Utilization of Probe Vehicle Data to Estimate Urban Traffic Conditions

Peng CAO

July 2014

Utilization of Probe Vehicle Data to Estimate Urban Traffic Conditions

Doctoral Dissertation

Submitted in Partial Fulfillment of the
Requirement for the Degree of
Doctor of Engineering

By
Peng CAO
July 2014

Academic Adviser:
Associate Professor Tomio MIWA

Department of Civil Engineering
Nagoya University
JAPAN

Abstract

Traffic congestion is a crucial problem that adversely and significantly affects the environment and transport efficiency in numerous cities. Efforts to solve this problem lead to the development of intelligent transportation systems (ITS). An essential input for ITS is accurate and reliable knowledge about traffic conditions. The objective of this research is to estimate traffic conditions from probe vehicle data for urban networks.

In traffic flow theory, flow, speed (or travel time) and density are three fundamental macroscopic variables representing traffic conditions in the spatial and temporal domains. Therefore, to provide comprehensive knowledge of urban traffic, this research explores estimation of various measurements of traffic conditions from these three aspects. Specifically, firstly methods are proposed to estimate traffic flow (including link flow and origin-destination (O-D) flow) from travel times of probe vehicles. Secondly link travel time distribution is modelled and discussed. Then methodology of traffic monitoring is explored.

In Chapter 1, the background of this research is firstly introduced. And then measurements of traffic conditions and characteristics of probe vehicle data are extracted and discussed. Further, challenges of estimating urban traffic conditions are summarized and analyzed. Then, research objectives are explicitly defined. This is followed by an organization of this thesis.

Chapter 2 gives a detailed review of literatures which most relate with this research. The mechanism of probe vehicle is firstly introduced. And then issues involved in probe vehicle data including data types, polling schemes, penetration and applications are reviewed. Finally, methodologies of estimating flow, travel time and density using probe vehicle data are reviewed, respectively.

Chapter 3 provides a methodology of estimating dynamic link flow from raw probe vehicle data. From the view of practical application, this methodology consists of three steps: travel time allocation, link performance function fitting, and link flow estimation. In the first step, method of proportional allocation is used to decompose probe travel time onto individual links. In the second step, link performance function is obtained from a speed-density function derived from Gazis's nonlinear follow-the-leader model. In the third step, a Bayesian method that incorporates prior distribution of link speed is applied. A traffic network called Kichijoji-Mitaka area is developed and validated in VISSIM according to a rather complete benchmark data set based on this area. The proposed method is tested by using simulation data in this network. Additionally, the effects of polling frequency and penetration of probe vehicle are tested and analyzed. Results show, the Bayesian method can give acceptable estimates of link flows. This methodology avoids of assumption of random sampling and is recommended to be put into practice for urban networks.

Chapter 4 describes a bi-level generalized least square (GLS) model to estimate dynamic O-D flow from estimated link flow in Chapter 3 and historical O-D flow. It is an extension of Tavana's model. Both the distance between the estimated and target OD matrices and the distance between the calculated and observed link flows are considered in the objective function. Hence the extended Bell algorithm is used to solve the upper level of this model. For the lower level, the dynamic traffic assignment (DTA) module in VISSIM is applied. Moreover, the GLS formulation can utilize the variance of estimated link flow from Chapter 3. However setting an identical variance of link flow for links decreases GLS to be ordinary least square (OLS) as Tavana's model. A case study on Kichijoji-Mitaka area validates the superior of the bi-level GLS to OLS. And it illustrates the application of microscopic traffic simulator like VISSIM in solving the lower level of bi-level estimation of O-D flow.

As a whole, the proposed methods in Chapter 3 and Chapter 4 are capable of estimating link flow and dynamic O-D flow using probe data in practice, such as in probe vehicle-based dynamic route guidance system, or in situation that link counts are not available.

Chapter 5 presents the formulation of link travel time distribution in the signalized road section by using truncated distribution. Link travel time is decomposed into time-in-motion and time-in-queue. The time-in-motion mainly depends on physical attributes of the link and traffic conditions, and is modelled by a truncated distribution. The time-in-queue is mainly related to signal timings and the signal offset between adjacent intersections, and is derived from hydrodynamic theory and horizontal queuing theory. The derived link travel time distribution is parameterized by fraction of stationary vehicles, red signal time, and motion behavior parameters including truncation points. These parameters are estimated from travel times provided by probe vehicles using a maximum likelihood estimator. It is shown that it's better to model time-in-motion with a truncated distribution instead of a non-truncated one using AICc and BIC criteria. And the results also validate the superiority of applying a truncated distribution to model the distribution of link travel times. It is recommended that truncation of travel time should be considered and which imposes effects on travel time variability and reliability.

Researches in Chapter 3-5 utilize link travel times provided by part of probe vehicles, which has relative high polling frequency. Link travel times become less accurate and reliable for lower polling frequency probe vehicle data because of the difficulties of map-matching and path travel time allocation. Therefore, Chapter 6 explores methodology of mining traffic information from scatted probe vehicle data regardless of polling frequency.

Chapter 6 describes the formulation of a joint probability distribution of vehicle

location and speed on arterial road using hydrodynamic theory and horizontal queuing theory. By assuming that sample vehicle locations are proportional to the traffic densities, the probability distribution function (PDF) of vehicle location is derived. Conditioned on certain location, speed is distributed according to a mixture distribution of distributions with different densities. Then the joint PDF of vehicle location and speed is derived using multiplication rule. Particularly, the joint PDF is derived for both under-saturated and saturated regimes. This model is parameterized by link parameters (cycle light time, red light time, etc.), parameters of driver behavior, and traffic states (densities and remaining queue length). To validate these distributions, a Kolmogorov-Smirnov test is performed using probe data collected during a field test in Toyota City. The numerical results validate the use of proposed models for both under-saturated and congested regimes. Additionally, the proposed models capture most features of vehicle location and speed distributions. Additionally, the derived models are simplified by neglecting some small details to improve the computation efficiency when applying it into network-level estimation. The proposed model is a brand new model which models location and speed distributions simultaneously. The derived joint PDF makes it possible to learn macroscopic traffic parameters and link parameters from location and speed data of probe vehicles.

Chapter 7 concludes this research by summarizing the proposed models and their performance, and pointing out the limitations of this research and indicating the future researches.

Acknowledgements

I would like to take this opportunity to express my sincere thanks to all people who have made contribution on completion of this dissertation.

First of all, I am very grateful to my academic supervisor Associate Professor Tomio MIWA for his continuously careful guidance, warmhearted encouragement and valuable suggestions in my research and my daily life of studying in Nagoya University. As I transferred my major subject from Logistics to Transportation, my knowledge was in short of transportation at the beginning of my doctoral course. I am very grateful to Associate Professor Tomio MIWA for he teaches me knowledge of transportation from basic concepts to advanced techniques step by step, and provides me various materials for learning. Besides, his instructions also include philosophy of life, principle of doing things and spirit of doing research. I will benefit from these in my future life.

I am very grateful to Professor Takayuki MORIKAWA as the vice academic supervisor for his great foresight guidance on determining the subject of my research and giving valuable suggestions in the process of studying my research. His solid knowledge of traffic engineering and economics helped me to broaden my view and consider problems from a comprehensive perspective. What's more, his life style and philosophy of doing things are worth of my continuously learning and appreciating.

I am very grateful to Professor Toshiyuki YAMAMOTO as the vice academic supervisor for his careful guidance on the knowledge of traffic modelling and research techniques. His insight comments helped me conduct the research in depth. And I greatly appreciate that he has given me many valuable comments and detail suggestions on the draft of this dissertation.

I also express my sincere appreciation to Professor Yukimasa MATSUMOTO in Meijo University, who is a member of my dissertation defense committee, for his valuable comments and warmhearted help. I am very grateful to Professor Hideki NAKAMURA, who is also a member of my dissertation defense committee, for his

valuable comments and teaching me knowledge of traffic engineering. And his spirit of doing research and serious attitude to research are worth of learning.

I also greatly appreciate the assistance from Lecturer Hitomi SATO during my study in Nagoya University. And I would like to express my thanks to Associate Professor Kai LIU in Dalian University of Technology for his valuable comments in the discussion of my research.

I would like to appreciate researcher Dr. Danpeng MA for his careful checking my thesis and valuable comments. Besides, I benefit a lot from his wide knowledge of humanities and social sciences. I would like to express my sincere thanks to my colleague and friend Mr. Jia YANG, since he gave me a lot of help in my living in Japan and he shared a lot of precious experiences with me which I will remember and cherish forever. I would like to appreciate researcher Dr. Rui MU for her help before and after I joined this lab. And we shared many happy memories together with Dr. Dawei LI who also deserves my thanks. Thanks were also extended to other students in our laboratory who had helped me during these three years on campus or outside of campus

I would like to express my sincere thanks to Professor Lixin MIAO, Associate Professor Benhe GAO and Dr. Qiang LI in Tsinghua University for their recommendation of studying in Nagoya University.

I would also express my deep gratitude to the Ministry of Education, Culture, Sports, Science and Technology in Japan, since the MEXT scholarship can guarantee me to concentrate myself to complete the thesis.

Finally, it is my greatest pleasure to dedicate this small achievement to my beloved parents. I have totally relied on the continuous support and encouragement of them throughout my study period with great love and patience.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Traffic congestion has been a crucial problem around the world, especially in cities where population and motorization increase rapidly. For example, the number of cars in Peking increased 800 000 new vehicles in 2010 from a total of 4 million vehicles at the beginning of this year. Both in perception and in reality, traffic congestion wastes resources and labors, causes environmental pollution and stress. According to an urban mobility report 2012 (Schrank *et al.*, 2012), it was estimated that traffic congestion caused Americans to spend 5.5 billion hours more in travel and to purchase an extra 2.9 billion gallons of fuel, which results a congestion cost of $121 billion in 498 U.S. Urban Areas. Besides direct economic loss, there are many kinds of externalities from congestion, including the negative impacts on public health and happiness because of air pollution and painful wait.

To alleviate traffic congestion, researchers and planners proposed and implemented multiple strategies including increasing transportation infrastructures and managing traffic efficiently. Among them, intelligent transportation systems (ITS) show significant effects and have been playing an important role in traffic control and management (Zhang *et al*., 2011).

1.1.1 Intelligent transportation systems

Intelligent transportation systems (ITS) is defined by the International Standards Organization Committee for ITS Standards (ISO TC204) as ‘‘information, communication and control systems in the field of urban and rural surface transportation, including intermodal and multimodal aspects, traveler information,

traffic management, public transport, commercial transport, emergency services and commercial services'' (Williams, 2008). The objective of ITS is to evaluate, develop, analyze and integrate new technologies to achieve traffic efficiency, save time, save energy, improve environment, enhance safety and comfort drivers, pedestrians, and other traffic groups (Vanajakshi, 2010)

U.S.A., Japan and Europe take the leaderships around the world in the development of ITS. Taking these countries as examples, we briefly introduce the history of ITS. As showed in Table 1.1, the history of ITS can be divided into three stages (Tokuyama, 1996). The first stage began from 1960s with the representatives of the Electronic Route Guidance System (ERGS) in the United States, the Japanese Comprehensive Automobile Traffic Control System in Japan, and a similar system named Autofahrer Leit and Information System (ALI) in Germany. These systems are the original prototype of ITS. Due to limitations of advanced technologies, they didn't result in practical applications. Benefiting from the technology reforms (e.g. the application of mass memory), the second stage from 1980s experienced a fast development. Multiple systems, including the U.S. Intelligent Vehicle Highway Systems (IVHS), the Road/Automobile Communication System in Japan (RACS), and the Program for a European Traffic System with Higher Efficiency and Unprecedented Safety (PROMETHEUS), were formulated and put into practical applications. From 1995, the ITS entered a new era with the fast development of communications, information technology, global poisoning system, etc. More and more advanced ITS have been carried out and are running in lots of cities. At present, we are facing good opportunities of the development for ITS because of various mature advanced technologies. However, due to the complexity, diversity and variability of real traffic conditions, developing a real 'intelligent' ITS is still a great challenge for traffic engineers and researchers.

Table 1.1 History of ITS development (Tokuyama, 1996)

	1st Stage	2nd Stage	3rd Stage
U.S.A.	ERGS ARCS UTCS	MOBILITU2000 IVHS	ITS
JAPAN	CACS	ARTS RACS VICS AMTICS SSVS	ITS
EUROPE	ALI	PROMETHEUS DRIVE I DRIVE II	PROMOTE TELEMATICS
Year	70 75 80	85 90 91 92 93 94 95	96

Note: ERGS-Electronic Route Guidance System, ARCS-Automatic Route Control System, UTCS-Urban Traffic Control System, CACS-Japanese Comprehensive Automobile Traffic Control System, ALI-Autofahrer Leit and Information System (ALI)

The core component of ITS is a Traffic Management Centre (TMC) which is “the hub of transport administration, where data is collected, and analyzed and combined with other operational and control concepts to manage the transportation network” (Vanajakshi, 2010). It is a center point communicating traffic information and the public, where traffic participants can regulate their activity decisions in real time according to the traffic conditions. Typically, several subsystems, such as Maintenance and Construction Management, transit management, Emergency Management and Information Service Provider, share the administration of transportation infrastructure and traffic information. Raw traffic data and road network information are input into center of traffic management, and are processed to obtain traffic conditions. The

derived traffic conditions are then transmitted to sub systems.

The effective performance of TMC and consequently ITS mainly depends on four factors: automated data acquisition, accurate analysis of data, fast data communication, reliable information to public/traveler. Among them, data acquisition is the foundation of all other factors, and critical to success for ITS.

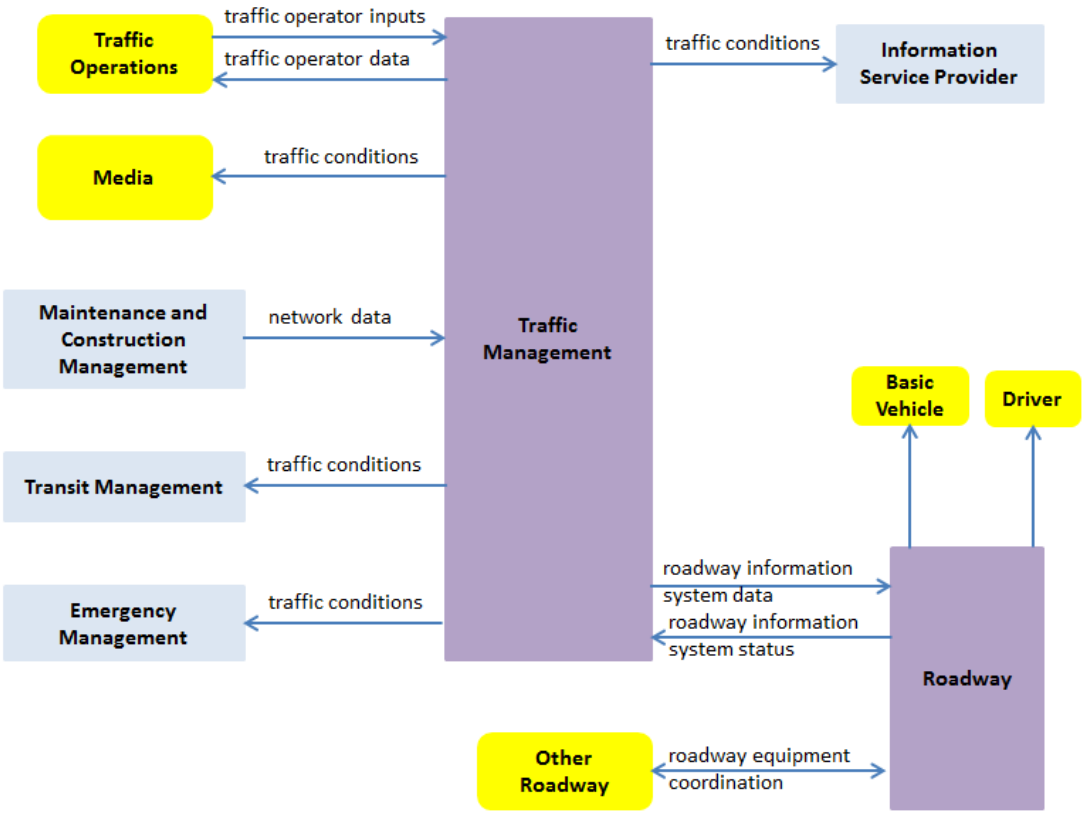


Figure 1.1 Schematic of a TMS (Vanajakshi, 2010)

1.1.2 Data acquisition methods

The basic mechanism of ITS is like a factory where raw materials are input and products are output. The quality of products highly depends on the materials. ITS could only be possible to have credible performance in controlling and managing traffic if high quality traffic data is input.

As shown in Figure 1.2, there are multiple methods of collecting traffic data including traditional manual counting, fixed sensors, and off-road technologies. Fixed

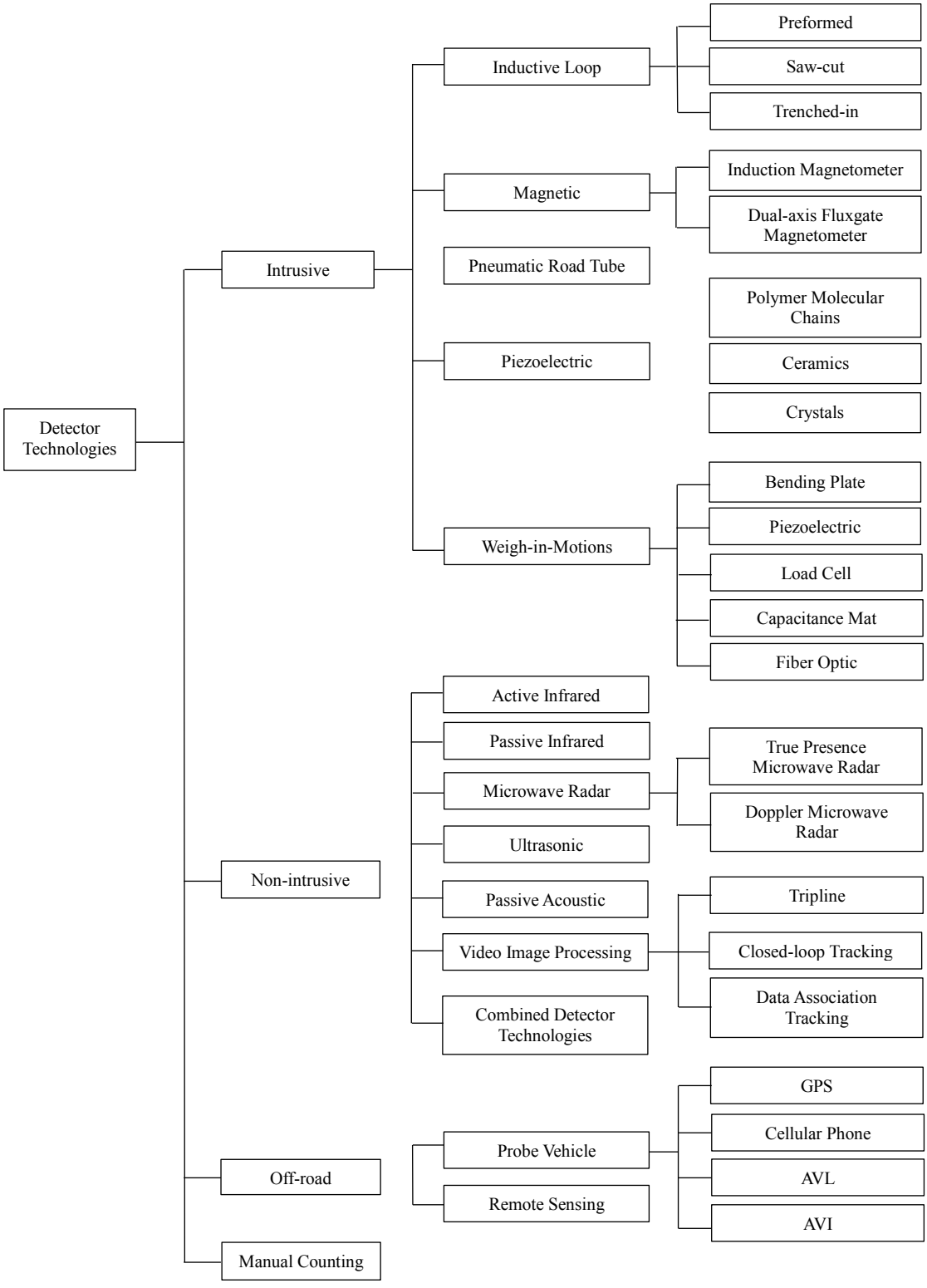


Figure 1.2 A family map of detector technologies of collecting traffic data (Martin, 2003)

sensors belong to on-road technologies and refer to detectors located along the roadside. Manual counting is the most traditional one. However, it is very inconvenient,

inefficiency and uneconomic. Fixed sensors can be divided into two classes: intrusive methods and non-intrusive methods. The intrusive methods collect data using a recorder receiving data from a sensor place on the road, while the non-intrusive methods refer to detectors based on remote observations. According to different technologies adopted, the intrusive and non-intrusive methods can be further divided into various types of detectors. Currently, off-road detectors include probe vehicle and remote sensing. Probe vehicle refers to a vehicle installed an in-vehicle devices which can record the vehicle information including time, location, speed etc. Technologies in probe vehicles include Global Position System (GPS), cellular phone, Automatic Vehicle Identification (AVI) and Automatic Vehicle Location (AVL) (Martin, 2003). In this thesis, probe vehicles refer GPS-based and cellular phone-based technologies for their wide applications. A family map of detector technologies is summarized in Figure 1.2. More details can be found in Martin (2003) and Leduc (2008).

