\max \sum_{m\in\Phi_{\underline{n}}} p_{\underline{n}}^m D_{\underline{n}}^m + \frac{1}{K}\sum_{k=1}^{K}[\sum_{n=\underline{n}+1}^{\overline{n}}\sum_{m\in\Phi_n} p_n^{m,k} D_n^{m,k}] \\
&q_n^{m,k} = q_{n-1}^{m,k} + D_{n-1}^{m,k} \quad \forall n \in [\underline{n}+1, \overline{n}], \forall m \in \Phi_n, \forall k \in [1,K] \\
&q_{\underline{n}}^{m,k} = q_{\underline{n}}^m \qquad \forall m \in \Phi_n, \forall k \in [1,K] \\
&\sum_{m\in\Phi_n} q_n^{m,k} + \sum_{m\in\Phi_n} D_n^{m,k} \le C_n \quad \forall n \in [\underline{n}, \overline{n}], \forall m \in \Phi_n, \forall k \in [1,K] \qquad (4.28) \\
&D_n^{m,k} = A_n^m \frac{1}{1+\exp(\alpha_n^m(\tau_n^m - \hat{\tau}_n^{m,k}) + \eta_n^m p_n^{m,k} + \gamma_n^m)} \quad \forall n \in [\underline{n}+1, \overline{n}], \forall m \in \Phi_n, \forall k \in [1,K] \\
&D_{\underline{n}}^{m,k} = A_{\underline{n}}^m \frac{1}{1+\exp(\alpha_{\underline{n}}^m(\tau_{\underline{n}}^m - \hat{\tau}_{\underline{n}}^{m,k}) + \eta_n^m p_{\underline{n}}^m + \gamma_{\underline{n}}^m)} \quad \forall m \in \Phi_{\underline{n}}, \forall k \in [1,K] \\
&\tau_n^m = \frac{\tilde{l}_m - l_n}{v_f} \quad \forall n \in [\underline{n}, \overline{n}] \quad \forall m \in \Phi_n \\
&\hat{\tau}_n^{m,k} = f_n^m(\hat{\rho}^k(\cdot, t+\sum_{i=\underline{n}}^{n-1}\kappa_i)) \quad \forall n \in [\underline{n}, \overline{n}] \quad \forall m \in \Phi_n, \forall k \in [1,K]
\end{aligned}
$$

The pre-determined strategy means that, at time $t$ when solving the sub optimal problem for toll entrance $n$, actually we will obtain one unique price vector for toll entrance $n$ and $K$ price vectors for the subsequent toll entrances along the characteristic line, where each price vector is the corresponding pricing strategy for one scenario. The reason why there is unique price vector for toll entrance $n$ is that the price decision of this toll entrance is made before knowing which scenario will occur. However for the subsequent toll entrances, before making the price decisions, we observe the scenario and thus choose the corresponding price vectors for this scenario. Therefore we generate $K$ price vectors for subsequent toll entrances when solving the

sub optimal problem for toll entrance $n$. For every chosen time interval, when the price is to be updated, we solve $N$ such sub optimal problems for all $N$ toll entrances to obtain the new prices.

In conclusion, to solve this problem, a Markov decision process is first formulated and the Bellman equation is derived. However due to the high dimension of both state variables and decision variables, the Bellman equation is very difficult to solve. Under the assumption of the independence between the general lane traffic evolution and the toll prices, the original problem is decomposed to $N$ sub problems and a discrete Markov decision process model is formulated for each of the $N$ toll entrances. Each sub problem solves the optimal prices for one toll entrance. Finally, a simulation based numerical algorithm is applied to obtain an approximate optimal solution to the original problem.

## 4.4 Numerical Case Study

This section will illustrate how the algorithm works to solve the optimal pricing problem for a multi-entry and multi-exit managed toll lane. In this numerical example, a 15-miles long managed toll lane is assumed to be located in parallel with a general lane. There are 4 toll entrances located at 0, 4, 8 and 12 miles, and 4 toll exits located at 3, 7, 11 and 15 miles from the start of the managed toll lane. The last toll exit is at the end of the managed toll lane.

The experiment will simulate the scenario of the managed toll lane system, including the evolution of traffic density on the general lane and the optimal toll prices on each individual toll entrance, during the morning peak hours from 5am to 10am. The traffic density evolution on the general lane is assumed to follow the stochastic model in equation (2.11) and the density evolution can be simulated according to the steps in Algorithm 2 in Section 2.3.2.3. The highway is equally divided to 75 cells and the length of each cell is 0.2 miles. In the following cells: 10, 13, 16, 19,

22, 25, 45, 48, 51, 54, 57 and 60, the parameters $a$, $b$ and $\sigma$ in the stochastic traffic flow model are assumed to evolve as shown in Figure 4.3. For the other cells, $a$ and $b$ are constantly equal to 0, while $\sigma$ follows the same evolution as shown in Figure 4.3. From the figures, we can see that at around 5:40am, $a$ starts to jump from 500 vehicles/mile/hour to 2400 vehicles/mile/hour, while $b$ starts to drop from $-10$ to $-15$ and $\sigma$ starts to rise from 5 to 15. Then at around 8:00am, those parameters gradually revert back to their normal values. The reason behind this assumption on parameter evolution is that starting from 5:40am, the traffic system starts to enters into the peak hour, and there are more vehicles entering and leaving the highway, so the absolute values of both $a$ and $b$ increase. In addition, the volatility $\sigma$ also increases because the congestion will cause more uncertainty on the traffic condition. When the congestion starts to diminish at around 8:00am, the absolute values of $a$, $b$ and $\sigma$ revert back to their normal values. Figure 2.10 in Section 2.4.2 also shows that, in reality, the absolute values of $a$, $b$ and $\sigma$ hover during peak hours.