Table 1.2 Data type provided by various detectors

Detector Type			**Volume/ Count**	**Speed**	**Travel time**	**Vehicle Type**	**Occupancy**	**Vehicle Presence**
Fixed sensors (intrusive)	Inductive loop		○	○	×	○	○	○
	Magnetic		○	○	×	○	○	○
	Pneumatic Road Tube		○	○	×	○	×	×
Fixed Sensors(non-intrusive)	Active Infrared		○	○	×	○	×	×
	Passive Infrared		○	○	×	○	○	○
	Microwave Radar	Doppler	○	○	×	○	○	U
		True Presence	○	○	×	○	○	○
	Ultrasonic		○	×	×	×	×	○
	Passive Acoustic		○	○	×	○	○	○
	Video Image Processing		○	○	×	○	○	○
Probe Vehicle	GPS		×	○	○	○	○	○
	Cellular Phone		×	○	○	○	○	○

Note: ○-can provide the type of data; ×-cannot provide the type of data; U-unknown

Since fixed sensors and probe vehicles are currently the main technologies for

collecting traffic data, we summarize various types of data provided by them by referencing (Martin, 2003; Leduc, 2008; Bennett, 2005; IMAGINE, 2006; Schmidt, 2005), as shown in Table 1.2.

Fixed sensors have been used to detector vehicle for a long time. Technologies in fixed sensors are mature, and researches on fixed sensors and analyses on fixed sensor data are rich. However, there are some unavoidable limitations: expensive fee of installation and maintenance, limited coverage, low travel time accuracy (derived from data), low precision for urban areas because of traffic interruptions.

In the research, we focus on the application of probe vehicle data, which is promising to be an alternative to existing data sources. Probe vehicles are becoming crucial in the development of ITS. The reason lies in that probe vehicle can cope with some limitations of fixed sensors and at the same time has its own characteristics. Compared with fixed sensors, probe vehicle technologies have many advantages:

- Large coverage area. Probe vehicles cover the entire network where a vehicle can travel on, and run in full time.
- Low cost of installation and maintenance. With no need to additional infrastructures, existing vehicles with only an in-vehicle device are used as probe vehicles, and these devices can be shared with other systems like scheduling system.
- Low cost per unit of data. Data can be collected easily at low cost after the necessary equipment is in place.
- Abundant and accurate data. Besides the information of time, location, speed, direction, more information such as acceleration, weather condition, fuel consumption can also be recorded without additional devices.
- No disruption of traffic. Existing vehicles are used as probe vehicles, which would not affect traffic conditions for the purpose of collecting traffic data.

- Strong anti-interference. Prove vehicle is seldom interfered by environmental conditions such as bad weather and dim light.

However, as an emerging and advanced technology, probe vehicle technologies are not mature and still in development. More details on probe vehicle technologies will be reviewed in Chapter 2.

1.1.3 Traffic conditions

Traffic condition is a reflection or snapshot of real traffic situations. There are lots of measures adopted by traffic engineering for describing traffic conditions: travel time, speed, volume, density, occupancy, level of service (LOS) (i.e. volume/capacity), delay, or queue length. Any one of them can only reflect one aspect instead of characterizing the full picture of traffic condition. All these measures actually belong to three categories by their definitions:

- **Flow rate** (volume, LOS, queue length). Whereas density is a spatial measurement, flow rate is a temporal measurement and is expressed in number of vehicles per hour.
- **Mean speed** (travel time, delay). Mean speed of a traffic stream is expressed in kilometers per hour. Travel time of a vehicle travelling a road is the product of the inverse of mean speed and length of this road.
- **Density** (occupancy). Density is typically expressed in number of vehicles per kilometer, and it measures how crowed a certain section of a road is.

Density k, flow rate q and mean speed v actually are three basic macroscopic variables in traffic flow theory. A unique relation called fundamental relation of traffic flow theory exists among them:

$$q = kv \tag{1.1}$$

In theory, this relation provides a close bond between these three variables: we can

calculate one from the other two. However, this is based on two assumptions: (1) continuous variables (Maerivoet, 2005); (2) traffic composition follow two assumptions: homogeneous traffic (i.e. the same type of vehicles) and stationary traffic (all the vehicle's trajectories should be parallel and equidistant (Maerivoet, 2005). In real traffic, these assumptions could not be completely reached. In reality, for one certain mean speed, there are multiple possible densities and flows. This is the same for certain flow and certain density. Additionally, without loss of generality, for example, certain mean speed and certain density does mean a unique flow in certain traffic condition. Considering the complexity of real traffic, it's better for a ITS to obtain all the three traffic flow measurements to comprehensively grasp the traffic conditions.

1.2 Problem statement

Along with the fast development of technologies of wireless communication, mobile internet, computer science, etc., types of probe vehicles become more abundant and their penetration is experiencing a fast growth around the world. Although estimating urban traffic conditions from probe vehicle data has been getting increasing attention, the existing researches have some limitations or local perspectives: the investigations are mainly using probe vehicle data as complement of traditional data, and only part of information (e.g. travel times) from probe vehicle data is utilized; as described in section 1.1.3, in existing literatures certain traffic quantities are used to measure traffic conditions instead of providing comprehensive measurements; the characteristics of probe vehicle data are not completely considered in the methodologies.

The reasons lie in that, on one hand, the urban traffic are mixed with non-recurrent and recurrent traffic, interrupted frequently by signal controllers or occasionally by pedestrians, and affected by frequent lane changing and commuting time. The

methodologies of estimating the urban traffic conditions have be able to deal with the extreme complexity of urban traffic. On the other hand, the aforementioned characteristics of probe vehicle data, including the diversity of data types, the lack of reliability and ubiquity, and the randomness of spatial-temporal coverage, make it challenging to be processed and applied in traffic estimating and monitoring. The proposed methods have to be able to properly handle probe vehicle data by carefully studying those features.

This study aims to develop methodologies for estimating comprehensive urban traffic conditions using probe vehicle data from the perspective of practical application for ITS. Particularly, in order to provide comprehensive measurements of traffic conditions, we conduct researches from three aspects as the three basic quantities in traffic flow theory. Considering the effects of polling frequency, we classify the information of probe vehicle data into types: travel time (the reliability and accuracy of travel time become lower with lower polling frequency) and distributed point with location and instantaneous speed (not affected by polling frequency). Therefore, the research issues we carry out are as follows:

1. **Flow**: link flow and origin-destination (O-D) flow estimation. This research aims to provide a methodology of estimating link flow (link-level traffic demand) and O-D flow (network- level traffic demand) from probe vehicle data. The methods of estimating sample link travel times from raw probe vehicle data would be discussed; the method of estimating link flow from link travel times would be investigated, and the application of prior information from achieved probe vehicle data would be analyzed; methods of estimating O-D flow would be studied and implemented; the effects of polling frequency on the accuracy of estimation results would be analyzed.
2. **Mean speed (travel time)**: link travel time distribution estimation. The

objective of this research is to model the probability distribution function (PDF) of link travel time, which is the foundation of measuring variability and reliability of link travel time. The composition of link travel time would be studied and modelled, and sample link travel times from probe vehicle data would be used for estimating the PDF of link travel time.

3. **Density:** traffic monitoring and network parameters estimation. The motivation of this research is mining traffic and network related information from the distributed probe points. The distribution characteristics of probe points would be investigated; the traffic densities on a signalized road section would be identified in time and space; the spatial-temporal relations between traffic on adjacent links of urban network would be studied; the parameters learning techniques would be investigated.

1.3 Organization of the thesis and contributions

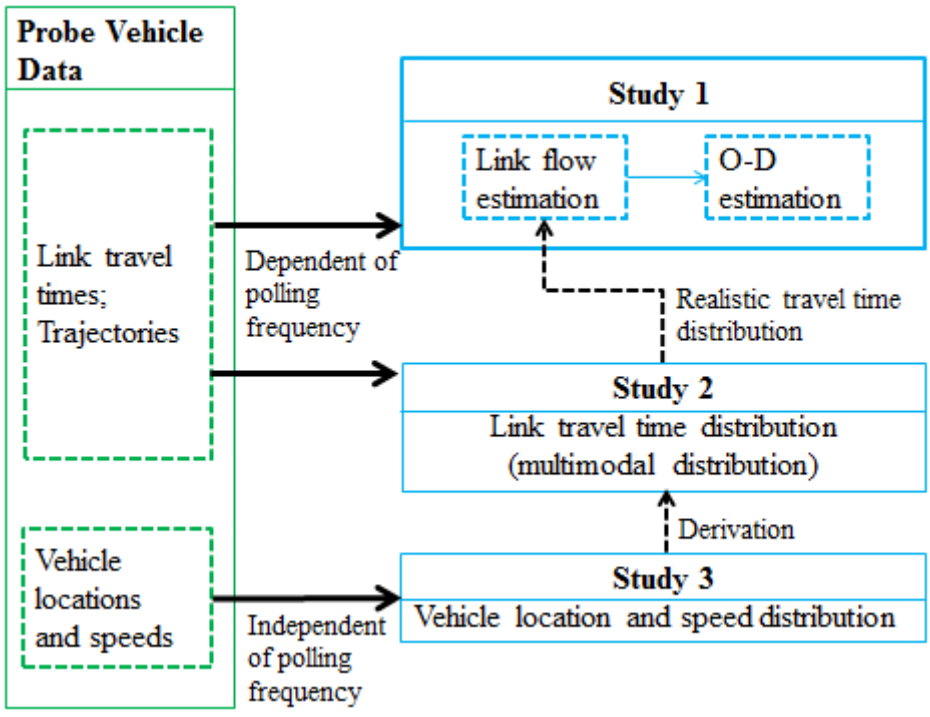


Figure 1.3 Flowchart of this research

As shown in Figure 1.3, firstly, we conduct study 1-link flow estimation and O-D

estimation. In this study, we assume that link travel time follows a normal distribution. Through this study, we find that the link travel time distribution plays an important role in estimation, and the effect of signal should be considered to obtain more accurate link travel time distribution.

So, the study 2-link travel time distribution is conducted. A multimodal distribution considering the effect of signal and truncation is derived. Using this model, it is expected to improve the link/O-D flow estimation in study1. However, results from both study 1 and 2 are dependent on polling frequency of prove vehicle data. If probe vehicle data are recorded with low polling frequency, the accuracy of results from study 1 and 2 will decrease.

So, we conduct study 3-estimationg vehicle location and speed distribution using information of locations and speeds from probe vehicle data, which is independent of polling frequency. Using this joint distribution, a better link travel time distribution may be derived in future for study 2 and then improve flow estimation in study 1. Or using Machine learning techniques, this distribution may be directly applied for improving link/O-D flow estimation in future for study 1.

The thesis consists of 7 chapters. The organization of this thesis and contributions are described hereafter.

In **Chapter 1**, the background of this research including intelligent transportation system, data acquisition methods and traffic conditions are introduced. This is followed by the problem statement and the organization of this thesis. **Chapter 2** gives a detailed literature review and assessment of the current state-of-the-art of probe vehicle technologies and traffic condition estimation.

Chapter 3 provides a methodology of estimating dynamic link flow from raw probe vehicle data. **Chapter 4** describes a bi-level generalized least square (GLS) model to estimate dynamic O-D flow from estimated link flow in **Chapter 3** and

historical O-D flow.

Publication: Cao, P., Miwa, T., Yamamoto, T. and Morikawa, T. (2013). Bilevel Generalized Least-Square Estimation of Dynamic Origin-Destination Matrix for Urban Network Using Probe Vehicle Data. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2333* (Presented at the 92nd annual meeting of the transportation research board, January 2013, Washington, D.C., USA.), Transportation Research board of the National Academies, Washington, D. C., 66–73.

Publication: Cao, P., Miwa, T., Yamamoto, T. and Morikawa, T. (2013). Estimation of Dynamic Link Flows and Origin-Destination Matrices from Lower Polling Frequency Probe Vehicle Data. *Journal of the Eastern Asia Society for Transportation Studies*, 10, 762–775.

Chapter 5 presents the formulation of link travel time distribution in signalized road section using truncated distribution.

Publication: Cao, P., Miwa, T. and Morikawa, T. (2014). Modeling Distribution of Travel Time in Signalized Road Section Using Truncated Distribution, *Procedia-Social and Behavioral Sciences,* Vol. 138, pp. 137-147.

Researches in **Chapter 3-5** utilize link travel times provided by part of probe vehicles, which has relative high polling frequency. Link travel times become less accurate and reliable for lower polling frequency probe vehicle data because of the difficulties of map-matching and path travel time allocation. Therefore, **Chapter 6** explores methodology of mining traffic information from scatted probe vehicle data regardless of polling frequency.

Chapter 6 describes the formulation of a joint probability distribution of vehicle location and speed on arterial road using hydrodynamic theory and horizontal queuing theory.

Publication: Cao, P., Miwa, T. and Morikawa, T. (2014). Use of Probe Vehicle

Data to Determine Joint Probability Distributions of Vehicle Location and Speed on an Arterial Road. In *Transportation Research Record: Journal of the Transportation Research Board*, Transportation Research board of the National Academies, Washington, D. C. (in press).

Chapter 7 describes conclusions and future researches.

References

A report to the ITS Standards Community ITS Standards Testing Program By Battelle Memorial Institute for US Department of Transportation (USDOT), Chapter 2.

Bennett, C.R., Chamorro, A., Chen C., de Solminihac, H., Flintsch, G.W. (2005). Data Collection Technologies for Road Management, Version 1.0, East Asia Pacific Transport Unit, The World Bank, Washington, D.C.

IMAGINE. (2006). Improved Methods for the Assessment of the Generic Impact Of Noise in the Environment. Collection Methods for Additional Data, Demand and Traffic Flow Management (WP2).

Leduc, G. (2008). Road traffic data: Collection methods and applications. Working Papers on Energy, Transport and Climate Change, 1, 55.

Maerivoet, S. and De Moor, B. (2005). Traffic flow theory. arXiv preprint physics/0507126.

Martin, P. T., Feng, Y. and Wang, X. (2003). Detector technology evaluation (No. MPC Report No. 03-154). Mountain-Plains Consortium.

Schmidt, M., Giorgi, L., Chevreuil, M., Paulin, S., Turvey, S. and Hartmann, M. (2005). GALILEO: Impacts on road transport, JRC-IPTS Technical Report EUR 21865.

Schrank, D., Eisele, B. and Lomax, T. (2012). TTI's 2012 urban mobility report.Texas A&M Transportation Institute. The Texas A&M University System.

Tokuyama, H. (1996). Intelligent Transportation Systems in Japan, US Department of Transportation-Public Roads, 60(2).

Vanajakshi, L., Ramadurai, G. and Anand, A. (2010). Centre of Excellence in Urban Transport IIT Madras.

Williams, B. (2008). Intelligent transport systems standards. Artech House.

Zhang, J., Wang, F. Y., Wang, K., Lin, W. H., Xu, X. and Chen, C. (2011). Data-driven intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 12(4), 1624–1639.

Chapter 2

Literature review

2.1 Introduction

Traffic engineers have witnessed the fast development of probe vehicle technologies in the past twenty years. Probe vehicle is promising to be an alternative for collecting vehicle data. We introduce the mechanisms of probe vehicle and review some crucial technologies in processing probe vehicle data in the chapter. Additionally, methods of estimating traffic conditions are reviewed from three aspects of flow, speed and density.

2.2 Probe vehicle technologies

2.2.1 Probe vehicle

Probe vehicle, also known as floating car, is a vehicle equipped a recording device receiving signals from GPS or antenna (GPS-based or cellular phone-based) and can provide series of points with information time stamp, speed and location etc. along the vehicle trajectories. There are multiple types of probe vehicles: taxis, delivery vehicles and buses with on-board device, vehicles using traditional cellular phone or smartphone as probe.

These types of probe vehicles can be categorized as: the GPS-based and cellular phone-based probe vehicle, depending on the principle of positioning. It should be noted that smartphone with integrated GPS is regarded as GPS-based.

For GPS-based probe vehicle (Figure 2.1), the vehicle position is determined using signals from at least four satellites firstly. This information will be corrected with the differential data that is calculated at the differential correction station. The location data together other information of the vehicle is packaged to be transmitted to

the control center. For cellular phone-based probe vehicle (Figure 2.2), the vehicle locations are determined by calls between the phone and two towers. This is done by a geolocating component of the system.

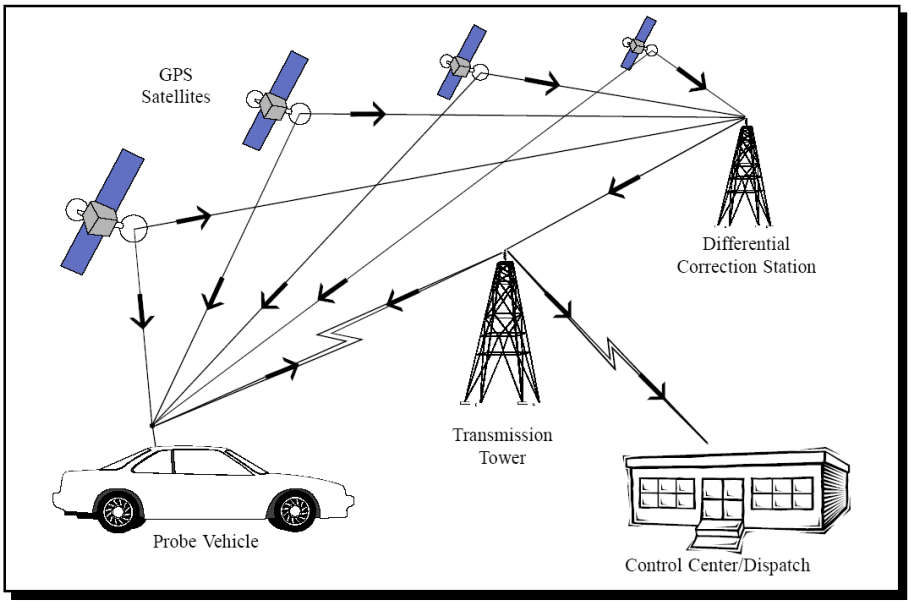


Figure 2.1 A typical GPS-based probe vehicle system (FHWA report, 1998)

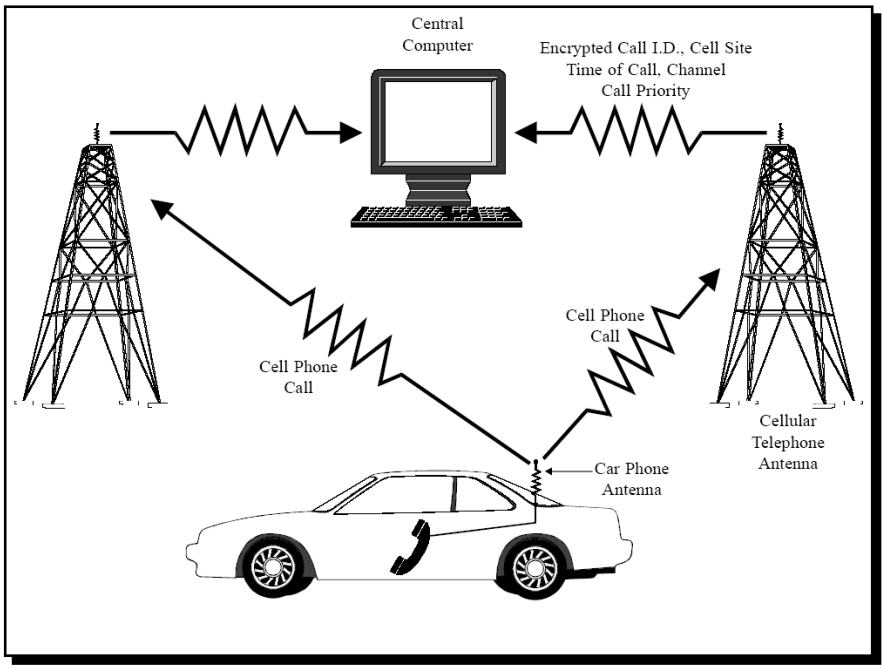


Figure 2.2 Cellular phone-based probe vehicle system (FHWA report, 1998)

2.2.2 Probe vehicle data

1) Data types

The original probe vehicles provide the basic information of time stamp, vehicle position and speed. Besides, more and more probe vehicles provide acceleration/deceleration, weather condition, ambient temperature and number of passengers. In recent years, some probe vehicle can also provide CAN (controller area network) data including fuel consumption, engine temperature etc. that reflect the status of a vehicle. These kinds of information are also called as extended probe vehicle data (xPVD) or extended floating car data (xFCD). Since no additional hardware is needed and only software is required, it is easy for probe vehicle to provide detail information about the vehicle or passengers. The aforementioned types of data can be directly recorded by on-board sensor in probe vehicle. By processing these data by some techniques, vehicle trajectory, O-D point, travel time, CO2 emissions etc. can be derived.

2) Polling schemes

Polling scheme is one crucial factor which essentially influences the data quality. The polling scheme determines the spatial distribution of probe points on the road network. Basically, there are two types of polling scheme: time-based and space-based. The former means probe vehicle reports a message every certain time interval. Probe points from this scheme are densely distributed on links close to signal controller or links of where the traffic is congestion. The later means probe vehicle sends a message every certain distance travelled by the vehicle. Probe points from this scheme are evenly distributed on links. Liu *et al*. (2007) compared the feasibility and efficiency of time-based and space-based polling schemes using probe vehicle data provided by Nagoya probe experiment. They concluded that, there are more missing data for space-based scheme than time-based due to communication congestion, and time-based data reflects various traffic conditions better than space-based data. Sometimes, another polling scheme, which is event-based, is also incorporated as a

complement. For example, cellular phone based probe vehicle send a signal when the phone is on call.

One crucial issue involved in polling scheme is polling frequency (alternative sampling rate). Polling frequency refers to the rate of sending a message per unit time. Polling frequency not only determines the data quality but also communication cost. Usually, polling interval, the inverse of polling frequency, is used in common. For the time-based polling scheme, the polling time interval can be pre-set to be a constant value. But for space-based and event-based scheme, the polling time interval is not a constant but random. In principle, high polling frequency means high quality data but with high payment for communications and fast battery consumption of sensors. Therefore, experimenters have to make a compromise between them. Except some particular experiments use high polling frequency (one data point every1-10s), most probe vehicles reports data every 30-60s or even longer, or unknown polling time interval because of randomly polling. And this trend will continue with the development of smartphone.

3) Penetration

Penetration is another crucial factor which influences the application of probe data. Penetration refers to the ratio of probe vehicle in total vehicle population. The higher penetration means more vehicles in the traffic provide their travelling information. On-board device-based probe vehicles such as taxis, FedEx, UPS or other delivery vehicles reach market penetration of 1% at most, and it will not increase a lot in near future because of the limited market demand for these vehicles. Cellular phone-based probe vehicles seem to have great potential to increase the total penetration of probe vehicles. Cellular phone reached extensive market penetration in many countries (Bar-Gera, 2007). For example, a survey conducted at the Phone corridor, Lyon, France reported that 77.4% of the vehicles had at least one cellular phone in 2000

(Ygance, 2001). In 2004, 83% of households owned one cellular phone and 53% owned two or more in Israel (Israeli Central Bureau of Statistics, 2006). In recent years, the extensive use of GPS-based smartphone around the world further increases the use of cellular phones as probes. Probe vehicles including all kinds of types would certainly rise to a high level of penetration.

4) Applications

The applications of probe vehicle data have been studied for more than decades. Along with the new types of data being collected by probe vehicles, probe vehicle data has proven the benefits in many issues and shown great potential in wider fields. Leduc (2008) summarized the areas of benefiting from the use of probe vehicle data.

- Congestion reduction
- O-D matrix estimation
- Travel time estimation/prediction
- Traffic queue detection
- Dynamic route guidance
- Incident management
- Dynamic network traffic control
- Vehicle fleet management (UPS, taxis, etc.)
- Fuel consumption and air emissions
- Digital map construction

These applications almost cover most key issues in transportation engineering. And the use of probe vehicle data is at different degrees to affect all the transportation actors: Government/public authorizes, location based service providers, logistics and fleet operators, consultants, map providers, and automobile manufactures and telecommunications.

2.2.3 Key technologies in processing probe vehicle data

1) Map-matching

Map-matching refer to the process of associating the reported points by probe vehicles onto the digital map of the road network. Map-matching arises from the reasons of unavoidable errors of positioning and making digital map. Usually, map-matching comprises two stages: path-inference and point matching. The former is to infer the links that the reported points are on, and the latter is to identify the locations on the inferred links that the reported points should be.

There exist lots of map-matching algorithms proposed by researchers categorized as probabilistic map-matching algorithms and advance map-matching algorithms using concept of a Kalama Filter, a fuzzy logic model and an interacting multiple model etc. The methods of map-matching are out of the scope of this thesis, details can be found in Quddus *et al*. (2007).

Besides the performance of the map-matching algorithms, the accuracy of digital map and the positioning of probe vehicles, polling frequency is another key factor that determines the accuracy of map-matching. The difficult of map-matching would dramatically increase and consequently the accuracy would reduce for probe vehicle data with lower polling frequency. The reason lies in that it is difficult to infer the real path since the vehicle may travel across several links between two consecutive probe points of lower polling frequency. The accuracy of map-matching would then affect the calculation of travel times. Therefore, polling frequency plays a very important role in map-matching and travel time calculation.

2) Travel time allocation

In principle, it's not necessary to allocate travel time from high polling frequency (e.g. one point every 1s) probe vehicle data. Whereas travel time allocation should be taken carefully for lower frequency data, since the two consecutive polled positions do not necessarily correspond to the end points of individual links. Hellinga *et al.* (2008)

proposed an analytical method of travel time allocation by recognizing that vehicles are more likely to incur stopping delay at the downstream rather than upstream end of a link, especially when the link is influenced by a traffic control device. Soon afterwards, Zheng and Van Zuylen (2012) presented a three-layer Artificial Neural Network (ANN) model and got higher accuracy estimates than Hellinga's model. However, both models are not capable for network application in which it's difficult to determine the parameters for them. In Hellinga's model, there are two unknown model parameters that are used to reflect the stopping likelihood pattern of a link. In Zheng's model, all parameters are learned from very high polling frequency data (one point every 0.3s), while this kind of data are not available in practice.

Intuitively, the error of the derived link travel time would increase when the polling interval becomes longer regardless of methods of travel time allocation. It is also validated by an empirical study (Liu *et al*., 2007).

2.3 Use of probe vehicle data to estimate traffic conditions

There are a large number of literatures dealing with the estimating traffic conditions from various data sources. The estimation method varies depending on the types of data. In this section, we only review the literatures close to our research topics: use of probe vehicle data to estimate traffic conditions. As we mentioned in problem statement in chapter 1, we concentrate on the discussion from three aspects of traffic conditions: flow, travel time and density.

2.3.1 Flow estimation

Generally, traffic flow includes link flow at the link-level and O-D flow at the network-level. The estimations of them are related with each other but the methods are different. We review these methods for them, respectively.

For link flow estimation, there are only a small amount of researches on the

estimation using probe vehicle data. It is because link flows usually are recorded by existing fixed sensors. However, for most links without fixed sensors on, the flow have to be estimated for better monitoring traffic conditions of these links. Probe vehicle data has been used to estimate link flows in a few studies. Yamamoto *et al.* (2009) proposed a Bayesian method to infer link flow from prior and current link speed distribution from link travel times provided by probe vehicles. However, the experiment results didn't validate the proposed method because of the limited data. Caceres *et al.* (2012) presented a set of models for inferring traffic volume by means of anonymous call data of cellular phone. In their models, the users' calling behavior and hourly intensity in calls and vehicles are contained. However, the inferred traffic volume actually is not link flow for a road segment but the number of vehicles from one cell to another.

More studies have been conducted to estimate O-D matrices from probe vehicle data. Van Aerde *et al*. (1993) estimated the level of penetration by calculating the ratio of probe vehicles in the population in aggregated time intervals for links, and then estimated dynamic O-D matrices using the origin and destination points of probe vehicles and the estimated penetration. Eisenman *et al.* (2004) proposed a method of estimating static O-D matrices based on traffic assignment, in which link choice ratios are inferred from probe route data. Ásmundsdóttir *et al.* (2010) discussed the rules of determining origins and destinations, route choice and trip length distribution within probe data from taxies, and then proposed a method of estimating dynamic O-D matrix from archived data and real time data. Nonetheless, the results didn't support the proposed method well. These methods attempt to develop a direct estimation process based on the assumption of random sampling. But in practice probe vehicles are from one or several types of vehicles (taxis, delivery vehicles, etc.) and consequently are not a random sample from the population. To avoid the assumption

of random sampling, Yamamoto *et al.* (2009) suggested a two-step indirect framework. In the first step, they inferred link flows from probe vehicle speed using two Bayesian methods; then they updated the target O-D matrix using an entropy maximization method. However, they didn't treat the variances of estimated link flows in a statistical manner and the results didn't validate the advantage of applying Bayesian methods in estimating the link flow.

2.3.2 Travel time estimation

Probe vehicle data has been widely applied in estimation of travel time, since sampled travel times can be directly provided by probe vehicles instead of other sensors.

Hellinga *et al.* (1999) derived analytical model based on queuing theory, and proved that bias in arrival time distributions and the distribution of probe vehicles in the total traffic would lead to a systematic bias in the estimation of the mean delay. In order to reduce the effect of this bias, Hellinga *et al.* (2002) proposed a methodology based on stratified sampling techniques. At the same time, Cheu *et al.* (2002) studied the probe vehicle sample size and population needed for arterial speed estimation. The results indicated that 4% to 5% of active probe vehicles are needed in network, or at least 10 probe vehicles must travel across a link in the estimation time interval, then the absolute error of estimated link speed can be less than 5 km/hour for more than 95% of the time.

In recent years, researchers developed multiple approaches which are model-based or data-driven. Zou *et al.* (2005) proposed a methodology of estimating arterial speed based on a sampling model using high frequency probe data provided by taxis. A mathematical model is proposed by Jula *et al.* (2008) to estimate travel times along the arcs and arrival times at the nodes of dynamic stochastic networks. The Kalman filter is applied to predict the future travel times along the arcs. Using the

"event-based" data including every vehicle actuations over loop detector, Liu (2009) presented a method to estimate arterial travel time by tracing a virtual probe vehicle. Jenelius *et al.* (2013) proposed a statistical model for estimating travel time for urban road network using low frequency probe vehicle data. By decomposing the trip travel times into link travel times and delays in intersection and using a spatial moving average (SMA) structure to express the correlation between travel times on different links, they presented a way to estimate the parameters including the correlation structure. The results revealed the potential of using probe vehicle data for monitoring the urban traffic.

Data-driven approaches are expected to solve the problem of travel time estimation by using large amount of data. Li and McDonald (2002) presented a pattern recognition method based on the analysis of speed-time profile. The driven pattern of a probe vehicle is classified by using fuzzy sets according to the features of the speed profile. Then the travel time is estimated by using corresponding methods for different driven pattern. Chan and Tam (2008) applied a method of nonparametric regression to estimate link travel times using an offline database consisting of average link travel times.

Besides the estimating the average value of travel time, researches are also conducted on the probability distribution of travel time. It is because travel time distribution determines the variability and reliability of travel time in the network performance.

The common way assumes that link travel time would follow a typical random distribution. May *et al.* (1989) assumed that travel time follow a normal distribution for a group of car commuters in north London. Rakha *et al.* (2010) also assumed a normal distribution for travel time on studying trip travel time reliability. Log-normal distribution was also applied in the analysis of road network reliability in Beijing in

Chen *et al.* (2007). Kaparias *et al.* (2008) proposed a new measure of travel time reliability based on log-normal distribution. Based on the assumption of log-normal distribution, Hollander (2008) estimated travel time distribution by repeated simulation. Additionally, Gamma distribution is assumed for travel time distribution in some researches. Polus (1979) studied route travel time distribution using a Gamma distribution. Guehthner (1985) used Gamma distribution to model the distribution of bus transit on-time performance.

As indicated by Lomax *et al.* (2003), a normal distribution is appropriate in the cases for simplifying the calculation process; when the characteristics of distributions like skewness cannot be ignored, the log-normal is more appropriate. However, since the urban traffic is complicated and varies along the time and space, a single random distribution may not be able to model the travel time satisfactorily. Therefore, some researches explored the application of composite distributions with a combination of more than one random distribution (May 1990, Feng *et al.* 2011).

In empirical experiments, the link travel times usually show multimodal distribution. It is because the traffic is interrupted by traffic signals. Feng *et al.* (2011) model arterial travel time distribution as a multi-component mixture of normal distributions. However, their model is actually data-driven and the effect of traffic signal wasn't explicitly considered. Hofleitner *et al.* (2012) formulate a analytical model parameterized by red signal time, cycle time etc., for arterial travel time distribution based on the hydrodynamic theory. However, delays due to traffic congestion and acceleration/deceleration are not considered in their research.

2.3.3 Density estimation

The researches on density estimation using probe vehicle are quit few, because the data provided by probe vehicles are quite different with density. However, since

collection of location- based data become abundant, density-based modelling using this kind of data would be meaningful. Hofleitner *et al.* (2010) formulated an analytical model for probability distribution of vehicle locations based on the observation: probe points are more likely distributed in dense traffic and on the road close to intersections. Together with the model of travel time distribution formulated in Hofleitner *et al.* (2012), they proposed a hybrid approach of flow modeling and machine learning for arterial travel time prediction. However, the travel times from low frequency data are less reliable and the more reliable information on speed are not applied in their researches.

References

Caceres, N., Romero, L. M., Benitez, F. G. and Del Castillo, J. M. (2012). Traffic flow estimation models using cellular phone data. *IEEE Transactions on Intelligent Transportation Systems*, 13(3), 1430–1441.

Chen, K., Yu, L., Guo, J. and Wen, H. (2007). Characteristics analysis of road network reliability in Beijing based on the data logs from taxis. In *86th Annul meeting of Transportation Research Board*, Washington, DC, USA.

Chan, K. S. and Tam, M. L. (2008). Using spatial travel time covariance relationships for real-time estimation of arterial travel times. In *87th Annul meeting of Transportation Research Board*, Washington, DC, USA.

Cheu, L. R., Xie, C. and Lee, D. H. (2002). Probe vehicle population and sample size for arterial speed estimation. Computer Aided Civil and Infrastructure Engineering, 17(1), 53-60.

Feng, Y., Davis, G. A. and Hourdos, J. (2011). Arterial travel time characterization and real-time traffic condition identification using GPS-equipped probe vehicles. In *90th Annul meeting of Transportation Research Board*, Washington, DC, USA.

FHWA report. (1998). Travel Time Data Collection Handbook, chapter 5: ITS Probe Vehicle Techniques.

Guehthner, R. P. and Hamat, K. (1985). Distribution of bus transit on-time performance. *Journal of Transportation Research Record No. 1202*, 1–8.

Hellinga, B. and Fu, L. (1999). Assessing expected accuracy of probe vehicle travel time reports. *Journal of Transportation engineering*, 125(6), 524–530.

Hellinga, B. R. and Fu, L. (2002). Reducing bias in probe-based arterial link travel time estimates. *Transportation Research Part C: Emerging Technologies*,10(4), 257–273.

Hellinga, B., Izadpanah, P., Takada, H. and Fu, L. (2008) Decomposing travel times measured by probe-based traffic monitoring systems to individual road segments. *Transportation Research Part C: Emerging Technologies*, 16(6), 768–782.

Hofleitner, A., Herring, R. and Bayen, A. (2012). Probability Distributions of Travel Times on Arterial Networks: Traffic Flow and Horizontal Queuing Theory Approach. In *91st Annul meeting of Transportation Research Board(No. 12-0798)*, Washington, DC, USA.

Herring, R., Hofleitner, A., Abbeel, P. and Bayen, A. (2010). Estimating arterial traffic conditions using sparse probe data. In *13th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 929–936.

Hollander, Y. and Liu, R. H. (2008). Estimation of the distribution of travel times by repeated simulation. *Transportation Research Part C: Emerging Technologies*, 16(2), 212–231.

Israeli Central Bureau of Statistics. (2006). Ownership of durable goods in deciles of households by net income per standard person. In: *Statistical Abstract of Israel*, No. 57.

Jenelius, E. and Koutsopoulos, H. N. (2013). Travel time estimation for urban road networks using low frequency probe vehicle data. *Transportation Research Part B: Methodological*, 53, 64–81.

Jula, H., Dessouky, M. and Ioannou, P. A., (2008). Real-time estimation of travel times along the arcs and arrival times at the nodes of dynamic stochastic networks. *IEEE Transactions on Intelligent Transportation Systems*, 9, 97–110.

Kaparias, I., Bell, M.G. H. and Belzner, H. (2008). A new measure of travel time reliability for in-vehicle navigation systems. *Journal of Intelligent Transportation Systems*, 12(4), 202–211.

Leduc, G. (2008). Road traffic data: Collection methods and applications. Working Papers on Energy, Transport and Climate Change, 1, 55.

Liu, H. (2009). A virtual vehicle probe model for time-dependent travel time estimation on signalized arterials. *Transportation Research Part C: Emerging Technologies*, 11–26.

Liu, K., Yamamoto, T. and Morikawa, T. (2007). Comparison of time/space polling schemes for a probe vehicle system. In *Proceedings of the 14th World Conference on Intelligent Transport Systems*.

Li, Y. and McDonald, M., (2002). Link travel time estimation using single GPS equipped probe vehicle. In: *The IEEE 5th International Conference on Intelligent Transportation Systems*, Singapore.

Lomax, T., Schrank, D., Turner, S. and Margiotta, R. (2003). Selecting travel reliability measures. Texas Transportation Institute, 474360-1.

Polus, A. (1979). A study of travel time and reliability on arterial routes. *Transportation*, 8, 141–151.

May, A. D. Traffic flow fundamentals. Prentice Hall, 1990.

May, A. D., Bonsall, P.W. and Marler, N.W. (1989). Travel time variability of a group of car commuters in north London. ITS Working Papers, University of Leeds, Great Britain, 277–279.

Quddus, M. A., Ochieng, W. Y., and Noland, R. B. (2007). Current map-matching algorithms for transport applications: State-of-the art and future research directions. *Transportation Research Part C: Emerging Technologies*, 15(5), 312–328.

Rakha, H., EL-Shawarby, I. and Arafeh, M. (2010). Trip travel-time reliability: issues and proposed solutions. *Journal of Intelligence Transportation Systems*, 14(4), 232–250.

Ygnace, J. (2001). Travel time/speed estimates on the French Rhone corridor network using cellular phones as probes. SERTI V Program, STRIP (System for Traffic Information and Positioning) Project, INRESTS-LESCOT 0201, Lyon, France.

Zheng, F. and Van Zuylen, H. (2012) Urban link travel time estimation based on sparse probe vehicle data. *Transportation Research Part C: Emerging Technologies*, 31, 145–157.

Zou, L., Xu, J. M. and Zhu, L. X. (2005) Arterial speed studies with taxi equipped with global positioning receivers as probe vehicle. *Proceedings of the 2005 International Conference on Wireless Communications*, Networking and Mobile Computing, 1343–1347.

Chapter 3

Estimation of link flow using link travel times by probe vehicles

3.1 Introduction

Previous method of estimating link flows using probe vehicle data not only utilized the accurate link travel times from high polling frequency probe vehicle points, but also ignored the effect of polling frequency on the derived link travel times (Yamamoto *et al.* 2009). However, it is less likely a probe vehicle directly records link travel times, since the two consecutive polled positions do not necessarily correspond to the end points of individual links. Methods have been proposed to decompose travel times measured by probe vehicles into individual road segments (Hellinga, 2008; Zheng and Van Zuylen, 2012). These researches indicate the link travel time become less reliable when it is derived from lower polling frequency probe vehicle points. This chapter analyzes the effects of polling frequency and method of decomposing travel time on the derived travel time, and then explores method of estimating dynamic link flows from probe vehicle data.

This chapter first discusses the issue of decomposing probe vehicle travel times into individual links, and then proposes a method of estimating dynamic link flows. Finally, in the experiment, the effect of polling frequency and penetration will be analyzed and discussed.

3.2 Methodology

The raw probe vehicle data are series of track points including location, time. Usually, it requires procedures of map-matching and travel time allocation before link travel time being obtained. This research starts from the map-matched probe data, discusses travel time allocation that is directly related with polling frequency. To build a bridge

between link travel time and link flow, we derive link performance function based on the density-versus-space mean speed (*k*-*v*) curve. Therefore, the methodology includes three steps: travel time allocation, link performance function fitting and dynamic link flows estimation.

3.2.1 Step1-travel time allocation

To our knowledge, most existing probe vehicle systems still use a simple but practical method to obtain link travel time, in which uniform motion is assumed (Miwa *et al*. 2004). From the perspective of practical applications, we also use this method described as follows.

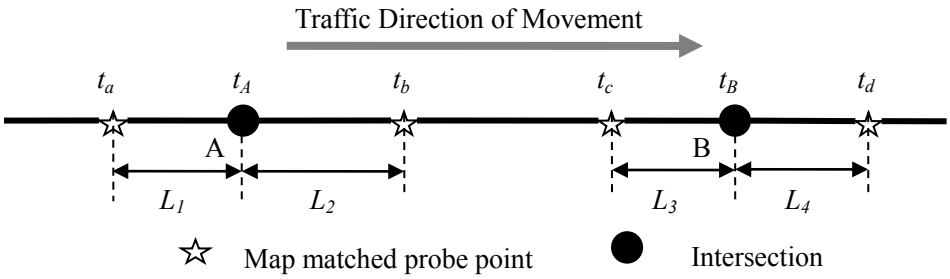


Figure 3.1 Calculation of link travel time using probe vehicle data

In Figure 3.1, we assume raw probe data have been map-matched onto road, A and B are intersections which are controlled by traffic signals. t_a, t_b, t_c, and t_d are the time recorded by the probe vehicle. t_A and t_B are the time a probe vehicle departs intersection A and B respectively, which need to be estimated.