The demand function for each toll station follows equation (4.11). Table 4.2 summarizes the values of all the parameters used in this experiment. $A$ is the demand matrix where $A_{ij}$ is the potential demand between entrance $i$ and exit $j$. We assume that the demand matrix $A$ is static. This assures that the congestion toll prices are determined by the traffic conditions and not due to the changes in demand. This then illustrates how the traffic state condition affects the optimal congestion prices. As we see in the utility function, $\alpha$ is the sensitivity to the time saving, $\eta$ is the sensitivity to the price, and $\gamma$ is the preference factor. They are assumed to be constant for different pairs of origin and destination. These parameters are heavily dependent on the income and sensitivity of time saving of the residents in the region. For simplification, we assume they are the same at each toll entrance. $C$ is the capacity where $C_i$ is the capacity right after toll entrance $i$. $v_f$ is the free flow speed on both the managed toll lane and the general lane.

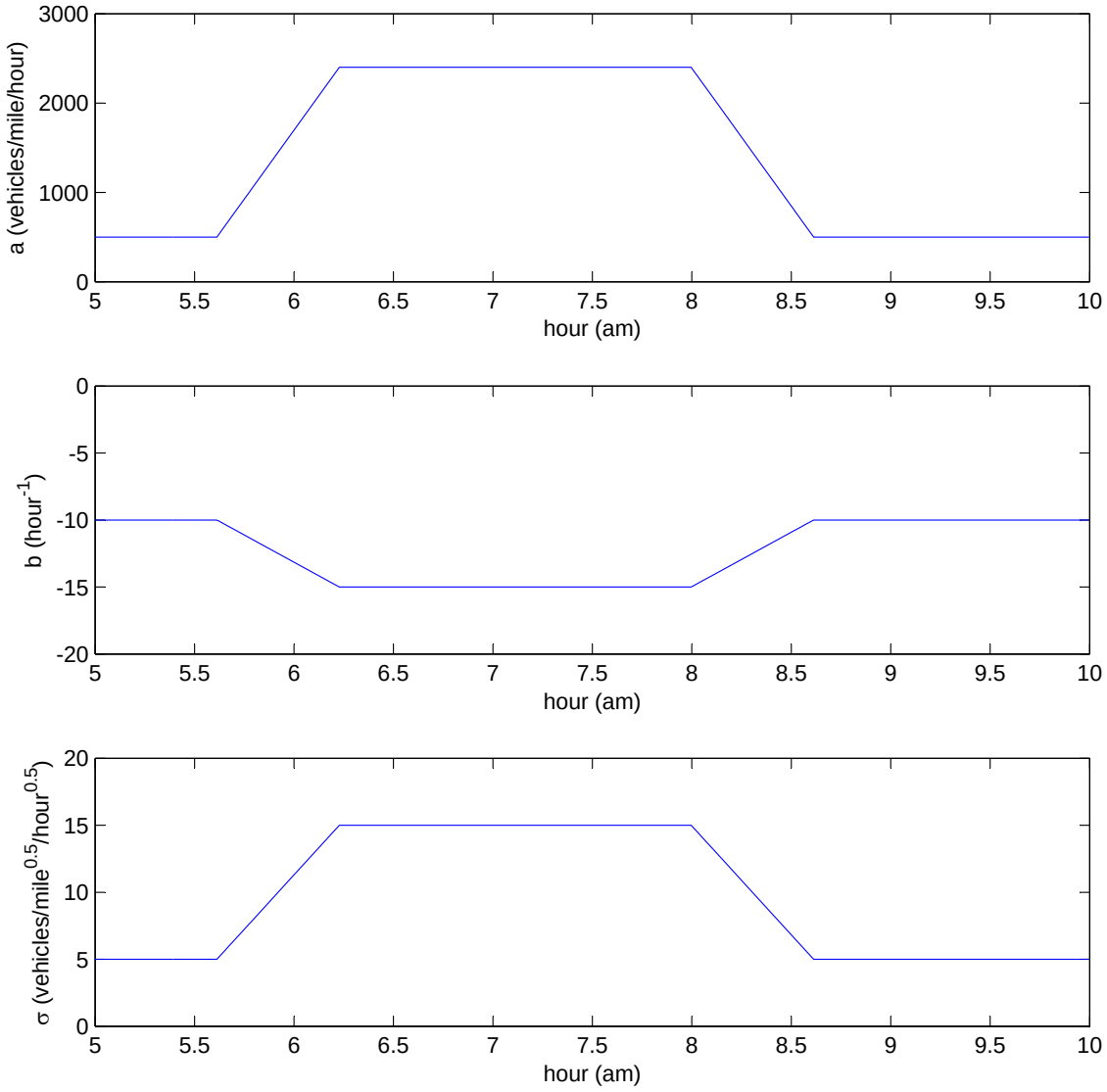



Figure 4.3: Parameters in the stochastic traffic flow model used for simulation.