Assume a probe vehicle travels in uniform motion between t_a and t_b, and between t_c and t_d, then t_A and t_B can be estimated by:

$$\widehat{t_A} = \frac{t_a * L_2 + t_b * L_1}{L_1 + L_2} \tag{3.1}$$

$$\widehat{t_B} = \frac{t_c * L_4 + t_d * L_3}{L_3 + L_4} \tag{3.2}$$

Then the travel time of link AB is calculated by:

$$\widehat{T_{AB}} = \widehat{t_B} - \widehat{t_A} \tag{3.3}$$

This method actually allocates travel time according to the distance between probe point and intersection point, thus we call it proportional allocation in this thesis. Proportional allocation can be easily applied to various kinds of probe vehicle systems in spite of probably resulting inaccurate link travel times. Moreover, it has a special function of reducing the variance of the derived link travel time. It can be illustrated with an example using Figure 3.1 in the following.

Note that the following calculation is just an illustration example. Without loss of generality, we assume the time of a vehicle entering in link AB t_A is known, the time of this vehicle leaving the link should be calculated using Equation (3.2). Intersection B is controlled by a traffic signal with 30s red light of 60s cycle. A probe vehicle enters link AB at $t_a = 100\text{s}$, and runs at a constant speed of 10m/s or stops at the intersection B waiting for the red light. And the polling frequency is one point every 60s. The length of link AB is 300m, values of L_3 and L_4 depend on the time when probe vehicle sends data. Since the red light is the main factor causing a vehicle to stop, we consider two extreme situations: no stop, stop when red light begins.

Situation 1: No stop

In this situation, a vehicle always maintains a speed of 10m/s, and experiences the minimum delay. A probe vehicle sends one point on $t_c = 115\text{s}$ and $t_d = 175\text{s}$, then we can calculate $L_3 = 150\text{m}$, $L_4 = 450\text{m}$, $t_B = 130\text{s}$, thus $T_{AB} = t_B - t_A = 30\text{s}$. If only t_a and t_d are assumed known, t_B is estimated using Equation (2), then $\widehat{t_B} = 130s$, $\widehat{T_{AB}} = 30s$. We can see the estimated link travel time $\widehat{T_{AB}}$ equals the true link travel time T_{AB} in this case

Situation 2: stop when red light begins

A probe vehicle travels at a speed of 10m/s, and then stops at intersection B waiting for 30s of red light, finally leaves link AB at the speed of 10m/s. In this situation, the vehicle experiences the maximum delay. The probe vehicle reports

$t_c = 115\text{s}$ and $t_d = 175\text{s}$. We can calculate $L_3 = 150\text{m}$, $L_4 = 150\text{m}$, $t_B = 160\text{s}$, thus $T_{AB} = t_B - t_A = 60$ s. Using Equation (3.2-3.3), we can estimate t_B as $\widehat{t_B} = 145s$, T_{AB} as $\widehat{T_{AB}} = 45s$ Thus, in this case the estimated link travel time $\widehat{T_{AB}}$ is smaller than the true link travel time T_{AB}.

From the above calculation, we know the true link travel time T_{AB} is distributed in range of $[30s, 60s]$, whereas the estimated link travel time using proportional allocation $\widehat{T_{AB}}$ is distributed in a smaller range of $[30\text{s}, 45\text{s}]$. Therefore, the variance of estimated link travel time using proportional allocation is smaller than that of true link travel time.

3.2.2 Step2-link performance function fitting

Link performance function builds a bridge between link travel time and link flow. The average and variance of link flow can be estimated from link travel time if the performance function is given. In traffic theory, there is a basic relation among three macroscopic variables which is formulated for uninterrupted flow. For urban traffic, since we only consider the macroscopic characteristics, the basic relation is applied:

$$q = k\bar{v} \tag{3.4}$$

where q is link flow, k is vehicle density, $\bar{v}$ is space mean speed.

The density k cannot be directly obtained from probe vehicle, but can be estimated from space mean speed using a *k*-*v* function. From Gazis's nonlinear follow-the-leader model (Gazis *et al*. 1961), we derive a *k*-*v* function:

$$\bar{v} = v_f exp(-\alpha \left(\frac{k}{k_j}\right)^{\beta}) \tag{3.5}$$

where k_j denotes the jam density, v_f denotes the free flow speed and α and β are parameters.

This formula is essentially a generalized version of the Underwood model (Underwood, 1961). As indicated by Wu *et al.* (2012), the fundamental diagram is

significantly affected by signal timings in the case of an urban road, and thus the speed-density relationship is different from that for a highway. It should be noted that the derived function is a little different from the one given in Yamamoto *et al.* (2009), where link capacity C replaces jam density k_j, and α and β have different meanings. The above formula is applied in this study, because k_j can be easily set for a particular link as the length of the link divided by the average vehicle spacing.

In a probe vehicle system, link travel times and thus link speeds can be easily obtained in real-time in aggregated time intervals. However, the observed link travel speed of probe vehicle is not necessarily identical to the mean speed, since the speed of a probe vehicle depends on the arrival flow rate and distribution, traffic signal timings and arrival time, which are random variables. So the link speed of probe vehicle i in time interval t is regarded as a random variable distributed around the mean speed $\bar{v}_t$:

$$v_t^i = \bar{v}_t + \varepsilon_t^i \tag{3.6}$$

where ε_t^i is an error term.

Assume that ε_t^i follows a normal distribution $N(0, s_t)$, then the probability density function of v_t^i can be given as

$$f\left(v_t^i \middle| \bar{v}_t, {s_t}^2\right) = \frac{1}{\sqrt{2\pi} s_t} exp\left\{-\frac{\left(v_t^i - \bar{v}_t\right)^2}{2{s_t}^2}\right\} \tag{3.7}$$

A Lagrangian likelihood function can be constructed for Equation (3.7) into which Equation (3.5) is substituted, then parameters v_f, α, β and s_t can be estimated using the maximum likelihood method. The reader is recommended to refer to Yamamoto *et al.* (2009) for more details of this procedure.

3.2.3 Step3-dynamic link flows estimation

To utilize prior information that can be obtained from archived probe data, a method based on Bayes' inference theory, namely the Bayesian method (BM), is used in this study. For each link at time interval t, the posterior mean and variance of the mean

speed v_1, ${\sigma_1}^2$ are given as

$$v_1 = \frac{{\sigma_0}^{-2} \cdot v_0 + n \cdot s^{-2} \cdot \hat{v}}{{\sigma_0}^{-2} + n \cdot s^{-2}} \tag{3.8}$$

$${\sigma_1}^{-2} = {\sigma_0}^{-2} + n \cdot s^{-2} \tag{3.9}$$

where v_0, ${\sigma_0}^2$ are the prior mean and variance of the mean speed, respectively, and $\hat{v}$ is the average link speed of probe vehicles.

Then using the posterior distribution of link speed, the link performance function and the relationship among link flow, traffic density and link speed as in Equation (4), the posterior mean and variance of link flow, $\bar{q}$ and ${\sigma_q}^2$, are given as

$$\bar{q} = \int_{v=0}^{\infty} K(v) \cdot v \cdot f(v|v_1, \sigma_1) dv \tag{3.10}$$

$${\sigma_q}^2 = \int_{v=0}^{\infty} (K(v) \cdot v - \bar{q}) \cdot f(v|v_1, \sigma_1) dv \tag{3.11}$$

where $K(v)$ is the function of density k with respect to speed v solved from Equation (3.5).

An ordinary method (OM) of link flow estimation is also employed to compare with the above Bayesian method. This is represented as

$$\bar{q} = \frac{1}{n} \sum_{i=1}^{n} v_i K(v_i) \tag{3.12}$$

$${\sigma_q}^2 = \frac{1}{n} \sum_{i=1}^{n} (v_i K(v_i) - \bar{q})^2 \tag{3.13}$$

where v_i is the link speed of probe vehicle i and n is the number of probe vehicles observed in a certain time interval.

Note that we obtain both the link flow and the variance using both BM and OM. The variances virtually reflect the difference in the reliability of link flow estimates among links. In practice, a huge number of data can be obtained after a probe vehicle system has been running for several months. This make it possible get enough data for particular time of particular day on particular road, thus provide enough prior information for link speed distribution.

3.3 Study Network

The network studied in this research is a western part of central Tokyo, namely Kichijoji-Mitaka area, which extends about 2 km from east to west and 1 km from north to south (Figure 3.2). This network consists of 138 links and 57 nodes, including four major north-south streets and two major east-west streets. Horiguchi *et al.* (1998) carried out a precise traffic survey on this network in morning peak period 7:00 am-10:00 am on 30 Oct. 1996, and made an open data set. Link volume on 70 links were observed and totally 16,043 vehicle trajectories are identified and, after data cleaning, link flows and O-D demands for each 10-minute period for the effective time interval 7:50 am to 10:00am are derived. There are 26 origins and 26 destinations identified in this network. In addition to these data, geometry of most intersections and all signal timings can be also found in the Kichijoji-Mitaka open data set. We use this network not only because of the rather complete data set, but also because there are multiple routes for many O-D pairs.

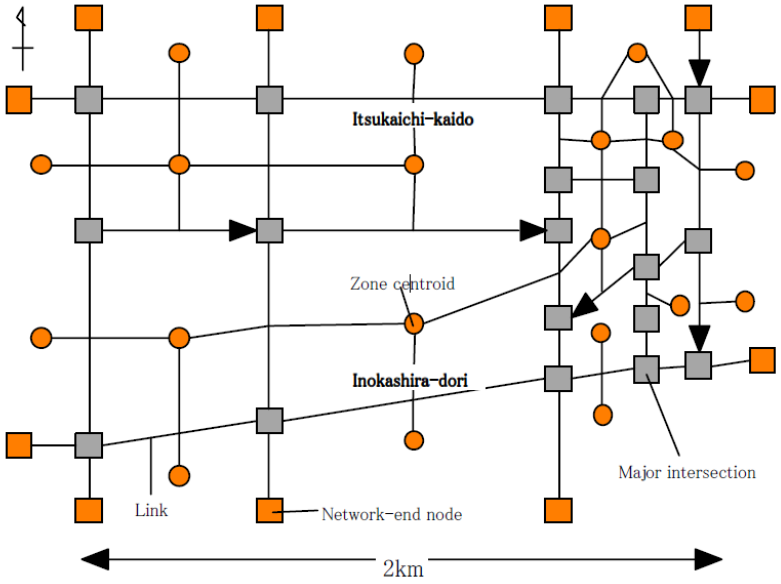
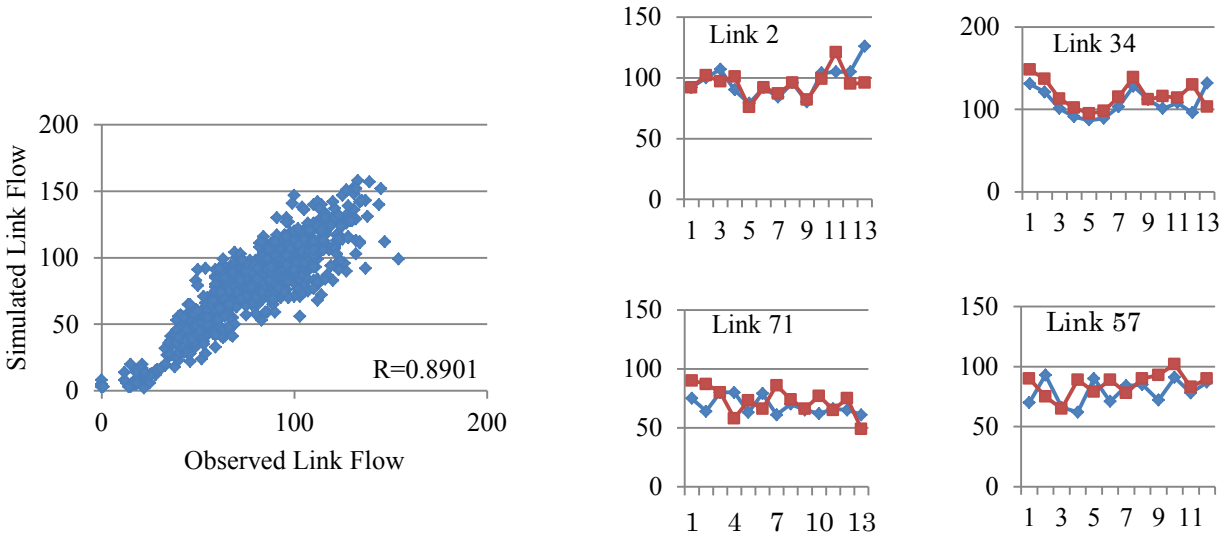


Figure 3.2 The Kichijoji-Mitaka network (Horiguchi *et al.*, 1998)

1.2. Traffic Simulator

In order to reproduce the real traffic conditions, we develop a microscopic traffic

simulator based on VISSIM using Kichijoji-Mitaka network and its data set. An equal scale road network is drawn on VISSIM. Network geometry including link lengths, number of lanes and link connectivity etc. is set the same as in reality. Parameters for the simulator, including signal timings and traffic controls are also set using the real data. The observed dynamic O-D matrix is mapped onto the developed network to validate the simulator.



(a), Scatter plot of observed and simulated link volume for each 10-minute period

(b), Link flows comparison for all time intervals. x-axis is the aggregated time interval, and y-axis is link flows per 10minuts. Blue line denotes observed flow, and red line denotes simulated flow

Figure 3.3 Validation of the developed simulator

The comparison between the developed simulator and the real traffic network is shown in Figure 3.3. Figure 3.3a is the scatter plot of observed link flows given by the Kichijoji Benchmark Data Set and simulated link flows obtained from the simulator for each 10-minute period. The linear correlation coefficient between them is 0.8901, the slope of the linear fitted line is 0.9793, and the root mean square error (RMSE) is 13.56 vehs/10mins. We also select four links and observe dynamic link flows for various time intervals (Figure 3.3b). In Figure 3.3b, the red plots represent observed link flows while

the blue plots represent simulated link flows. As we can see, the simulated link flows fits the observed values well. These results suggest that the simulator is able to reproduce real traffic conditions with high accuracy, although a better simulator would be obtained if more parameters of the true network, such as lane widths, stop line locations and road gradients were available.

3.4 Numerical experiment and results

3.4.1 The assumptions for simulation

Utilizing the developed VISSIM simulator, we can simulate the traffic within certain ratio of probe vehicles. Before implementing this simulator, we make assumptions of system as follows:

- Single class vehicles. All vehicles are assumed vehicles with standard vehicle properties like length and width. Multi-class vehicles in the original network have been converted using vehicle conversion factors (Horiguchi *et al.*, 1998).
- Driving behavior. We use the psycho-physical model of Wiedemann 74 (1974), which classifies four driving states: free driving, approaching, following and braking. Wiedemann 74 is applicable for inner urban road traffic. We assume free lane changing for vehicles which would change lane for larger traveling space or higher traveling speed. This means overtaking is permitted at any lane as long as conditions are satisfied.
- Dynamic traffic assignment. All vehicles are not equipped route guidance devices, thus route choice decision is only based on cost in previous iteration. We assume that driver chooses not only the optimal route but also a series of feasible routes.

The simulation time is the morning peak period from 7:50 am to 10:00 am, the length of aggregate time interval is assumed 10 minutes, and thus there are 13 time interval during the simulation time.

3.4.2 The estimation results for BM and OM

In principle every link has its own performance function, which is affected by factors like number of lanes, width and length of link and signal timings at intersections. However, in practice there is usually not enough data for determining performance function for each link. Considering the practical application, we choose number of lanes as the only criterion that distinguishes one link from another. Thus, two types of link are identified for the study network according to the number of lanes: roads with one lane and roads with two lanes in one direction. Then the link performance functions and *k*-*v* functions for them are obtained using method described section 3.2.2. The calibrated link performance functions and *k*-*v* functions are then used to estimate dynamic link flows using Bayesian Method (BM) and Ordinary Method (OM) in section 3.2.3. In the Bayesian method, prior distributions of vehicle speed are aggregated over each 10 minutes. Since there is some probability that no probe vehicle passes along some links in certain time intervals, especially when the probe ratio is low, the proportion of records being estimated (PRBE) is counted to indicate the different estimation abilities of BM and OM. Additionally, if no probe vehicle is passes in some time interval, the OM estimated link flow is set at 0 for that link. To examine the usage of prior information in calculating estimation results, three prior distributions of vehicle speed are assumed in BM: A (aggregated over each 10 minutes); B (aggregated over each 1 hour); and C (aggregated over 2 hours). A, B and C represent three levels of accuracy of prior information, with A the highest accuracy, B the second highest and C the lowest.

Table 3.1 shows the estimation results of BM and OM for various probe ratios. Two kinds of RMSE are calculated: RMSE for all links and RMSE for travelled links. The former is the RMSE calculated for all links in all time intervals, while the latter is the RMSE computed for the links in time intervals when at least one probe vehicle

travels.

Table 3.1 Link flow estimation results for BM and OM

Probe ratio			0.01	0.03	0.05	0.07	0.1	0.2	0.3	0.4	0.5
PRBE (%)	BM		100	100	100	100	100	100	100	100	100
	OM		52.64	86.37	93.30	95.82	97.69	99.23	99.45	99.78	99.89
RMSE for all links (vehs)	BM	A	53.20	34.41	30.74	29.74	29.65	32.22	33.66	34.57	35.10
		B	55.55	35.59	31.88	31.13	32.62	32.13	33.82	34.79	36.13
		C	55.96	34.63	31.11	32.88	35.54	32.85	34.31	34.99	35.35
	OM		58.44	35.69	31.34	30.10	31.02	32.36	32.22	32.47	32.46
RMSE for travelled links (vehs)	BM-A		19.72	27.82	28.36	28.67	29.78	32.29	33.73	34.60	35.11
	OM		21.54	28.32	28.77	29.06	31.07	32.46	32.30	32.50	32.47

The values of PRBE demonstrate, as expected, that the BM gives an estimation result even if no probe vehicle uses a certain link in a certain time interval as long as the prior information is given, whereas the OM can only estimate those links passed by at least one probe vehicle.

From the RMSE for all links, the BM results indicate that accuracy is best with the highest level of prior information, A, especially for low probe ratio data. As the probe ratio increases, the RMSE of BM decreases to 29.65 vehs when the probe ratio equals 0.1, and then begins to increase. There is no significant benefit as the probe ratio rises. The reason for this may be, as Equation (3.8-3.9) implies, that the current information provided by probe vehicles dominates the posterior distribution of probe vehicles when more probe vehicles are used. A comparison of BM and OM results shows that, when the probe ratio is lower than 0.2, BM with prior information A (BM-A) gives a better estimation than OM, but a worse one when the probe ratio is higher than 0.2.

The RMSE values for travelled links that are also displayed in Table 3.1 enable BM and OM to be compared. In this calculation, the same links set for both BM and OM is considered for each probe ratio case. The results show the RMSE of BM-A is smaller than that of OM when the probe ratio is less than 0.2 but bigger at higher probe ratios. This is consistent with the previous observation and arises because OM has the

advantage of making use of all information from the observed probe vehicles, whereas BM can only take advantage of part of the current information.

3.4.3 The effect of polling frequency

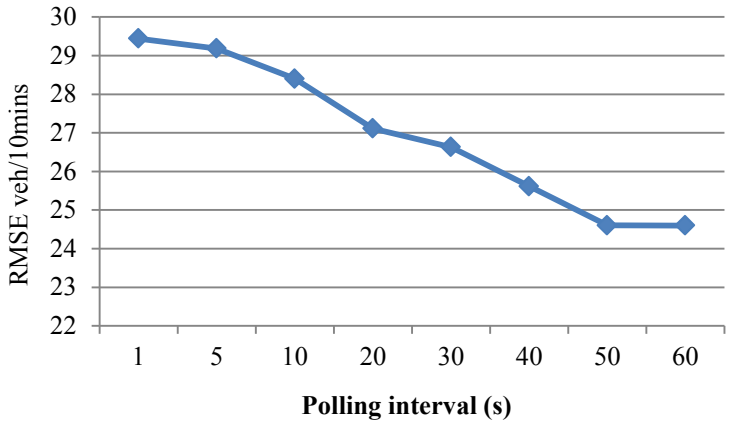


Figure 3.4 RMSE of estimated link flows

The probe ratio is set as 0.1 in the experiment. Figure 3.4 shows the RMSE of estimated link flows using various polling frequencies probe data. As we can see, the polling frequency does affect the accuracy of link flow estimation. Along with the polling interval becoming longer (i.e. the polling frequency becoming lower), the RMSE of estimated link flows decreases before polling interval is 50s, later begins to slow down. Further, we obtain acceptable estimates of link flows with RMSE 24.6 veh/10 minutes when polling interval is longer than 50s. In other words, using the proposed Bayesian method, the longer polling interval would produce higher accuracy estimated link flows.