Figure 4.4 shows the evolution of the traffic density on the general lane, and Figure 4.5 illustrates the speed evolution on the general lane. From the figures, we can see that congestion on the general lanes starts at around 6:00am and continues to deteriorate until around 8:30am due to hover of the absolute values of parameters $a$, $b$ and $\sigma$. From around 8:30am, the congestion starts to alleviate because the parameters revert back to their normal values and the general lane reaches the free flow condition at around 9:00am.

| Parameters | Values |
|---|---|
| Demand Matrix | $A = \begin{bmatrix} 1400 & 1400 & 1400 & 1400 \\ 0 & 1400 & 1400 & 1400 \\ 0 & 0 & 1400 & 1400 \\ 0 & 0 & 0 & 1400 \end{bmatrix}$ vehicles/hour |
| Time Sensitivity | $\alpha = 7$ hour$^{-1}$ |
| Price Sensitivity | $\eta = 1.5$ \$$^{-1}$ |
| Preference Factor | $\gamma = 0.69$ |
| Capacity Vector | $C = [1800 \quad 1800 \quad 1800 \quad 1800]^T$ vehicles/hour |
| Free Flow Speed | $v_f = 65$ miles/hour |

Table 4.2: Parameters in the numerical experiment

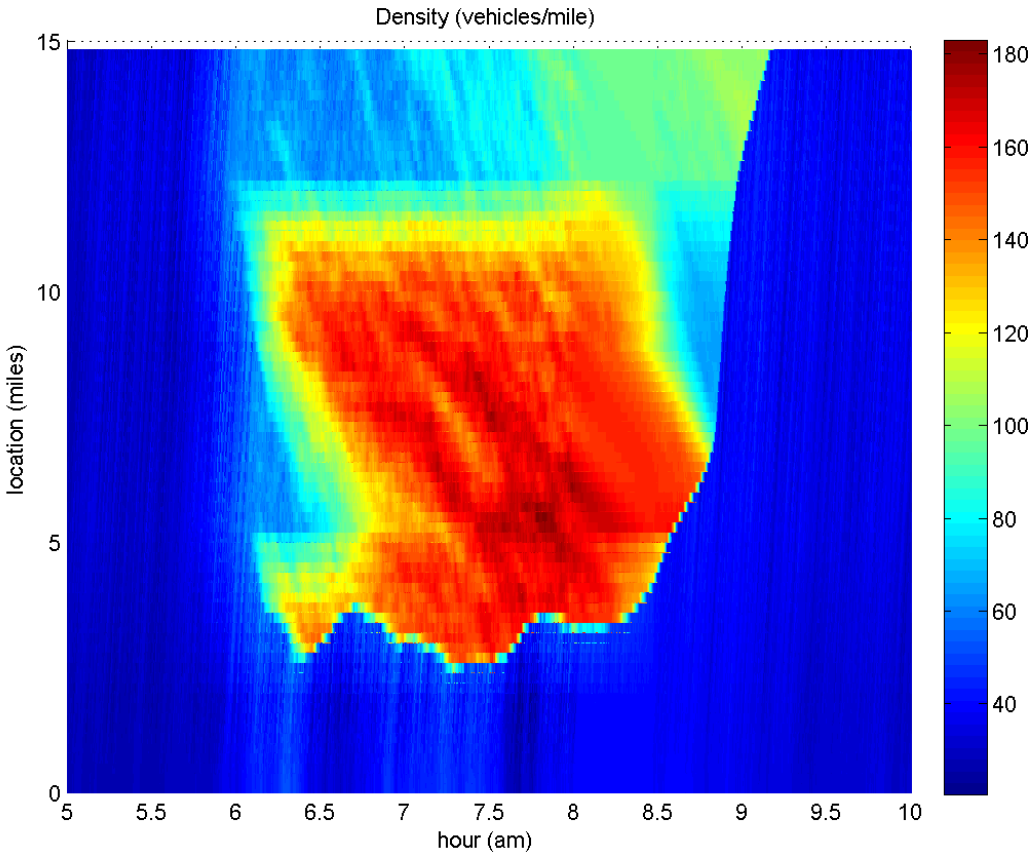



Figure 4.4: Density evolution on the general lane

Under the general lane traffic condition in Figure 4.4 and Figure 4.5, the optimal toll prices of each individual toll entrances are shown in figure 4.6. From the figures, we can see that the prices increases when the general lane becomes congested and decreases when the general lane is running smooth.

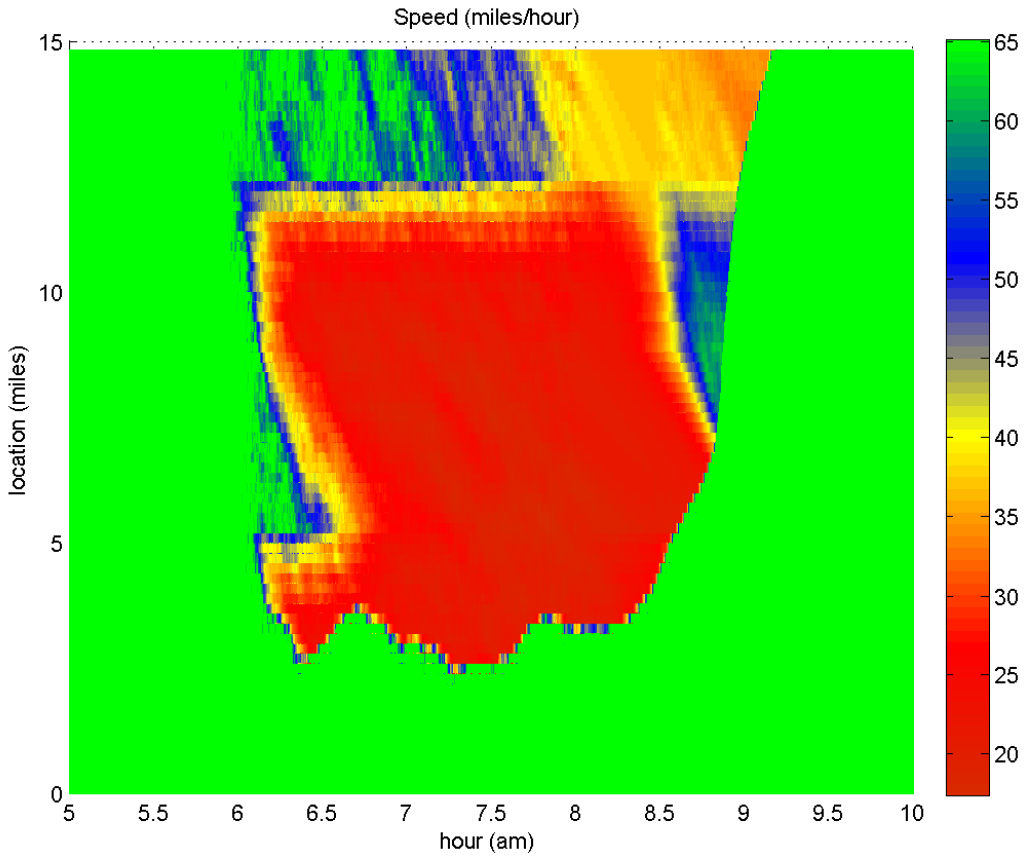



Figure 4.5: Speed evolution on the general lane

Figure 4.7 plots the evolution of the volume on the toll lane under the optimal toll pricing strategy. From this figure, we can see that the volume along the toll lane is maintained below the volume capacity, and traffic along the toll lane is maintained at the free flow condition during the peak hour. Since the traffic on the toll lane travel under free flow speed, we can see there are characteristic lines with slope $v_f$ in Figure 4.7.

Figure 4.8 plots the cumulative revenue of each individual toll entrance, we can see that the revenue increases faster during the peak hours from 6:00am to 8:00am, which means that more congested on the general lane, the more is the revenue collected at the toll entrances.