In order to get insight of estimation results, we plot scatter figures of true link flow and estimated link flow for four different polling intervals: 1s, 10s, 30s and 60s (Figure 3.5). In Figure 3.5, we also demonstrate the distribution of estimate with different σ_1 value ranges (the variance of posterior link speed distribution calculated by Equation 8b) denoted by different tags. We can observe from these figures that the scatter plots become more concentrated on the diagonal line for each σ_1 value range with the

increase of polling interval, and the correlation coefficient R becomes larger. Therefore, we can obtain the conclusion that the estimate link flows are more accuracy for longer polling intervals, which consists with that from Figure 3.4. Additionally, from all plots in Figure 3.4 we can see the trend that the σ_1 increase with the decrease of link flow and this trend is clearer for longer polling interval. The reason can be found in Equation (3.9). There are more probe vehicles observed in certain time interval for links with higher volume of flow. The increase of number of probe vehicles n would bring the decrease of σ_1 in Equation (3.9).

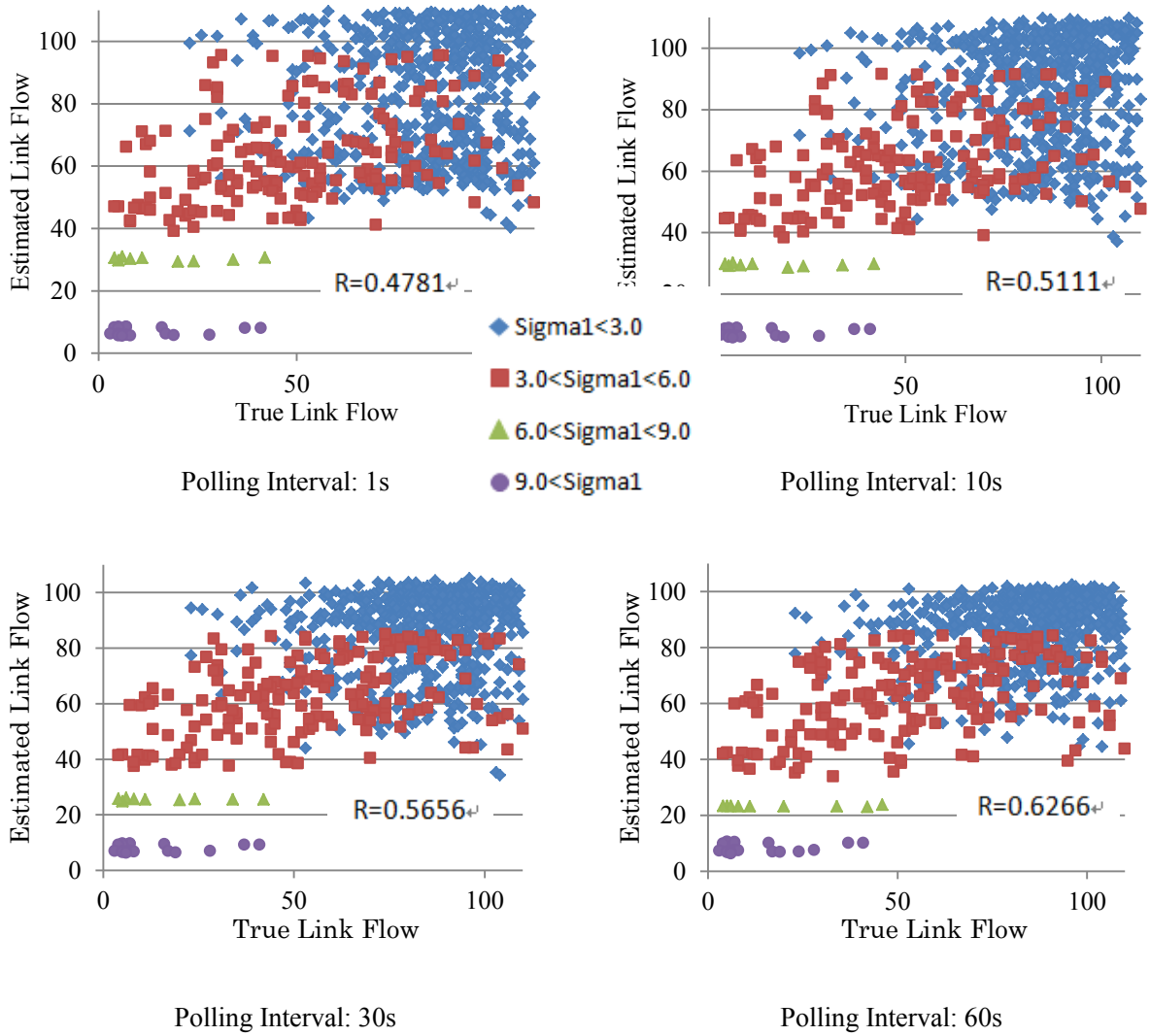


Figure 3.5 Scatter plots of true and estimated link flows for various polling intervals

These results show that the link flow estimates are more accurate for longer polling intervals than shorter polling intervals. The reason lies in the proportional

allocation in dealing with signalized intersection. Short polling interval probe data might incorrectly reflect the traffic condition in some specific cases. For example, even in the free flow situation, a probe vehicle, which arrives at the intersection when the red signal begins, will report a long link travel time (free travel time + red signal time, thus low link speed). The reported link travel time from short polling interval reflects actual travel time of the vehicle but it will lead misunderstanding of the traffic situation in some cases. Suppose that penetration of probe vehicle is 0.1 and there are 100 vehicles passing an intersection in one cycle. Even when 50 vehicles experience stop and 50 vehicles don't, it's possible that 8 probe vehicles experience stop and 2 probe vehicles don't stop. In this situation, the travel times from shorter interval data will lead biased traffic condition. In the long polling interval case, on the other hand, although the observed travel times will be relatively inaccurate, they are evened link by link and vehicle by vehicle and they can reflect actual traffic condition better than that in short polling interval case. As illustrated in section 3.2.1, the variance of derived link travel times from longer polling interval data using proportional allocation is smaller than that from shorter polling interval data regardless of free flow or congested flow. From these analyses, we can find that our conclusion is more suitable for the lower penetration of probe vehicle, since the randomness whether a vehicle experiences a stop is larger for small number of probe vehicles. Therefore, the proposed method is applicable for practice application, because penetration of probe vehicle is usually low in reality. Although our method performs better for longer polling interval (lower polling frequency), it doesn't mean that very low polling frequency data (e.g. one point every 2-5 minutes) can still give better results. In reality, polled points become sparser and map-matching becomes much more difficult and for lower frequency probe vehicle data, which results less reliable information. Based on this consideration, it’s better for the proposed Bayesian method to be applied in the situation of lower polling frequency

probe vehicle data (e.g. one point every 30-60s).

3.5 Summary

In this chapter, method of estimating dynamic link flows using probe vehicle data is explored. The probe data is assumed the map-matched points on the road network. We proposed a three-step methodology including: travel time allocation, link performance function fitting and dynamic link flow estimation.

In the first step, method of proportional allocation is used to decompose probe travel time onto individual links. This method can be easily carried out for network application, and it has a function of reducing the variance of the derived link travel time. In the second step, link performance function is obtained from a derived speed-density function. The speed–density function is derived from Gazis's nonlinear follow-the-leader model. The key feature of this function is that it only has four parameters, which are estimated using maximum likelihood method. We only classify links into two types by their number of lanes in the experiment from a view of practical application. If link performance function is known for each link, the estimate would be better. In order to estimate link flow, a Bayesian method that incorporates prior distribution of link speed and an ordinary method are applied in third step. It has been shown that the Bayesian method can effectively use the prior distribution of vehicle speed accumulated from archived probe vehicle data, and produce an acceptable link flow estimate even if there is no probe vehicle observed in that link. In addition to the average value of link flow, Both the Bayesian method and ordinary method estimate the variance of estimated link flow at the same time (see Equation 3.11 and 3.13).

References

Gazis, D. C., Herman, R. and Rothery, R. W. (1961). Nonlinear Follow-the-Leader Models of Traffic Flow. *Operation Research*, 9(4), 545–567.

Hellinga, B., Izadpanah, P., Takada, H. and Fu, L. (2008). Decomposing travel times measured by probe-based traffic monitoring systems to individual road segments. *Transportation Research Part C: Emerging Technologies*, 16(6), 768–782.

Horiguchi, R., Yoshii, T., Akahane, H., Kuwahara, M., Kitakura, M., Osaki, H. and Oguchi, T. (1998). A Benchmark Data Set for Validity Evaluation of Road Network Simulation Models. *Proceedings of the 5th World Congress of Intelligent Transport Systems*.

Miwa, T., Sakai, T. and Morikawa, T. (2004). Route Identification and Travel Time Prediction Using Probe-Car Data. *International Journal of ITS Research*, 2(1).

Underwood, R. T. (1961). Speed, Volume, and Density Relationship: Quality and Theory of Traffic Flow. Yale Bureau of Highway Traffic 141–188 (New Haven, Connecticut).

Wu, X., Liu, H. X. and Geroliminis, N. (2011). An Empirical Analysis on the Arterial Fundamental Diagram. *Transportation Research Part B: Methodological,* 45(1), 255–266.

Yamamoto, T., Miwa, T., Takeshita, T. and Morikawa, T. (2009). Updating Dynamic Origin-Destination Matrices Using Observed Link Travel Speed by Probe Vehicles. *Transportation and Traffic Theory 2009: Golden Jubilee*, 723–738.

Zheng, F. and Van Zuylen, H. (2012). Urban link travel time estimation based on sparse probe vehicle data. *Transportation Research Part C: Emerging Technologies*, 31, 145–157.

Wiedemann, R. (1974). Simulation des Straenverkehrsflusses. *Schriftenreihe des instuts fur verkehrswesen*, Universitat Karlsruhe, Heft 8.

Chapter 4
Application of a bi-level GLS model for estimating dynamic origin-destination flow

4.1 Introduction

In this chapter, we explore the estimation of O-D flows from probe vehicle data. We apply the idea of two-stage proposed by Yamamoto *et al.* (2009). For the first stage, the dynamic link flows are estimated by using the Bayesian method (BM) and ordinary method (OM), which are presented in chapter 3. For the second stage, O-D flows are estimated from estimated link flows. However, instead of using an entropy maximization method in Yamamoto *et al.* (2009), we describe a dynamic traffic assignment (DTA)-based bi-level generalized least squares (GLS) model in this chapter. This is an extension of the iterative bi-level estimation framework proposed by Tavana *et al.* (2001a), and we adopt the same notation for variables in the model formulation that follows.

4.2 Methodology

4.2.1 Bi-level GLS estimator

We consider a traffic network where L is the number of sensed links, and I and J are the numbers of origins and destinations, respectively. We are interested in finding a feasible vector O-D demand D for Γ aggregated time intervals, given a target demand vector $\widehat{D}$, and observed link flow vector $\hat{V}$ for T observation time intervals. The assignment of the O-D matrix onto the links in the network is made according to the link-flow proportion matrix $P = \{p_{(l,t)(\tau,i,j)}\}, l = 1,2,\cdots,L; t = 1,2,\cdots,T; \tau = 1,2,\cdots,\Gamma; i = 1,2,\cdots,I; j = 1,2,\cdots$, where each element $p_{(l,t)(\tau,i,j)}$ in the matrix represents the proportion of aggregated demand flow $d_{(\tau,i,j)}$ in aggregated time interval τ that flows

on link l during observation time interval t. Further, $\hat{v}_{(l,t)}$ is the element of $\hat{V}$ representing the observed link flow for link l during observation time interval t, and $\hat{d}_{(\tau,i,j)}$ is the element of $\hat{D}$ representing the target O-D demand for trips originating in zone i in aggregated time interval τ with destination j. It is noteworthy that the duration of the aggregated time interval can be one or several departure time intervals, and the departure time interval is equal to the observation time interval.

It has been pointed out that the error between the assigned link flow V and observed link flow $\hat{V}$ is a combined error E arising from inconsistencies in traffic assignment, measurement of traffic volumes and aggregating demand (Tavana *et al.*, 2001a). That is:

$$\hat{V} = PD + E \tag{4.1}$$

or

$$\hat{v}_{(l,t)} = \sum_{\tau,i,j} p_{(l,t)(\tau,i,j)} \cdot d_{(\tau,i,j)} + \varepsilon_{(l.t)} \tag{4.2}$$

where $V = PD$ and combined error E is a random vector with variance-covariance matrix W.

Tavana *et al.* formulated a bi-level GLS model that minimizes the above combined error E in objective function (Tavana *et al.*, 2001a):

$$min\, F(D) = (\hat{V} - PD)^T W^{-1} (\hat{V} - PD) \tag{4.3}$$

$$\text{subject to } P=\text{assignment } D \text{ from DTA} \tag{4.4}$$

$$D \geq 0$$

In reality a target O-D matrix (e.g. historical O-D demand from survey) when available can help to identify the estimated O-D matrices. Also, there are errors between the target and estimated O-D.

$$\hat{D} = D + \Pi \tag{4.5}$$

or

$$\hat{d}_{(\tau,i,j)=} d_{(\tau,i,j)} + \eta_{(\tau,i,j)} \tag{4.6}$$

where error Π is a random vector variance-covariance matrix Z.

The overall objective is thus to minimize the above two sets of errors in Equations (4.1 and 4.5). A weighted formulation is adopted to combine the two kinds of deviations, with $(1-\omega)$ and ω, respectively, reflecting the decision maker's preference or perceived importance for observed link flows and the target O-D matrix. Generally speaking, if no target O-D matrix or an unreliable O-D matrix is provided, the value of ω should be small and vice versa. Usually, if no further perceptual information about observed link flows and the target O-D matrix is known, ω is given a value of 0.5 for both terms to indicate no preference.

Similarly to the static case, the dynamic bi-level GLS estimator can be formulated as

$$min\, F(D) = \omega\left(\hat{D} - D\right)^T Z^{-1}\left(\hat{D} - D\right) + (1-\omega)(\hat{V} - PD)^T W^{-1}(\hat{V} - PD) \quad (4.7)$$

subject to P=assignment D from DTA (4.8)

$$D \geq 0$$

For the objective function, if both Z and W are set to be the identity matrix I, our formulation will drop back to the OLS presented by Zhou *et al.* (2003). Actually, W was set to I in Tavana's experiments, so the benefit of W has not been validated. In this research, the merits of both *Z* and *W* will be carefully considered and implemented.

As already noted, this proposed model is a DTA-based bi-level GLS estimator. The upper level is a constrained minimization problem with objective function GLS. Both the distance between the estimated and target O-D matrices and the distance between the calculated and observed link flows are considered in the objective function. The target O-D matrices can be obtained from census data. For example, the Ministry of Land, Infrastructure, and Transport of Japan (MLIT) conduct a census every five years with a sample rate of 3%, and this can provide hourly O-D matrices (Miwa *et al.*, 2010).

The lower level is a user equilibrium-based DTA process that is the same as Tavana's and Zhou's work (Tavana *et al.*, 2001a; Zhou *et al.* 2003).

4.2.2 Solution Method

The iterative solution procedure presented by Zhou *et al.* (2003) is applied to the proposed bi-level GLS model. This procedure is briefly outlined here, followed by a detailed discussion of the upper-level and lower-level procedures.

Step 1: (Initialization) k=0. Start from the target O-D matrix, obtain link-flow proportions P_0 from DTA simulator.

Step 2: (Optimization) Substituting the link-flow proportion matrix P_k, solve the upper level to obtain demand D_k.

Step 3: (Simulation) Using demand D_k, run the DTA simulator to generate new link-flow proportions P_{k+1}.

Step 4: (Evaluation) Calculate the deviation between simulated and observed link flows, and the deviation between estimated and target O-D matrices.

Step 5: (Convergence test) If estimated demand is stable or if no significant improvement is the overall objective, stop; otherwise k=k+1 and go to Step 2.

Specifically, regarding the upper level, the problem can be solved by forming the Lagrangian equation.

$$L(D, \Lambda) = F(D) + \Lambda^T(0 - D) \tag{4.9}$$

where Λ (lambda) is a vector of Lagrange multipliers with element $\lambda_{(\tau,i,j)}$.

Following a derivation process similar to that of Bell (1991) and Tavana *et al.* (2001a), we obtain the following equation:

$$D^* = G^{-1}(2\omega Z^{-1}\hat{D} + 2(1-\omega)P^T W^{-1}\hat{V} + \Lambda) \tag{4.10}$$

where $G = 2\omega Z^{-1} + 2(1-\omega)P^T W^{-1} P$.

If Λ is correctly obtained, the optimal estimated dynamic O-D matrix D^* is

accordingly solved. A constrained algorithm for the one-level GLS estimator with a lower limit as proposed by Bell (1991) and extended by Tavana *et al.* (2001a) is applied to solve Equation (4.10) by the following procedure.

Step 1 (Initialization)

Set $\Lambda = 0$ (unconstrained estimation)

Step 2 (Iteration)

Repeat calculation of D using Equation (4.10)

For $\tau = 1,2,\cdots,\Gamma; i = 1,2,\cdots,I \; and \; j = 1,2,\cdots J$

If $d_{(\tau,i,j)} < 0$ then

Set $\lambda_{(\tau,i,j)} = \lambda_{(\tau,i,j)} + (0 - d_{(\tau,i,j)})/g_{(\tau,i,j)}$

If $d_{(\tau,i,j)} > 0$ then

Set $\lambda_{(\tau,i,j)} = max\left(0, \lambda_{(\tau,i,j)} + \frac{0-d_{(\tau,i,j)}}{g_{(\tau,i,j)}}\right)$

until convergence

where $g_{(\tau,i,j)}$ is the principle element of G^{-1}.

For the lower level, we use the DTA module in VISSIM as the simulator. The reasons for this choice are twofold: first, microscopic models are more accurate in simulating real traffic than mesoscopic or macroscopic models (Jeihani, 2010) and while VISSIM is microscopic, DYNASMART-P and DynaMIT are mesoscopic; second, commercial packages like VISSIM have already been employed by planners and researchers, and are always available to any new users. A detail review of DTA computer packages can be found in Jeihani (2010).

4.3 O-D flow estimation results

In the O-D matrix estimation experiments, the probe ratio is assumed to be 0.1 and the matching link flow estimation results are used as the observed link flows defined in the bi-level GLS estimator. The variances as calculated from Equation (3.11) and Equation

(3.13) are used as the variances of the observed link flows in matrix W of Equation (4.7). A noise level of 50% is added to the true O-D matrix and the resulting O-D matrix is used as the target matrix. Therefore, the 50% dispersion in true demand is regarded as the variance of the target O-D demand in matrix Z of Equation (4.7). Co-variances of both the observed link flow and the target O-D demand are assumed to be 0. The value of ω is assumed to be 0.5 as in most studies. The aggregated time interval for O-D demand is taken to be the same as the departure time interval and observation time interval for link flows, which is 10 minutes. Therefore, there are 13 departure time intervals (denote as, DepIn1, DepIn2 …) in the simulation period from 7:50 am to 10:00 am. The first (DepIn1) and last (DepIn13) time intervals are not considered because of traffic initialization and ending.

Figure 4.1 shows the performance of GLS-BM, in which the estimated link flows obtained by BM are used as the observed link flows. The RMSE of O-D matrix doesn't decrease and remains roughly the same distance from the true O-D matrix for all time intervals; however, the calculated link flow is slightly improved by the GLS estimator for most time intervals. Then, we can say the initial O-D matrix is improved structurally. To explain this, we take a simple example. If a true O-D is [5,10], one estimated O-D is [7,11] and another estimated O-D is [4,12], then the RMSE for both estimated O-D matrices are the same, but they are different matrices and have different calculated link flow. In this case, the O-D matrix having small RMSE of calculated link flow is better.

Figure 4.2 shows the GLS estimation results when the link flows estimated by OM are used. It is clear that the estimation results fluctuate through the iterative process. Further, neither O-D flows nor link flows are improved by GLS for all time intervals. It is also clear that the estimation process is stable with GLS-BM while with GLS-OM it is unstable. The reason for this may lie in two differing characteristics of BM and OM:

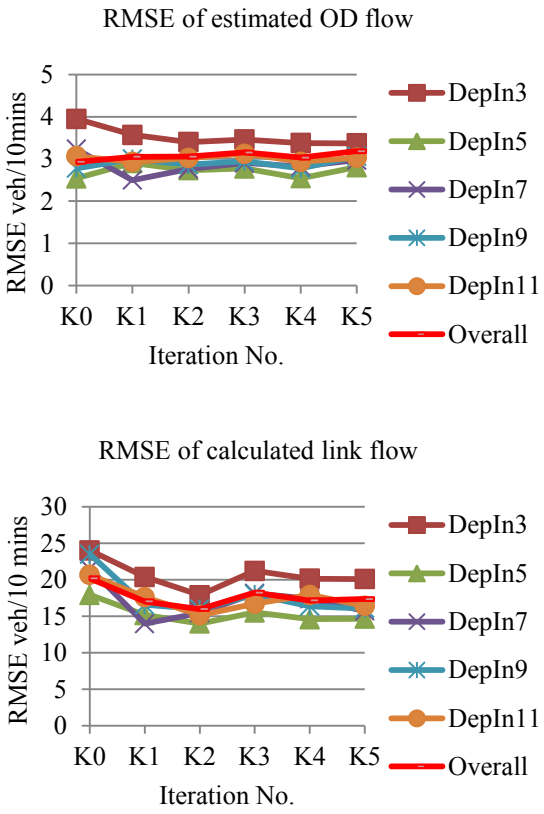


Figure 4.1 Estimation performance of GLS-BM

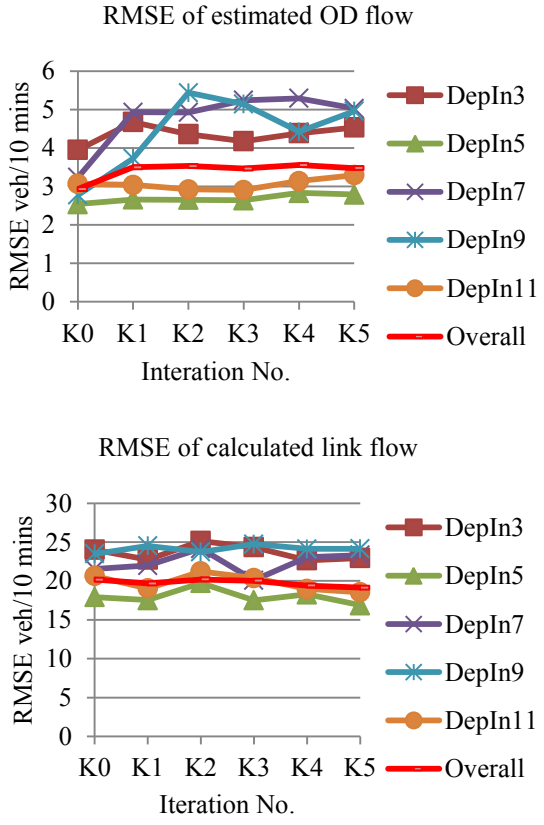


Figure 4.2 Estimation performance of GLS-OM

1) BM can give an acceptable estimated link flow even if there is no probe vehicle while OM cannot; and 2) the variance given by BM reflects the weights of links better than OM.