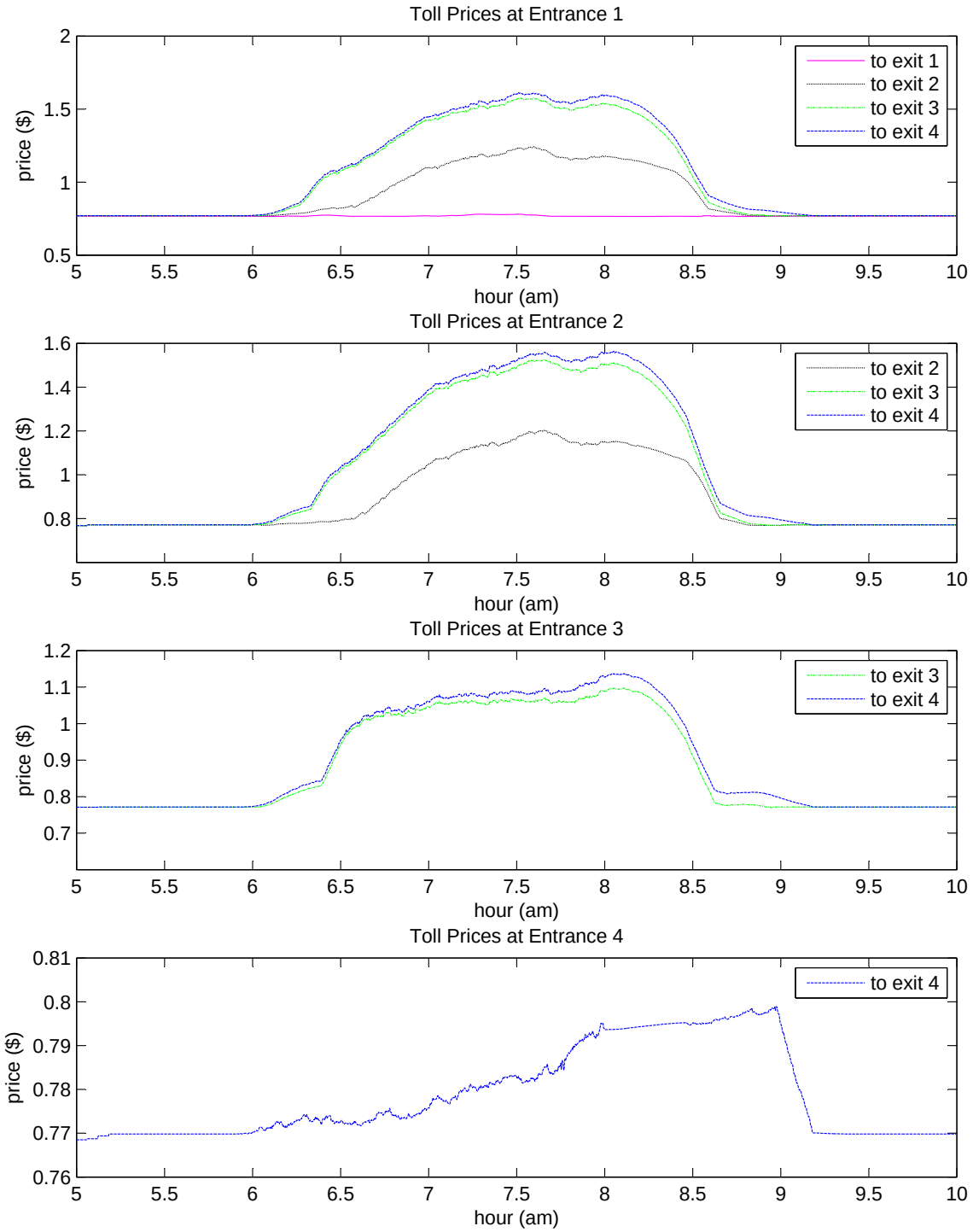



Figure 4.6: Optimal toll prices on each individual toll entrance

## 4.5 Conclusion

In this chapter, we investigate a dynamic distance-based pricing strategy for managed toll lane with multiple toll entrances and exits. The mathematical model devel-