To validate the proposed bi-level GLS estimator, the bi-level Ordinary Least-Square estimator, in which Z and W are both set to be the unit matrix, is also applied and implemented. Therefore, with link flows estimated using the various methods described, there are four O-D estimation variants in the validation: GLS-BM, GLS-OM, OLS-BM and OLS-OM. In addition, an OLS estimator using the true link flow (OLS-TrueLF) is also tested for comparison.

Figure 4.3 shows the estimation results for these variants. Among them, the performance of GLS-BM and GLS-OM has already been discussed in the section above. The RMSE of both estimated O-D flow and calculated link flow for OLS-BM rises slightly with the number of iterations. The OLS-OM variant has the same trend, but is slightly inferior to OLS-BM. When the true link flow is used in OLS, the estimation results are improved slightly for estimated O-D flow and markedly for calculated link flow. This confirms that the reliability of observed data in O-D matrix estimation plays an important role, as Yamamoto *et al.* pointed out (Yamamoto *et al.*, 2009).

It should be noted that, in both diagrams in Figure 4.3, using the link flow estimated by BM produces better O-D flow estimation results than that by OM no matter whether GLS or OLS are applied. This observation is consistent with that in link flow estimation, where the RMSE of estimated link flow by BM is 29.65 vehs/10mins, better than the 31.07 vehs/10mins of the estimate by OM (Table 3.1). What is more, O-D flows estimated by GLS are more accurate than those by OLS, which confirms that GLS has better ability to estimate O-D flows than OLS.

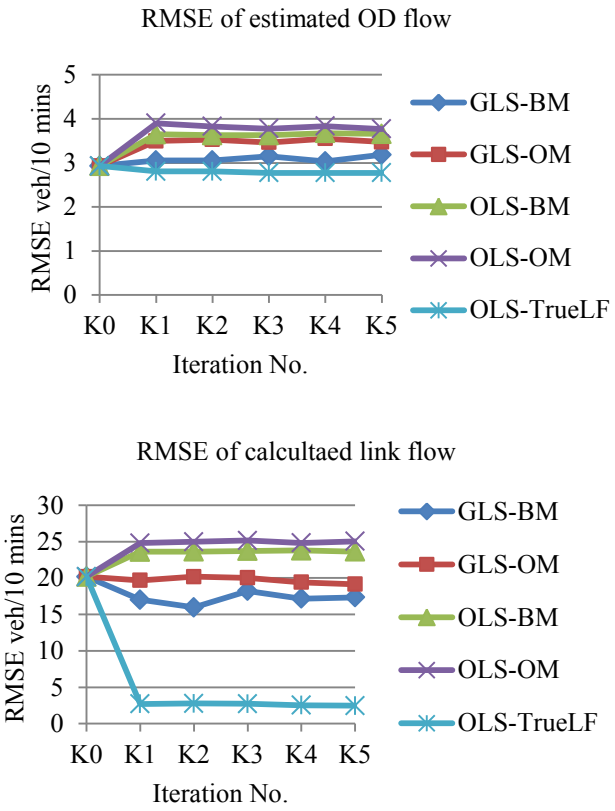


Figure 4.3 Comparative performance of various methods

4.4 Convergence of extended Bell algorithm

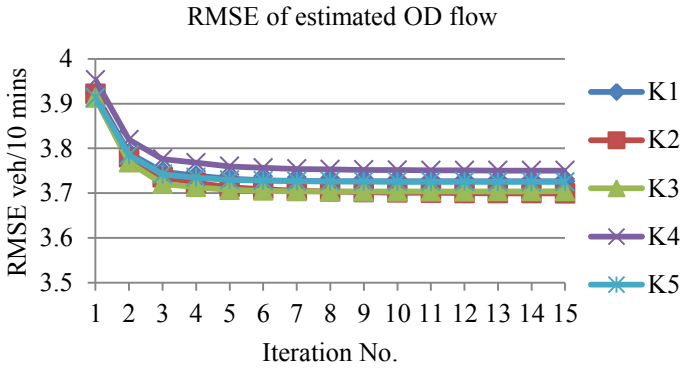


Figure 4.4 Convergence process of extended Bell algorithm for OLS-BM

For each iteration of the bi-level program, the upper-level constrained minimization model is solved by the extended Bell algorithm. This is also an iterative process. Figure 4.4 shows the results given by the extended Bell algorithm for five iterations of the bi-level model for the case of OLS-BM. It can be seen that the RMSE of estimated O-D flow decreases as more iterative cycles are performed and convergence is reached in

less than 10 iterations. In the experiments with other variants, the extended Bell algorithm exhibits similar convergence behavior to that shown in Figure 4.4.

4.5 Discussion of results

The link flows estimated by BM and OM in chapter 3 are used as the observed link flows in the proposed bi-level GLS estimator to obtain O-D matrix estimations. The results successfully validate BM as superior to OM in link flow estimation for dynamic O-D matrix estimation. This is also true if GLS is substituted by OLS. However, it seems it is difficult to improve the target O-D matrix using link flows estimated by either BM or OM. There are two possible reasons for this. Firstly, the estimated link flows provide information that is no more accurate than the target O-D matrix, because the initial RMSE of link flow assignments from the target O-D matrix is 20.17 vehs/10mins (Figure 4.3), but the RMSEs of estimated link flows for BM and OM are 29.65 vehs/10mins and 31.07 vehs/10mins, respectively (Table 3.1). This is also verified by the demonstration that both O-D matrix and link flows are improved when the true link flow is used in OLS. Secondly, the apportioning of link-flows is dependent on O-D demand, because the study concerns a congested network during the morning peak. The congestion effects make the link flows nonlinear with O-D demand.

Convergence of the extended Bell algorithm is demonstrated in this study through validation experiments. This convergent behavior has not been shown in other studies of dynamic O-D estimation. In Tavana's analysis, convergence of the Bell algorithm is not guaranteed in both static and dynamic O-D estimation (Tavana, 2001b). Convergence is achieved in this work because, in contrast with Tavana's work, we incorporate the distance between estimated and calculated O-D flows in the objective function. This results in matrix G in Equation (4.10) being a positive definite matrix, leading to convergence of the solution procedure.

4.6 Summary

Probe vehicle systems have been increasingly implemented in metropolitan areas as a way to collect traffic information and provide dynamic route guidance. In this study, we propose a method for estimating dynamic O-D matrices from probe vehicle data from the perspective of practical applications.

In the first step of proposed method, link flows are inferred from a prior distribution of vehicle speed and current probe vehicle data by using Bayesian inference. It has been shown that BM can effectively use the prior distribution of vehicle speed accumulated from archived probe vehicle data, and produce more accurate link flow estimate than ordinary method when probe ratio is less than 0.2. This feature makes BM more applicable in practice, as the probe ratio is low in most current probe systems. Additionally, the variance of estimated link flow by BM can better capture the difference in the reliability of link flow estimate among links than OM. In the second step, a DTA-based bi-level GLS estimator is formulated to estimate dynamic O-D matrices from the estimated link flows and historical O-D matrices. It is an extension of Tavana's model, both the distance between the estimated and target O-D matrices and the distance between the calculated and observed link flows are considered in the objective function. Hence the extended Bell algorithm can be used to solve our model. Moreover, the GLS formulation can utilize the variance of estimated link flow from the first step, while setting an identical variance of link flow for links decrease GLS to be OLS in Tavana's model. A case study on a mid-size signalized urban network illustrates that BM is superior to OM and GLS to OLS.

To make the proposed method more capable in application, we chose a commercial system VISSIM as the DTA simulator for a DTA-based dynamic O-D matrix estimation model. As a whole, the proposed method can be applied in practice

for estimating dynamic O-D matrices using probe data, such as in probe vehicle-based dynamic route guidance system, or in the situation that link counts are not available.

Although only links equipped with fixed detectors were used in the case study, the proposed method does not in fact depend on whether a fixed detector is present in a link or not. Since the number of probe vehicles traveling on a link changes with time, different link sets can be picked out for different time intervals in future applications of the method.

References

Bell, M. G. (1991). The Estimation of Origin-Destination Matrices by Constrained Generalized Least Squares. *Transportation Research Part B: Methodological*, 25(1), 13–22.

Jeihani, M. (2010). A. Review of Dynamic Traffic Assignment Computer Packages. *Journal of the Transportation Research Forum*, 46(2).

Miwa, T., Y. Okada and T. Morikawa. (2010). Applying a Structured Dispersion Parameter to Multiclass Stochastic User Equilibrium Assignment Model. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2196, Transportation Research board of the National Academies, Washington, D.C., 142–149.

Tavana, H., Mahmassani, H. S. (2001a). Estimation of Dynamic Origin-Destination Flows from Sensor Data Using Bi-level Optimization Method. *Transportation Research Board CD-ROM Preprints (TRB Paper No. 01-3241)*, National Research Council, Washington, D. C.

Tavana, H. (2001b). Internally-Consistent Estimation of Dynamic Network Origin-Destination Flows from Intelligent Transportation Systems Data Using Bi-Level Optimization. Ph.D. Dissertation, the University of Texas at Austin.

Yamamoto, T., Miwa, T., Takeshita, T. and Morikawa, T. (2009). Updating Dynamic Origin-Destination Matrices Using Observed Link Travel Speed by Probe Vehicles. *Transportation and Traffic Theory 2009: Golden Jubilee*, 723–738.

Zhou, X., Qin, X. and Mahmassani, H. S. (2003). Dynamic Origin-Destination Demand Estimation with Multiday Link Traffic Counts for Planning Applications. In *Transportation Research Record: Journal of the Transportation Research Board*, 1831, 30–38.

Chapter 5
Modeling link travel time distribution by incorporating truncated distribution

5.1 Introduction

Travel time is one of the most important measures for evaluating the performance of a traffic system. It is a concept that is well understood by both traffic engineers and the public. There are multiple factors that affect travel time, including driver behavior, road attributes, traffic conditions, and signal timings. As a result, travel times along road sections are randomly distributed. Therefore, travel time distribution is an unavoidable issue in the researches of travel time estimation, travel time variability and travel time reliability.

Hofleitner *et al.* (2012) formulate free-flow time and stopping time resperctively, and then combine them to obtain link travel time distribution. However, delays due to traffic congestion and acceleration/deceleration are not considered in their research. In this study, we decompose travel time in signalized road section into time-in-motion (the time for which a vehicle is actually moving) and time-in-queue. The time-in-motion, which contains free-flow time and delays due to traffic congestion and acceleration/deceleration, mainly depends on the physical attributes of the road and traffic conditions. The time-in-queue mainly relates to signal timings and the signal offset between adjacent intersections. In this way, we first model the time-in-motion and time-in-queue separately, and then from the results derive the link travel time distribution.

A further motivation for this work comes from the fact that travel time is distributed over a limited range instead of over the whole positive domain of the real axis. Wang *et al.* (2012) first introduced a truncated distribution into traffic quantity

modeling. If truncation is ignored, bias may arise in estimating the average and variance of travel time. In this chapter, truncation of vehicle time-in-motion is analyzed and then a truncated distribution is used to model time-in-motion and link travel time.

To analyze the effect of polling frequency of probe data on calibrating the proposed link travel time distribution, the Kolmogorov-Smirnov test is performed to test the estimated models of link travel time distribution from probe data with various polling frequencies.

This chapter is organized as follows. Section 5.2 explains the concept of a truncated distribution and the estimation of its parameters. In Section 5.3, the probability distribution function of link travel time is derived using the truncated distribution. In Section 5.4, numerical experiments are carried out and the results are analyzed. Finally, in Section 5.5, a summary of the conclusions of the work is presented.

5.2 Truncated distribution

5.2.1 Definition

In statistics, a truncated distribution is a conditional distribution that results from restricting the domain of a probability distribution. The following discussion is limited to the case of a continuous random variable, since travel time is continuous. Suppose a random variable *X* has infinite support (i.e. non-zero-valued to infinity) according to some probability density function (PDF) $f(x, \Theta)$ (where Θ represents the vector of parameters). Conditional on the value of *X* being limited in range $[a, b]$, the variable $X|a \leq x \leq b$ follows a truncated distribution with the following PDF:

$$g(x, \Theta, a, b) = \frac{f(x,\Theta)}{\int_a^b f(t)dt}, x \in [a, b] \tag{5.1}$$

The threshold points a, b are called truncation points. In the particular case where

$a = -\infty$, the truncated distribution becomes a right-truncated distribution. A left-truncated distribution can be similarly obtained. In comparison with the non-truncated distribution, the truncated distribution has a limited domain $[a, b]$ and two additional parameters, a, b.

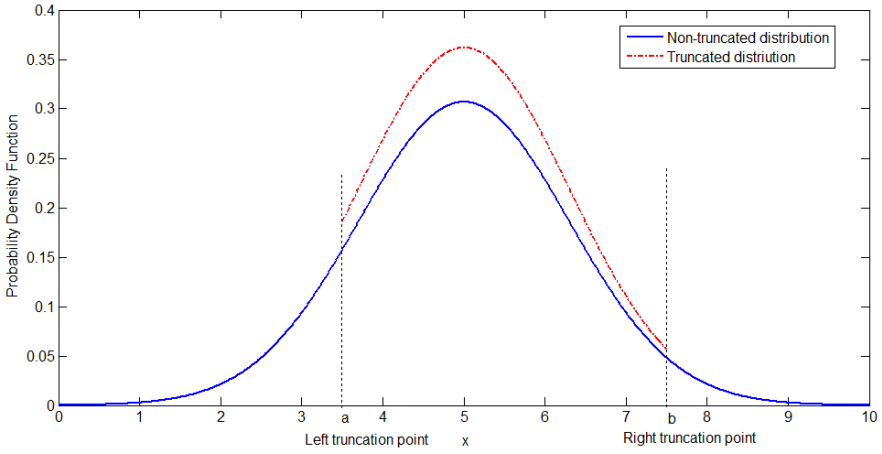


Figure 5.1 Non-truncated and truncated distributions

5.2.2 Parameter estimation

In literatures relating to parameter estimation, it is generally assumed that the two additional parameters in a doubly truncated distribution (i.e. the truncation points) are known (Johnson *et al.* 1994; Cohen 1959). However, the truncation points of travel times are not available in practice. We propose a simple method of estimating these truncation points and provide a proof for it. The other parameters of the truncated distribution can be obtained using methods described in Johnson *et al.* (1994) and Cohen (1959).

Assume a sample series of travel time variables $X_1, X_2, \cdots, X_n$ with observations $x_1, x_2, \cdots, x_n$. These variables all follow the same truncated travel time distribution with PDF $g(x, \Theta, a, b)$ and cumulative distribution function (CDF) $G(x, \Theta, a, b)$ denoted by $F_X(x)$. Then the estimated a and b are given by:

$$\hat{a} = min\{x_1, x_2, \cdots, x_n\} \tag{5.2}$$

$$\hat{b} = max\{x_1, x_2, \cdots, x_n\} \tag{5.3}$$

Explanation: Define a random variable $Y = min\{X_1, X_2, \cdots, X_n\}$, then the CDF of Y can be derived: $F_Y(x) = P\{min\{X_1, X_2, \cdots, X_n\} \leq x\} = 1 - (1 - F_X(x))^n$. We have $F_Y(a) = 0$, since $F_X(a) = 0$ for a truncated distribution. The probability of Y falling within a small interval $[a, a + \Delta]$ is $P\{a \leq Y \leq a + \Delta\} = F_Y(a + \Delta) - F_Y(a) = 1 - (1 - F_X(a + \Delta))^n$. If we set a confidence level $1 - \alpha$, we can calculate a minimum sample size $n = \lceil log^{\alpha}_{1-F_X(a+\Delta)} \rceil$ from $P\{a \leq Y \leq a + \Delta\} > 1 - \alpha$. For example, if $\alpha = 0.05$, $F_X(a + \Delta) = 0.05$, then the minimum sample size is 59. This means, in the example case, that as long as the sample size is greater than 59, we have a confidence of 0.95 using $min\{x_1, x_2, \cdots, x_n\}$ as an estimate of a. Using a similar argument, we obtain the same result for estimating b as $max\{x_1, x_2, \cdots, x_n\}$.

After a, b have been obtained, the other parameters Θ can be estimated by the maximum likelihood (or more conveniently the log-likelihood) estimator (MLE).

$$\hat{\Theta} = arg\,max\ \sum_i lng(x_i, \Theta, \hat{a}, \hat{b}) \tag{5.4}$$

In traffic research, the normal distribution and lognormal distribution are frequently utilized. Further information on truncated distributions can be found in the literatures (Cohen 1959; Johnson *et al.* 1994; Wang *et al.* 2012).

5.3 Modeling distribution of travel time

In arterial networks, traffic is driven by the formation and dissipation of queues at intersections. On an urban road, a vehicle is either in motion at a certain speed or stationary in a queue waiting at a red signal. The distributions of time-in-queue and time-in-motion are first formulated independently and then combined using convolution.

5.3.1 Probability distribution of time-in-queue

To model the probability distribution of time-in-queue, we adopt an analytical model derived by Holfleitnet *et al.* (2012) based on several assumptions: 1) the fundamental

diagram is triangular; 2) Traffic are stable during each estimation interval; 3) arrivals are uniform.

A fraction η of the vehicles entering the link would experience a stop. The remaining fraction $1-\eta$ of the vehicles travel though the link without stopping. For vehicles reaching the queue at an intersection, time-in-queue is uniform under the assumption that the arrival of vehicles is uniform in time. The uniform distribution has support $[0, R]$ corresponding to the minimum and maximum queuing times, where R is red time. For vehicles that do not experience a stop and travel straight through the intersection, the time-in-queue is 0.

Therefore, the time-in-queue of vehicles going through an intersection is a random variable T_q with a mixed distribution of two components. The first component represents vehicles that experience a stop with uniform distribution on $[0, R]$. The second component represents the vehicles that do not experience a stop with a mass distribution in 0. We note $\mathbf{1}_A$ the indicator function of set A,

$$\mathbf{1}_A(x) = \begin{cases} 1, x \in A \\ 0, x \notin A \end{cases}$$

The Dirac distribution centered in x is denoted $Dir_{\{x\}}(*)$.

The probability distribution of time-in-queue with respect to time-in-queue t_q is then formulated as:

$$h(t_q) = (1-\eta) Dir_{\{0\}}(t_q) + \eta \frac{1}{R} \mathbf{1}_{[0,R]}(t_q) \tag{5.5}$$

where

$$Dir_{\{0\}}(t_q) = \begin{cases} 1, t_q = 0 \\ 0, t_q \neq 0 \end{cases}, \quad \mathbf{1}_{[0,R]}(t_q) = \begin{cases} 1, t_q \in [0, R] \\ 0, t_q \notin [0, R] \end{cases}$$

5.3.2 Probability distribution of time-in-motion

In principle, a vehicle on the road can be in only one of two states: in motion or stopping in a queue. Hellinga *el al.* (2008) decomposed total travel time T_t experienced by a

vehicle on a link into four constituent parts: free-flow travel time T_f, delay due to traffic congestion T_c, deceleration and acceleration time T_{da}, and time-in-queue T_q. The first three parts make up the time-in-motion T_m. The sum of the first two parts is similar to the travel time on a freeway, which is factored by link attributes (such as link length, number of lanes), driver behavior, and level of traffic congestion. Time-in-queue is mainly caused by the presence of traffic signals. To simplify the modeling process, we also regard deceleration and acceleration time as one component of time-in-motion. This simplification actually has little influence on the calculation of total travel time because deceleration and acceleration time are relatively minor compared with the other components. Thus,

$$T_m = T_f + T_c + T_{da} \tag{5.6}$$

We formulate the probability of time-in-motion as a truncated distribution with PDF $g(t_m, \Theta, a, b)$. Specifically, a truncated normal distribution and a truncated lognormal distribution are tested in a later section.