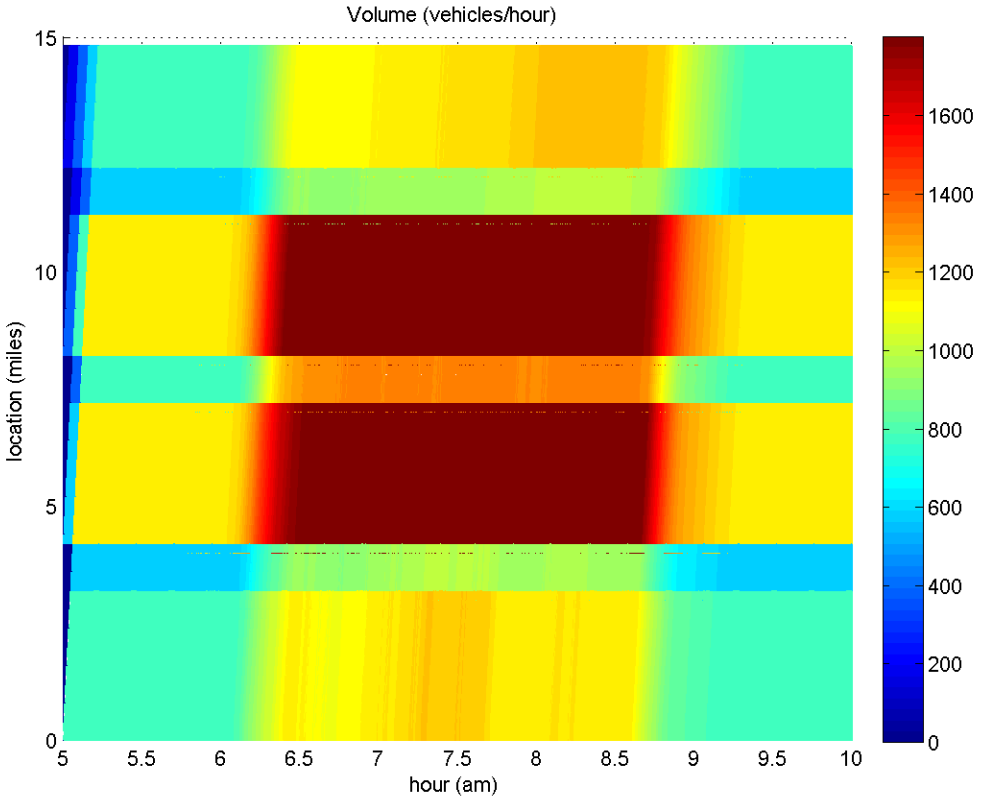



Figure 4.7: Volume evolution on the toll lane

oped in this chapter is very general and can be easily applied in practice. We use a stochastic partial differential equation model to describe the traffic evolution of the general lane. However the general pricing model developed in this paper is not limited to specific traffic flow model and is readily adapted to other macroscopic traffic models, such as the classical LWR model. Any model that fits the real data can be used in the pricing model as long as the traffic flow model has the ability to predict the travel time. This pricing model can be used with other objective functions, such as maximizing the total throughput, and thus a comparison of the pricing strategies under different objectives can be carried out. In another direction, an empirical study of the pricing model can be examined as well: First obtain real data on both toll prices and traffic flow; calibrate the stochastic model by using the real data; plug in calibrated traffic model into the pricing model to obtain the optimal prices, and

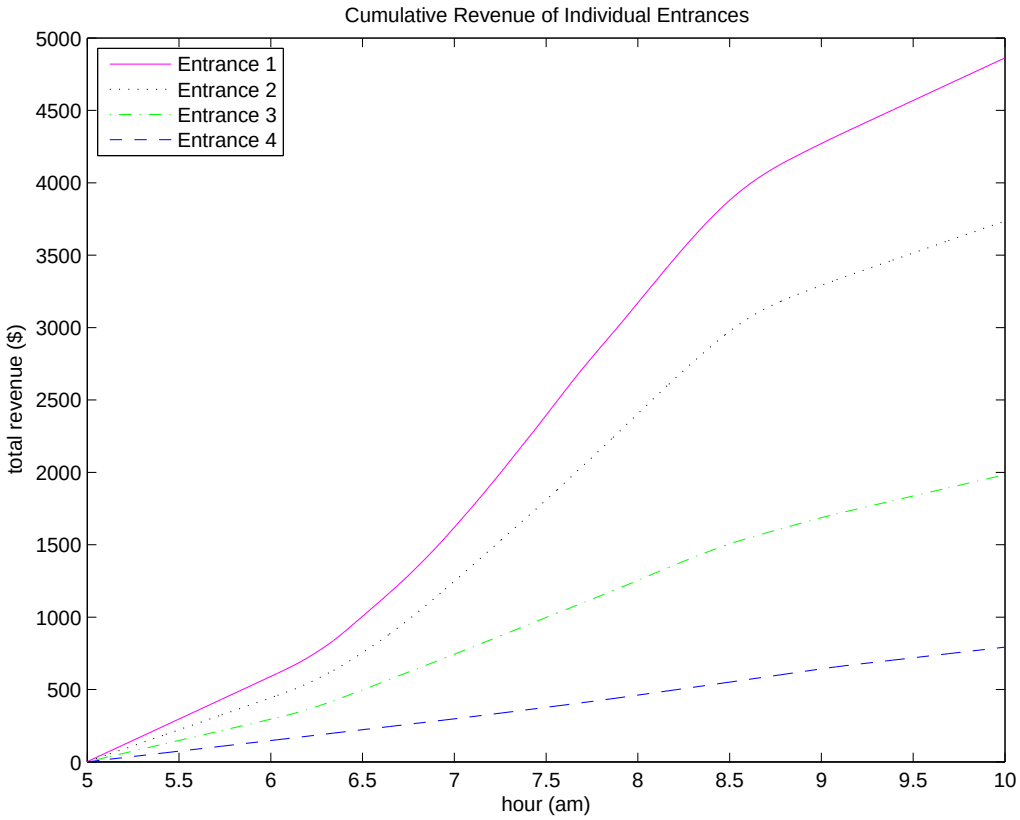



Figure 4.8: Cumulative revenue collected by each individual toll entrance

compare theoretical prices with the real prices. A simulation can also be used to get an idea of the increment in revenue by implementing the developed optimal pricing strategy.

# CHAPTER V

# Conclusions and Future Work

## 5.1 Conclusions and Summary of Contributions

This dissertation systematically studies and discusses an innovative stochastic traffic flow model and its application. The three main chapters in this dissertation are closely connected. In Chapter II, we propose the stochastic traffic flow model and investigate the off-line calibration algorithm for this model. The numerical procedure to predict the future traffic state based on this stochastic model is examined as well. Based on the result of model validation using real highway data, the stochastic model outperforms, in terms of prediction accuracy, the traditional macroscopic traffic flow model, and has been proved to be an accurate model capturing the characteristic of traffic flow evolution. Then in Chapter III, the on-line calibration algorithm for this stochastic traffic flow model is proposed. Various filtering techniques is analyzed and their performance on the speed of tracking the change of parameters are studied and compared. The numerical results show that the on-line calibration algorithm not only captures the change of model parameters but also improves the prediction accuracy of the traffic flow model. Last in Chapter IV, we formulate a general mathematical model for the problem of optimal dynamic distance-based congestion pricing. This mathematical model incorporates the traffic flow model proposed in Chapter II and the on-line calibration technique developed in Chapter III. The optimal pricing strat-

egy under a specified objective is explored. The integration of these three chapters makes the analysis of this stochastic traffic flow model systematic and complete.