5.3.3 Link travel time distribution

As discussed above, total link travel time should be the sum of the free-flow travel time T_f, delay due to traffic congestion T_c, deceleration and acceleration time T_{da}, and time-in-queue T_q. Thus, we have:

$$T_t = T_m + T_q \tag{5.7}$$

Assume T_m and T_q are independent variables, then the PDF of total travel time T_t equals the convolution of the PDFs of T_m and T_q. According to the linearity of convolution, we can derive the PDF of T_t:

$$p^{tr}(x) = (1-\eta)g(x, \Theta, a, b) + \frac{\eta}{R}\int_{-\infty}^{+\infty} \mathbf{1}_{[0,R]}(x-z)g(z, \Theta, a, b)dz \tag{5.8}$$

The integration is not null if and only if $x - z \in [0, R]$. Since $g(z, \Theta, a, b)$ is not equal to zero for $z \in [a, b]$, the integration is not null if and only if $z \in [x - R, x] \cap$

$[a, b]$. We assume $b - a \leq R$, since time-in-motion generally varies within a narrow range under the assumptions of stationary of traffic. Then the truncated model of the PDF for link travel time reads:

$$p^{tr}(x) = \begin{cases} 0, \quad if\ x < a \\ (1-\eta)g(x,\Theta,a,b) + \frac{\eta}{R}\int_a^x g(z,\Theta,a,b)dz\,,\ if\ x \in [a,b] \\ (1-\eta)g(x,\Theta,a,b) + \frac{\eta}{R}\int_a^b g(z,\Theta,a,b)dz\,,\ if\ x \in (b, a+R] \\ (1-\eta)g(x,\Theta,a,b) + \frac{\eta}{R}\int_{x-R}^b g(z,\Theta,a,b)dz\,,\ if\ x \in (a+R, b+R] \\ 0, \quad if\ x > b+R \end{cases} \tag{5.9}$$

This derived PDF has a support of $[a, b + R]$. This is true in reality in that link travel time only changes over a limited range that is the subset of the domain of positive real numbers.

To validate the application of this truncated distribution, we present the model proposed by Hofleitner *et al.* (2012) as an example of using non-truncated distributions. Without considering delay due to traffic congestion or deceleration and acceleration time, Hofleitner *et al.* (2012) use the following equation:

$$T_t = T_f + T_q \tag{5.10}$$

As with the above derivation process, Hofleitner *et al.* (2012) obtained the PDF by assuming that T_f follows a non-truncated distribution $g(x, \Theta)$ (Hofleitner's model):

$$p^{nt}(x) = \begin{cases} 0, \quad if\ x < 0 \\ (1-\eta)g(x,\Theta) + \frac{\eta}{R}\int_0^x g(z,\Theta)dz\,,\ if\ x \in [0,R] \\ (1-\eta)g(x,\Theta) + \frac{\eta}{R}\int_{x-R}^x g(z,\Theta)dz\,,\ if\ x > R \end{cases} \tag{5.11}$$

5.3.4 Fitting the travel time distribution

The derived PDF $p^{tr}(x)$ is parameterized by the fraction of stopping vehicles η, signal red time R, and the motion behavior parameters left truncation point a, right truncation point b, and Θ. In comparison, the PDF $p^{nt}(x)$ using the non-truncated distribution has the same parameters except a, b.

Then, the estimation of the parameters is done by maximizing the log-likelihood

of the link travel times. The estimation problem is given by:

$$maximize \quad \sum_i lnp^{tr}(x_i, \eta, R, \Theta, a, b) \tag{5.12}$$

The optimization problem is not convex and can be restrained to a small scale with a feasible set by placing additional constraints and bounds that yield physically acceptable values. In most cases, link travel times only are available using probe vehicle technology or other traffic information collecting tools. Thus, a grid search is implemented to obtain the best estimates of parameters. In Section 5.4.3, a method is introduced for reducing the computational burden by using the characteristics of vehicles' time-in-motion. In cases like high-frequency probe data (where vehicle information consisting of location, time and speed is recorded every second), the time-in-motion can be easily extracted. Truncation points a, b can be first estimated by using Equations (5.2) and (5.3), then the other parameters can be estimated using Equation (5.12) based on the estimated a, b.

5.4 Numerical experiments and results

5.4.1 Simulation network

We consider an ideal left-hand traffic network with four identical intersections. The network is modeled in VISSIM, as showed in Figure 5.2. The intersections are managed by signal controllers with identical signal timings. The three links between the intersections are each 300m long. Traffic entering a link from one end comes from three directions: left, straight and right. It then leaves the link in three directions. Vehicles that exit by turning left or going straight ahead are controlled by one signal. The link travel times of these vehicles are used in later experiments. A constant flow of 500 veh/h is input into the network at each entrance, and 10% probe vehicles are assumed in the traffic. In the simulation, vehicle data are recorded every 0.1s so that accurate travel times can be obtained for each vehicle.

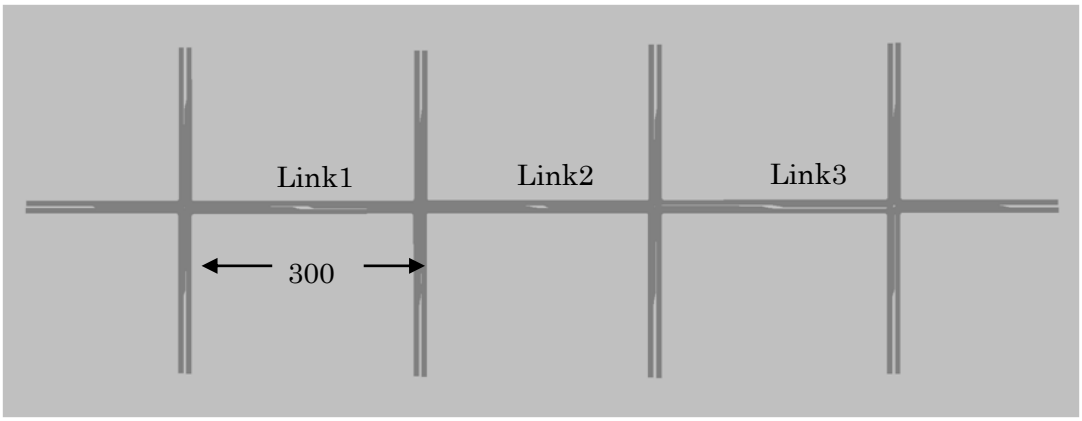


Figure 5.2 A four-intersection simulation network in VISSIM

5.4.2 Truncation of time-in-motion

A set of link travel times is obtained by simulating the above network. There is a stationary flag indicating whether a vehicle is stationary or not. The time-in-queue of a vehicle is calculated by observing the duration of the stationary flag. Time-in-motion is then extracted as link travel time minus time-in-queue. Histograms of time-in-motion for the three test links is illustrated in Figure 5.3. All times in motion clearly fall within a limited range for all links. Further, the truncation of times in motion is quite significant on the lower side.

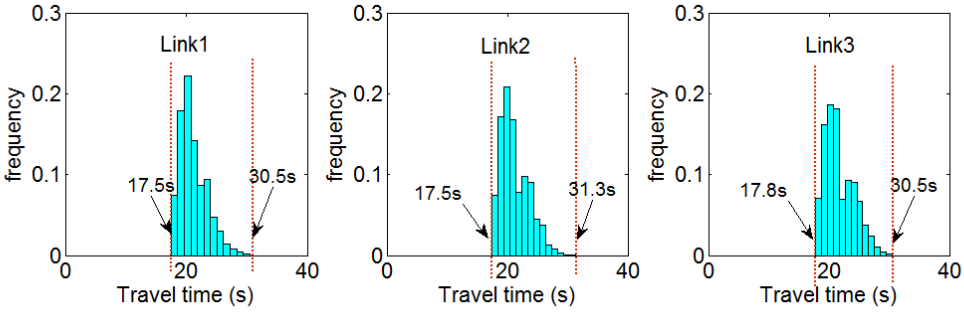


Figure 5.3 Histograms of time-in-motion for various links

To explicitly analyze this truncation, times in motion are fitted using non-truncated distributions and truncated distributions. That is, we model the time-in-motion distribution using the normal distribution, the lognormal distribution, and the respective truncated versions of these distributions. The distributions are measured with a small-sample-size corrected version of the Akaike Information

Criterion (AICc) and the Bayesian Information Criterion (BIC). AICc and BIC are two measures of the relative goodness of fit of a statistical model. The preferred model is the one with the minimum AICc and BIC value.

$$AICc = 2k - 2\,ln(L) + \frac{2k(k+1)}{n-k-1} \tag{5.13}$$

$$BIC = -2 \cdot lnL + kln(n) \tag{5.14}$$

where k is the number of parameters in the model, n is the sample size and L is the maximized value of likelihood function for the estimated model.

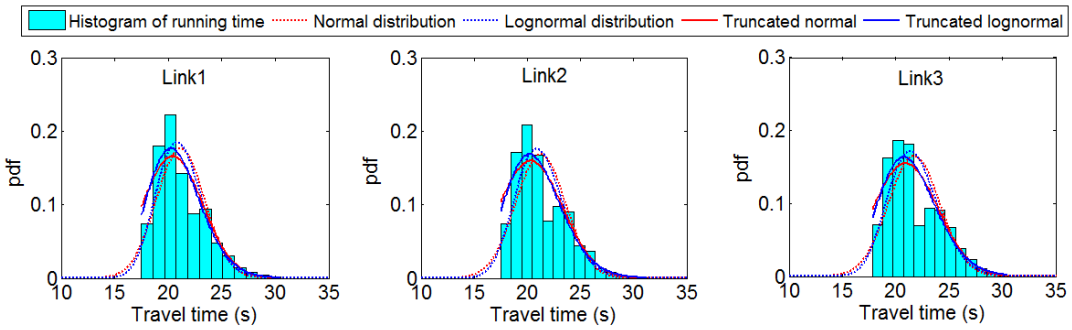


Figure 5.4 Fitted distributions of time-in-motion for various links

The fitted time-in-motion distributions are shown in Figure 5.4. It is clear that there are significant differences between the truncated and non-truncated distributions for lower times in motion, but almost no differences for higher values of time-in-motion. This feature tells us that careful consideration of left truncation is needed, while right truncation can be neglected because it is almost insignificant statistically.

The values of AICc and BIC for the candidate models are showed in Table 1. Values of AICc and BIC for the truncated distributions are all smaller than for the non-truncated distributions. This suggests that the truncated distributions are preferred as a model for time-in-motion. This is also consistent with the observations made about Figure 5.4. Additionally, it seems that the truncated lognormal distribution is slightly superior to the truncated normal distribution. The reason for this may be that the distributions of time-in-motion have a certain degree of skew, which is better captured

by the truncated lognormal distribution. However, the difference is not sufficient to reach the general conclusion that a truncated lognormal distribution is better than a truncated normal distribution for modeling time-in-motion, since the difference here is not significant statistically..

Table 5.1 Values of AICc and BIC for candidate models of time-in-motion distribution

Link ID	Measure	Normal distribution	Lognormal distribution	Truncated normal	Truncated lognormal
Link1	AICc	9971	9785	9597	9552
	BIC	9982	9797	9620	9575
Link2	AICc	11105	10922	10661	10636
	BIC	11116	10934	10685	10659
Link3	AICc	11108	10946	10703	10679
	BIC	11119	10958	10726	10703

5.4.3 Link travel time distribution

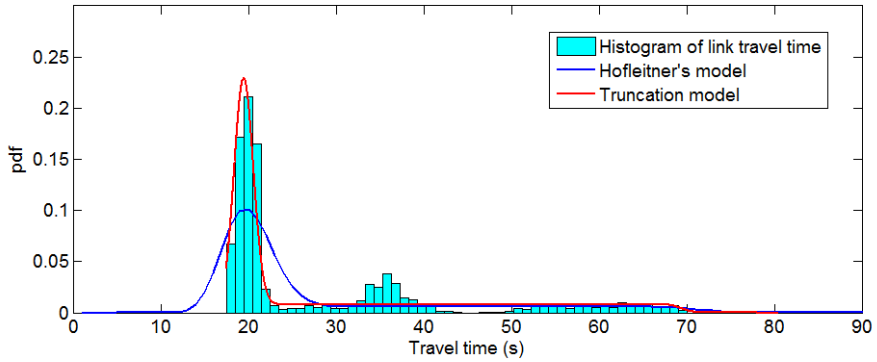


Figure 5.5 Estimated link travel time distribution

From Equation (5.7) ($T_t = T_m + T_q$), we obtain $min\{T_t\} = min\{T_m + T_q\} = min\{T_m\} + min\{T_q\}$. Since the minimum time-in-queue T_q is 0, then we have $min\{T_t\} = min\{T_m\}$. Using Equation (5.2), the left truncation point a is estimated as $\hat{a} = min\{T_t\}$. We know from Section 5.4.2 that the right truncation of time-in-motion is not significant. Thus, we can set the right truncation point b to a large number. Since we assumed $b - a \leq R$ in Section 3.3, b is set at $\hat{a} + R$. After determining a and b in this way, we can estimate the other parameters using MLE as described in Section

5.3.4.

The normal distribution is used to model free time in Hofeitner's model. In this truncation model, the truncated normal distribution is used to model time-in-motion. In Figure 5.5, we fit the travel time distribution for link 1 using both Hofeitner's model and the truncation model. The truncation model (AICc: 13272; BIC: 13238) provides better performance than Hofeitner's model (AICc: 14505; BIC: 14482). In particular, the truncation model fits the lower travel times well. However, the peak in travel time seen between 30s and 40s is not matched by either model. The reason for this may be that link flows do not fit the hypothesis of constant arrivals due to light synchronization (Bails *et al.* 2012).

5.4.4 Model estimation using probe vehicle data

In practice, the polling interval (PI) of most probe vehicles varies from 1s per point to 60s per point. The recorded travel time from probe vehicle should be allocated to individual links to obtain link travel time, since the two consecutive polled positions do not necessarily correspond to the end points of individual links. Consequently, the polling frequency would directly affect the accuracy of allocated link travel time, and the accuracy becomes poorer with lower polling frequency (Hellinga *et al.* 2008). To analyze the effects of polling frequency on the estimation of proposed model, we first calculate link travel time assuming uniform motion between consecutive polled positions (Miwa *et al*. 2004) (this travel time allocation method is called proportional allocation in this thesis), then estimate model parameters using MLE as described in Section 5.3.4.

The estimated link travel time distributions from probe data with various polling frequencies (or polling intervals) for link 1 is plotted in Figure 5.6. We make the Kolmogorov-Smirnov (K-S) tests for each estimated distributions by testing the

hypothesis *H_0: the link travel times are distributed according to the estimated distributions*. The K-S test is a standard non-parametric test to state whether samples are distributed according to a hypothetical distribution. This test is based on the K-S statistics which is computed as the maximum difference between the empirical and the hypothetical cumulative distributions. The test provides a p-value which informs on the goodness of the fit. Low p-values indicate that the data does not follow the hypothetical distribution. We reject hypothesis H_0 for p-values inferior to the significant level α. The parameter α is commonly set to 0.05 or 0.1.

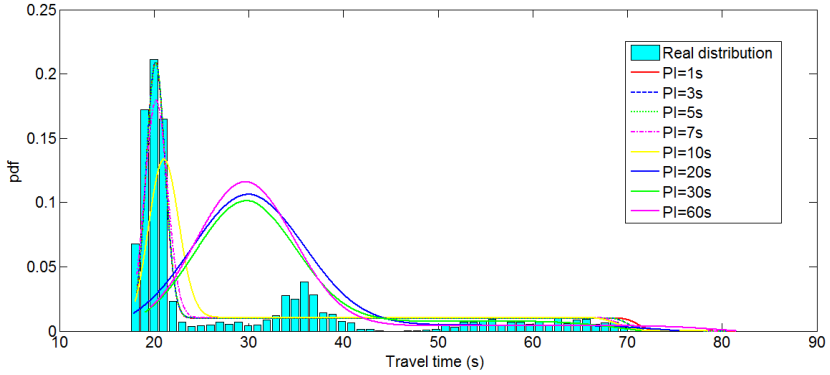


Figure 5.6 Estimated link travel time distributions from probe data with various polling frequencies

Table 5.2 The K-S test for estimated link travel distributions

	Polling interval							
	1s	3s	5s	7s	10s	20s	30s	60s
p-value	0.2587	0.2944	0.2936	0.3265	0.0412	5.13×10^{-12}	9.43×10^{-10}	6.40×10^{-11}

The results of tests show the estimated distributions from probe data with polling interval less than 7s pass the test for both $\alpha = 0.05$ and $\alpha = 0.1$, while the estimated distribution only pass the test for $\alpha = 0.1$ and fails the test for $\alpha = 0.05$ if the polling interval is 10s (Table 5.2). Additionally, the estimated distributions are significantly different from the real distribution when the polling interval is longer than 20s. The results of K-S test are also consistent with the observations in Figure 5.6.

Therefore, we can conclude that the link travel time distribution can be estimated at an acceptable level if the polling interval of probe data is less than 10s. However, the

results essentially depend on the accuracy of allocated link travel time. As mentioned above, proportional allocation is applied to calculate link travel time from probe data in the research. If more accuracy link travel times are obtained from an advanced method of travel time allocation, the link travel time distribution probably can be estimated from probe data with longer polling interval.

5.5 Summary

In this chapter, a truncated distribution is used to model the link travel time distribution in a signalized section of road. The probability distribution function of travel time is derived and parameterized by fraction of stationary vehicles, red signal time, and motion behavior parameters including truncation points. These parameters are estimated from travel time data using a maximum likelihood estimator.

We decompose link travel time into time-in-motion and time-in-queue, modeling them independently. The time-in-motion, which is similar to travel time on a freeway, mainly depends on physical attributes of the link and traffic conditions. The time-in-queue is mainly related to signal timings and the signal offset between adjacent intersections.

We introduce a simple method of estimating truncation points based on a relatively small sample size. Time-in-motion is then modeled using both non-truncated distributions (normal distribution and lognormal distribution) and truncated distributions (truncated normal distribution and truncated lognormal distribution). Using AICc and BIC criteria, we show that it is better to model time-in-motion with a truncated distribution instead of a non-truncated one. A comparison is carried out between the derived travel time distribution and an example model using a non-truncated distribution. The results validate the superiority of applying a truncated distribution to model the distribution of link travel times.

The Kolmogorov-Smirnov test is performed to test the estimated models of link travel time distribution from probe vehicle data with various polling frequencies. The results show, the link travel time distribution can be estimated from probe data that polling interval is less than 10s when the significance level is set to 0.01.

Following the suggestions given by Wang *et al.* (2012), we recommend that truncation should be considered for many of the quantities (e.g. travel time, speed and demand) used in traffic engineering. Truncated distributions should be used in any case where truncation cannot be neglected.

References

Bails, C., Hofleitner, A., Xuan, Y. and Bayen, A. (2012). A Three-Stream Model for Arterial Traffic. In *Proceedings of the 91st Transportation Research Board Annual Meeting (No. 12-1212)*.

Cohen, A. C. (1959). Simplified estimators for the normal distribution when samples are singly censored or truncated. *Technometrics*, 1(3), 217–237.

Hellinga, B., Izadpanah, P., Takada, H. and Fu, L. (2008). Decomposing travel times measured by probe-based traffic monitoring systems to individual road segments. Transportation Research Part C: Emerging Technologies, 16(6) , 768–782.

Hofleitner, A., Herring, R. and Bayen, A. (2012). Probability Distributions of Travel Times on Arterial Networks: Traffic Flow and Horizontal Queuing Theory Approach. In *Proceedings of the 91st Transportation Research Board Annual Meeting (No. 12-0798)*.

Johnson, N., Kotz, S. and Balakrishnan, N. (1994). Continuous Univariate Distributions, Vol. 1-2. New York: John Wiley & Sons, 80–249

Liu, H. X. and Ma, W. A. (2009). virtual vehicle probe model for time-dependent travel time estimation on signalized arterials. *Transportation Research Part C: Emerging Technologies*, 17(1), 11–26.

Miwa, T., Sakai, T. and Morikawa, T. (2004). Route Identification and Travel Time Prediction Using Probe-Car Data. International Journal of ITS Research, Vol. 2, No. 1.

Ramezani, M. and Geroliminis, N. (2012). Estimation of Arterial Route Travel Time Distribution with Markov Chains. In *Proceedings of the 91st Transportation Research Board Annual Meeting (No. 12-0614)*.

Wang, Y., Dong, W., Chin, S., Papageorgiou, M., Rose, G. and Young, W. (2012). A note on speed modeling and travel time estimation based on truncated normal and log-normal distribution, In *Proceedings of the 91st Transportation Research Board Annual Meeting (No. 12-2534)*.

Zheng, F. And Van Zuylen, H. (2012). Urban link travel time estimation based on sparse probe vehicle data. *Transportation Research Part C: Emerging Technologies*, 31, 145–157.