The major contributions of this dissertation are summarized as follows:

1. This dissertation introduces, for the first time, Brownian sheet into the transportation engineering area to model the stochastic behavior of traffic flow evolution. This state-of-the-art modeling methodology makes the traffic flow model more robust and flexible than traditional deterministic macroscopic traffic flow models. In the quantitative finance area, since Fischer Black and Myron Scholes applied the geometric Brownian motion to model the evolution of stock price and derived the option pricing formula(*Black and Scholes*, 1973), many other derivative pricing problems have been solved based on this model. Therefore a good underlying model can be a significant for the development for this area.

2. This dissertation presents a complete and detailed procedure on traffic flow modeling: model assumption, model calibration and model validation. There is not much work which uses real traffic data to validate a traffic flow model. This dissertation uses a case study to explicitly explain and validate each step in traffic modeling. It also describes simulation can be used to give the prediction on future traffic state based on the stochastic traffic flow model.

3. This dissertation reviews the various filtering algorithms that can be used for on-line traffic state and parameter estimation. It compares and summarizes the performance of different on-line filtering algorithms, and can be used as a comprehensive document in this area for future research.

4. To the author's best knowledge, this dissertation presents the first ever attempt to rigorously investigate the problem of dynamic distance-base optimal congestion pricing for multi-entry and multi-exit managed lanes. It formulates a general framework for this problem, and the solution of the optimal pricing

strategy can be adapted to any preferred objective function and underlying traffic flow model. Under this general framework, the model can be extended easily. The flexibility and the efficiency of the developed algorithms for the solution of this model makes it very adaptable to on-line use.

## 5.2 Future Work

To further investigate this traffic flow model and its applications, there are many more topics that can be explored. Several directions for future research related to this dissertation are described as below:

1. Analytic solution of the SPDE.

   In this dissertation, we numerically solve the stochastic partial differential equation using a simulation based Godunov's scheme. It has been shown that weak solutions exist, but it is not known what the properties (like pdf etc.) of the resulting random variables are. Getting an analytic solution would be very helpful. The analytic solution gives the probability distribution of the future traffic state explicitly without the large computation cost incurred by a simulation. In financial engineering, the analytic solution of the geometric Brownian motion results in the closed-form formula of the option price, as found by Black-Scholes. The analytical solution of the stochastic partial differential equation will open doors to many other interesting research areas.

2. On-line calibration incorporating Lagrangian measurement.

   In this dissertation, we developed an on-line calibration algorithms using Eulerian measurement. In the future, as more and more Lagrangian measurement become available, the investigation of filtering algorithm using the combination of Eulerian and Lagrangian measurements becomes meaningful and important.

We believe that the incorporation of Lagrangian measurement will make the estimation more accurate.

3. Verification of the optimal pricing strategy under practical operation.

   In this dissertation, we mathematically formulated the dynamic optimal pricing model and derive the methods to output the optimal pricing strategy. However in this model, some parameters, such as the parameters in the demand function, still need to be calibrated using practical operational data. Thus it will be important to validate these assumptions with the performance of the pricing strategy developed from this mathematical model in real highway operation.

# BIBLIOGRAPHY

# BIBLIOGRAPHY


Arnold, E. (1998), Ramp metering: a review of the literature, *Virginia Transportation Research Council Technical Report*.

Arulampalam, M., S. Maskell, N. Gordon, and T. Clapp (2002), A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking, *IEEE Transactions on Signal Processing*, *50*(2), 174–188.

Black, F., and M. Scholes (1973), The pricing of options and corporate liabilities, *The Journal of Political Economy*, pp. 637–654.

Cameron, G., and G. Duncan (1996), Paramics-parallel microscopic simulation of road traffic, *The Journal of Supercomputing*, *10*(1), 25–53.

Choa, F., R. Milam, and D. Stanek (2004), Corsim, paramics, and vissim: What the manuals never told you, *Proceedings of the 9th TRB Conference on the Application of Transportation Planning Methods*.

Chung, C., and W. Recker (2011), State-of-the-art assessment of toll rates for high-occupancy and toll lanes, *Transportation Research Board 90th Annual Meeting*.

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

Daganzo, C. (1995), Requiem for second-order fluid approximations of traffic flow, *Transportation Research Part B: Methodological*, *29*(4), 277–286.

Doucet, A., N. De Freitas, and N. Gordon (2001), *Sequential Monte Carlo methods in practice*, Springer Verlag.

FHWA (2008), *Status of the Nation's Highways, Bridges, and Transit: 2008 Conditions and Performance*.

Godunov, S. (1959), A difference scheme for numerical computation of discontinuous solution of hydrodynamic equations, *Math Sbornik*, *47*, 271–306.

Gunnarsson, G. (2006), *Stochastic partial differential equation models for highway traffic*, ProQuest.

Halati, A., H. Lieu, and S. Walker (1997), Corsim-corridor traffic simulation model, *Traffic Congestion and Traffic Safety in the 21st Century: Challenges, Innovations, and Opportunities*, pp. 570–576.

Hegyi, A., D. Girimonte, R. Babuska, and B. De Schutter (2006), A comparison of filter configurations for freeway traffic state estimation, in *Proceedings of the 2006 IEEE conference on Intelligent Transportation Systems*, pp. 1029–1034, IEEE.

Hirsch, C. (1990), *Numerical computation of internal and external flows. Volume 2: Computational methods for inviscid and viscous flows*, Wiley Series in Numerical Methods in Engineering.

Hoogendoorn, S., and P. Bovy (2001), State-of-the-art of vehicular traffic flow modelling, *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, *215*(4), 283–303.

Julier, S., and J. Uhlmann (2004), Unscented filtering and nonlinear estimation, *Proceedings of the IEEE*, *92*(3), 401–422.

Kalman, R., et al. (1960), A new approach to linear filtering and prediction problems, *Journal of Basic Engineering*, *82*(1), 35–45.

Lax, P. (1972), The formation and decay of shock waves, *The American Mathematical Monthly*, *79*(3), 227–241.

LeVeque, R. (1992), *Numerical methods for conservation laws*, Birkhäuser Verlag.

LeVeque, R. (1998), Balancing source terms and flux gradients in high-resolution Godunov methods: The quasi-steady wave-propagation algorithm, *Journal of Computational Physics*, *146*(1), 346–365.

Lighthill, M., and G. Whitham (1955), 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.

Lou, Y., Y. Yin, and J. Laval (2011), Optimal dynamic pricing strategies for high-occupancy/toll lanes, *Transportation Research Part C: Emerging Technologies*, *19*(1), 64–74.

Mihaylova, L., R. Boel, and A. Hegyi (2007), Freeway traffic estimation within particle filtering framework, *Automatica*, *43*(2), 290–300.

Payne, H. (1971), Models of freeway traffic and control., *Mathematical models of public systems*.

Ribeiro, M. (2004), Kalman and extended Kalman filters: Concept, derivation and properties, *Institute for Systems and Robotics*, p. 43.

Richards, P. I. (1956), Shockwaves on the highway, *Operations Research*, *4*, 42–51.

Schrank, D., and T. Lomax (2009), 2009 Urban Mobility Report, *Texas Transportation Institute*.

Smith, B., and M. Demetsky (1994), Short-term traffic flow prediction: neural network approach, *Transportation Research Record: Journal of the Transportation Research Board*, *1453*, 98–104.

Stathopoulos, A., and M. Karlaftis (2003), A multivariate state space approach for urban traffic flow modeling and prediction, *Transportation Research Part C: Emerging Technologies*, *11*(2), 121–135.

Sun, X., L. Muñoz, and R. Horowitz (2003), Highway traffic state estimation using improved mixture Kalman filters for effective ramp metering control, in *42nd IEEE International Conference on Decision and Control*, vol. 6, pp. 6333–6338.

Van Der Voort, M., M. Dougherty, and S. Watson (1996), Combining kohonen maps with ARIMA time series models to forecast traffic flow, *Transportation Research Part C: Emerging Technologies*, *4*(5), 307–318.

Walsh, J. (1986), An introduction to stochastic partial differential equations, *Lecture Notes in Math*, *1180*(265), 439.

Wang, Y., and M. Papageorgiou (2005), Real-time freeway traffic state estimation based on extended Kalman filter: a general approach, *Transportation Research Part B: Methodological*, *39*(2), 141–167.

Whitham, G. (1974), *Linear and nonlinear waves*, Wiley (New York).

Wong, G., and S. Wong (2002), 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.

Work, D., O. Tossavainen, Q. Jacobson, and A. Bayen (2009), Lagrangian sensing: traffic estimation with mobile devices, in *Proceedings of the 2009 conference on American Control Conference*, pp. 1536–1543.

Xiao, H., H. Sun, B. Ran, and Y. Oh (2003), Fuzzy-neural network traffic prediction framework with wavelet decomposition, *Transportation Research Record: Journal of the Transportation Research Board*, *1836*, 16–20.

Yang, H., W. Xu, B. He, and Q. Meng (2010), Road pricing for congestion control with unknown demand and cost functions, *Transportation Research Part C: Emerging Technologies*, *18*(2), 157–175.

Yang, Q., and H. Koutsopoulos (1996), A microscopic traffic simulator for evaluation of dynamic traffic management systems, *Transportation Research Part C: Emerging Technologies*, *4*(3), 113–129.

Zhang, G., Y. Wang, H. Wei, and P. Yi (2008), A feedback-based dynamic tolling algorithm for high-occupancy toll lane operations, *Transportation Research Record: Journal of the Transportation Research Board*, *2065*, 54–63.

Zhang, H. (1998), A theory of nonequilibrium traffic flow, *Transportation Research Part B: Methodological*, *32*(7), 485–498.

Zhang, H. (2000), Structural properties of solutions arising from a nonequilibrium traffic flow theory, *Transportation Research Part B: Methodological*, *34*(7), 583–603.