Chapter 6

Modeling joint distribution of vehicle location and speed on signalized road

6.1 Introduction

In numerous cities around the world, traffic congestion is a crucial problem that adversely and significantly affects the environment and transport efficiency. Therefore, intelligent transportation systems (ITS) are being developed to alleviate traffic congestion. An essential input for ITS is accurate and reliable knowledge about traffic conditions, but monitoring traffic conditions over large road networks has proven to be a significant challenge (Stathopoulos and Karlaftis, 2003; Hellinga *et al.*, 2008; Hofleitner *et al*. 2012a), in part due to unavoidable limitations (cost, limited coverage, etc.) of most existing technologies (loop detectors, radars and video cameras) for collecting traffic data. Fortunately, this situation has been changing due to the availability of a new information collection technique-the probe vehicle.

Fitted with a positioning device, a probe vehicle records information such as time, location and speed at a series of points along the route travelled. Unlike stationary sensors, probe vehicles can collect traffic data for any part of the network where they operate, and continuously collect data at any time that they are active (Jenelius and Koutsopoulos, 2013; Cao *et al*., 2013). As an advanced traffic information collection technology, probe vehicles have been experimented and studied for two decades (Van Aerde *et al.*, 1993; Miwa *et al.*, 2004). Especially in recent years researches based on probe vehicle data began to be widely carried out in both academe and industry. At the same time, the characteristics of probe vehicle data have been experiencing changes:

- Penetration. Probe vehicle data comes from three types of sources (Hofleitner *et al*. 2012): fleet data from dedicated vehicles (FedEx, UPS, taxis, etc.),

vehicle re-identification (RFID, magnetic signature) and participatory sensing (GPS enabled smartphone). The penetration of the former two is typically small (less than 1%) due to the limited types of vehicles and cost considerations; while the latter greatly enriches the probe vehicle data, and will certainly rise to a high level of penetration as a result of fast development of the mobile Internet (Hunter *et al.*, 2013; Bierlaire *et al.*, 2013; Bar-Gera, 2007).

- Polling frequency. Although it is easier to control the sampling schedule of fleet data, their polling frequencies are still low and uncertain most of the time due to cost considerations, the multiplicity of protocols (time-based and distance-based) or lost data. Participatory sensing and vehicle re-identification yield even more random, sparse and uncontrollable polling frequencies. Therefore, polling frequencies are often very low (less than one data point per minute) (Hofleitner *et al*. 2012a*,* Jenelius and Koutsopoulos, 2013; Hunter *et al.*, 2013; Lou *et al.*, 2009) and are variable for most probe vehicles (Miwa *et al.*, 2004).
- Accuracy. The accuracy of GPS positioning in civilian use has improved from 100 meters in the 1990s to 7.8 meters in 2008 (GPS.gov, 2013). The next-generation of GPS (GLONASS in Russia, Galileo in Europe, etc.) in the near future is expected deliver an accuracy of 1 meter, which makes it possible to accurately map-match recorded point onto its real location. In addition, Witte and Wilson (2004) found the recorded speed errors of roughly 5% on average, with the largest error at high speeds and when few GPS satellites are visible.

The aforementioned features of probe vehicle data, including merits of increasing penetration and improving accuracy but also demerits of the variety of data types,

lower polling frequency and randomness of the corresponding spatio-temporal coverage, make it challenging for fully characterizing macroscopic traffic model parameters and monitoring traffic conditions for large arterial network.

In light of these challenges, Hofleitner *et al.* used a well-suited statistical approach: traffic patterns are learned from past data, and then real-time data is fused with the learned patterns to identify the current state of traffic (Hofleitner *et al*. 2012a*)*. The core of this approach is modeling the observed variables with macroscopic traffic model parameters and link parameters. In Hofleitner *et al*. (2012a), Hofleitner *et al.* modeled the travel time distribution between any two points on a link. Soon later, in order to utilize both travel time data and location data from probe vehicles, they modeled the probability distribution of vehicle location and incorporated it into their approach (Hofleitner *et al*. 2012b). However, travel times from lower polling frequency probe data are less reliable due to the increasing difficulty of route identification (Hofleitner *et al*. 2012a; Hunter *et al.*, 2013; Zheng and Van Zuylen 2012). And they haven't made use of the more reliable information on speed. Therefore, this research focuses on modeling location and speed as observed variables, and expects to mine information from archived reports of location and speed instead of travel time from probe vehicles.

Extending a location distribution derived by Hofleitner *et al.* (Hofleitner *et al*. 2012b), we model the joint probability distribution function (PDF) of vehicle location and speed on an arterial road segment using hydrodynamic theory and horizontal queuing theory. In this model, we specifically consider the effects of signal controller (including deceleration and acceleration, queuing) on the distributions of location and speed on arterial roads. This probabilistic model is parameterized by link parameters (red signal time, cycle time, and critical flow density), driving behavior (average deceleration and acceleration, average travel speed and variation of arrival flow and

dissipating flow) and traffic state (arrival flow density) that are learned from historical probe location and speed data by maximum likelihood estimation.

This chapter is organized as follows. Firstly, the problem is briefly described. Secondly, traffic dynamics and model assumptions are explained. Thirdly, the PDF of vehicle location and speed is modeled for the general case, the under-saturated regime, and the saturated regime. Then, a method of estimating parameters is introduced. This is followed by validation of the proposed models using actual probe data. Finally, our conclusions and future work are elaborated.

6.2 Problem statement

For a signalized road segment of length of L, vehicles from upstream travel downstream across the intersection under the control of a traffic signal. Some vehicles stop at the stop line, forming a queue when the signal is red, and the queue then dissipates during the green signal time. The process of queue formation and dissipation is repeated cycle by cycle due to the recurrence of red and green signals. Probe vehicles are randomly distributed among the traffic, stochastically reporting at intervals with information about time, position and speed. After map-matching onto the road segment, a probe point can be denoted by a three-dimensional coordinate (t,x,v), where x represents the horizontal distance from the map-matched point to the stop line at the downstream intersection. We assume that probe points are recorded randomly in the spatial-temporal domain and are mutually independent, which is likely in line with the actual situation. Therefore, it is reasonable to take location x and speed v as random variables.

In principle vehicles travel freely and fast before they approach the stop line, and then more slowly before a queue, becoming stationary in the queue at the stop line. Therefore, vehicles generally spend less time per unit distance upstream in a road

segment than downstream. This means it is more likely for a reported point to be located downstream in a segment. Further, the instantaneous speed is possibly close to the free-flow speed upstream in a segment, while it ranges from zero to the free-flow speed at the downstream end due to deceleration, acceleration and stopping. An example distribution of problem point locations and speeds is drawn in Figure 6.1. The objective of this chapter is to derive the probability distribution function (PDF) $p(v,x)$ using macroscopic traffic model parameters and link parameters.

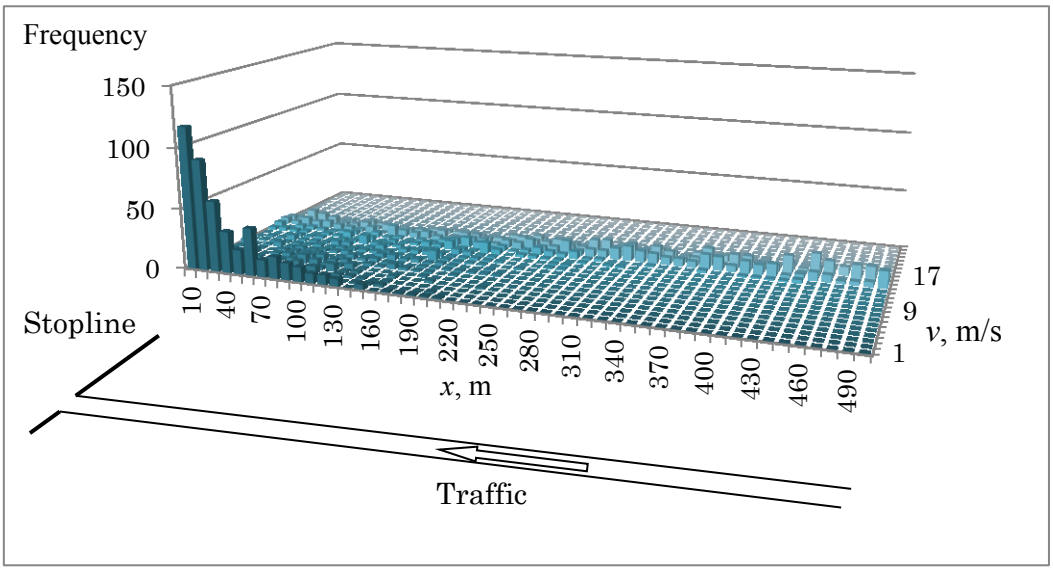


Figure 6.1 Histogram of sample probe points from simulation along a signalized road

6.3 Traffic dynamics and model assumptions

In traffic flow theory, there are three fundamental macroscopic variables of *flow* $q(x,t)$ (veh/s), *density* $\rho(x,t)$ (veh/m), and *speed* $v(x,t)$ (m/s), which represent traffic conditions in the spatial and temporal domains. The relation among these three variables is given (Lighthill and Whitham, 1955) as:

$$q(x,t) = \rho(x,t)v(x,t) \tag{6.1}$$

In arterial networks, traffic is driven by the formation and the dissipation of queues at intersections. The dynamics of queues are characterized by shocks, which form at the interfaces of traffic flows with different densities (Lighthill and Whitham,

1955). Before detailing our model derivation, we make the following assumptions on the dynamics of traffic flow:

1) Stable traffic. During each time interval of interest, signal timings (red light time R and cycle time C) and the arrival traffic density ρ_a are constant. With these assumptions, traffic dynamics are stable and periodic with period C in the estimation interval.
2) Deceleration and acceleration are considered. A vehicle enters a queue with deceleration a before it coming to a halt. It then dissipates from the queue with acceleration b. a and b are assumed constant for all vehicles in the estimation interval. This treatment of deceleration and acceleration is closer to reality than ignoring them in most related studies on arterials (Hofleitner *et al*. 2012a; Bar-Gera, 2007; Ban *et al.*, 2011; Mehran *et al.*, 2012; Hofleitner *et al.*, 2011; Liu *et al.*, 2009; Ban *et al.*, 2009; Skabardonis and Geroliminis, 2005; Hao *et al.*, 2012)
3) No spillover. A queue will not extend beyond the link length, which means that traffic entering a link is not affected by a queue. Actually, as long as the road is not heavily congested or the link length is very short, spillover will not occur.
4) Differences in driving behavior are modeled. The free flow speed is not the same for all vehicles. The arrival free flow speed and the dissipating free flow speed are modeled by normal distributions $N(\mu_a, \sigma_a)$ and $N(\mu_d, \sigma_d)$, respectively.

Let u denote the speed at which a queue forms, w denote the speed of queue dissipation, and Rankine-Hugoniot jump conditions (Evans, 1998) are used to express them as

$$u = \frac{\rho_a v_f}{\rho_{max} - \rho_a} \text{ and } w = \frac{\rho_c v_f}{\rho_{max} - \rho_c} \quad (6.2)$$

The patterns of traffic in such a network can be divided into the under-saturated regime and the congested regime, depending on whether a queue remains when the traffic signal changes from green to red. We model both of these in the following section.

6.4 Modeling the joint PDF model of vehicle location and speed

We first derive the joint probability distribution of vehicle locations and speeds for the general case without considering the traffic regime. The derivation shows our general approach, which is then applied to the under-saturated and congested traffic regimes.

6.4.1 General case

In probability theory, the joint distribution of location and speed can be derived using the multiplication rule:

$$p(v, x) = p(x)p(v|x) \quad (6.3)$$

where $p(x)$ is the PDF of location, and $p(v|x)$ is the PDF of speed given location x. As long as $p(x)$ and $p(v|x)$ are obtained, $p(v, x)$ can be derived using Equation (6.3).

The derivation of $p(x)$

Under the assumption of stable traffic, the traffic density at location x is time periodic with period *C* (Hofleitner *et al*. 2012b). Thus, the average density $d(x)$ at location x in one cycle time is:

$$d(x) = \frac{1}{C}\int_0^C \rho(x, t)dt \quad (6.4)$$

According to the model assumptions, the density $\rho(x, t)$ at location x and time t might be in one of five situations, numbered 1 to 5 for convenience: (1), $\rho_1 = \rho_a$, when vehicles arrive at the link and are not influenced by the queue; (2), $\rho_2 = \rho_{de}$,

when vehicles are decelerating for the queue; (3), $\rho_3 = \rho_{max}$, when vehicles are stationary in the queue. (4), $\rho_4 = \rho_{ac}$, when vehicles are accelerating from the queue; and (5), $\rho_5 = \rho_c$, when vehicles are dissipating from the queue and traveling with free flow speed. Thus, the average density at location x is

$$d(x) = \sum_{i=1}^{5} \alpha_i(x)\rho_i \tag{6.5}$$

where $\alpha_i(x)$ represents the fraction of time that the density is equal to ρ_i at location x, and is calculated as

$$\alpha_i(x) = \frac{t_i(x)}{C} \tag{6.6}$$

where $t_i(x)$ is the duration for which the density is equal to ρ_i at location x in one cycle.

Since probe vehicles record vehicle locations randomly, more densely populated areas of the link will have more location measurements. At location x, the PDF $p(x)$ of vehicle location is proportional to the average density $d(x)$. Using Equations (6.5) and (6.6), we have

$$p(x) = \frac{d(x)}{\int_0^L d(x)dx} = \frac{\sum_{i=1}^{5} t_i(x)\rho_i}{\int_0^L \sum_{i=1}^{5} t_i(x)\rho_i dx} \tag{6.7}$$

In the under-saturated and congested regimes, the computation of $t_i(x)$, $i = 1, \ldots, 5$ enables the derivation of the probability distribution of vehicle locations.

The derivation of $p(v|x)$

As the above statement, there are five possible traffic situations with different densities at location x. Assume the speed distribution of a traffic situation with density ρ_i is $g_i(v)$, then the probability distribution of speed given location x is a mixed distribution with five components. Thus, $p(v|x)$ is obtained as

$$p(v|x) = \sum_{i=1}^{5} \beta_i(x) g_i(v) \tag{6.8}$$

where $\beta_i(x)$ is the fraction of the number of sample points with density is equal to ρ_i at location x, and is calculated as

$$\beta_i(x) = \frac{n_i(x)}{\sum_{i=1}^{5} n_i(x)} \tag{6.9}$$

where $n_i(x)$ is the number of sample points that density is equal to ρ_i.

Consequently, we can derive the expression of $n_i(x)$ to obtain $\beta_i(x)$. Consider a tiny area $[x, x+\Delta x]$ around location x, the velocity of vehicle is v_i when the density is equal to ρ_i, the during time that density is equal to ρ_i at location x in one cycle is $t_i(x)$.Then, the accumulated number of vehicles travelling across the tiny area is $A_i = q_i t_i(x) = \rho_i v_i t_i(x)$. The total vehicle time traveled (VTT) is $VTT_i = A_i \frac{\Delta x}{v_i} = \rho_i t_i(x) \Delta x$. Assume the average polling rate is δ $(\text{time} \cdot \text{vehicle/point})$, then the number of sample points at location x is:

$$n_i(x) = \frac{VTT_i}{\delta} = \frac{\rho_i t_i(x) \Delta x}{\delta} \tag{6.10}$$

Then,

$$\beta_i(x) = \frac{\rho_i t_i(x)}{\sum_{i=1}^{5} \rho_i t_i(x)} \tag{6.11}$$

This method of deriving $\beta_i(x)$ will be applied to derive the speed distribution in deceleration and acceleration areas in the next section.

Using Equation (6.8, 6.11), the PDF of speed given location x is

$$p(v|x) = \frac{\sum_{i=1}^{5} \rho_i t_i(x) g_i(v)}{\sum_{i=1}^{5} \rho_i t_i(x)} \tag{6.12}$$

Finally, using Equations (6.3, 6.7, and 6.12), the joint probability distribution of vehicle location and speed in the general case is

$$p(v, x) = \frac{\sum_{i=1}^{5} \rho_i t_i(x) g_i(v)}{\int_0^L \sum_{i=1}^{5} t_i(x) \rho_i dx} \tag{6.13}$$

In the under-saturated and congested regimes, the computation of $t_i(x)$, ρ_i and $g_i(v), i = 1, \ldots, 5$ enables the derivation of the joint probability distribution of vehicle location and speed. Specifically, in situation 1, vehicles travel with arrival density ρ_a without influence of queues, and the speed distribution $g_1(v)$ is assumed to be normal $N(\mu_a, \sigma_a)$ to model the differences in driving behavior; in situation 3,

vehicles are stationary in a queue with a maximum density ρ_{max}, and the speed distribution $g_3(v)$ is a mass distribution in 0 $Dir_{\{0\}}(v)$; in situation 5, vehicles dissipate from the queue and travel with critical density ρ_c (when the flow rate is maximum), and the speed distribution $g_5(v)$ is assumed to be normal $N(\mu_d, \sigma_d)$ to model the differences in driving behavior. In situation 2 (deceleration) and situation 4 (acceleration), traffic densities vary with time. We should derive the average density ρ_{de} and speed distribution $g_{de}(v)$ for situation 2, and the average density ρ_{ac} and speed distribution $g_{ac}(v)$ for situation 4 in both the under-saturated and congested regimes.

6.4.2 Under-saturated regime

In the under-saturated regime, a queue that forms at a red signal is fully cleared during the green period of the same cycle. Figure 6.2 shows a space-time diagram of vehicle trajectories for one cycle in an under-saturated traffic regime for an arterial road segment. This diagram is divided into five regions for different traffic densities and patterns. As we can see, the first vehicle arriving in a particular cycle begins to decelerate before the light changes to red and stops at the stop line when the red light have last $v_f/2a$ seconds. Therefore, the maximum time that a vehicle is stationary in a queue is $R - v_f/2a,$ instead of $R.$ The queue forms at a rate of u and dissipates at a rate of $w.$ As described in the previous section, the five regions of the diagram represent arriving traffic, a deceleration area, the queue, an acceleration area and dissipating traffic. The deceleration area links the arriving traffic to the queue, while the acceleration area links the queue with the dissipating traffic.

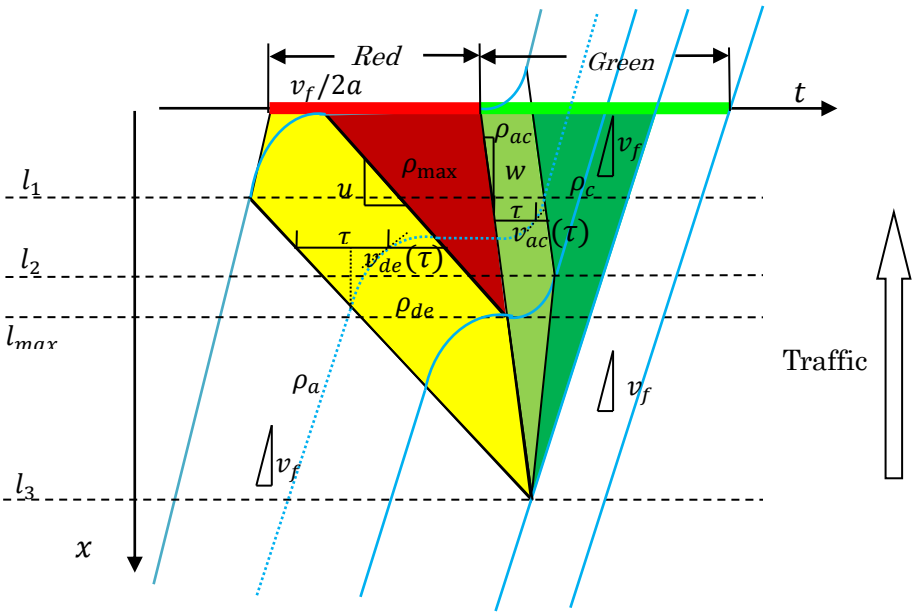


Figure 6.2 Space time diagram of vehicle trajectories in under-saturated traffic regime

In the deceleration area, the speed of a vehicle falls from the free speed to zero and traffic density changes with time from ρ_a to ρ_{max}. Similarly, in the acceleration area, vehicle speed increases from zero to the free speed and traffic density changes with time from ρ_{max} to ρ_c.

In order to derive the average density ρ_{de} and ρ_{ac}, and the speed distribution $g_{de}(v)$ and $g_{ac}(v)$, in these two areas, we first derive an expression for the density function $\rho_{de}(\tau)$ and $\rho_{ac}(\tau)$, where τ represents the duration of the deceleration area or acceleration area at location x (Figure 6.2).

Taking the deceleration area that x is located in $[l_1, l_2]$ as an example (Figure 6.2), we can easily obtain an expression of $v_{de}(\tau)$ using the basic mathematical relationship of distance, speed and acceleration. It is given as

$$\frac{\frac{v_f^2 - v_{de}^2(\tau)}{2a}}{u} + \frac{v_f - v_{de}(\tau)}{a} = \tau \tag{6.14}$$

Solving Equation (14), we obtain

$$v_{de}(\tau) = -u + \sqrt{(u + v_f)^2 - 2au\tau} \tag{6.15}$$

Using the commonly used density-speed function proposed by Greenshields,

$$\rho = \rho_{max}\left(1 - \frac{v}{v_f}\right) \tag{6.16}$$

And denoting $K_{de}(\tau)$ as

$$K_{de}(\tau) = 1 + \frac{u}{v_f} - \sqrt{\left(\frac{u}{v_f} + 1\right)^2 - \frac{2au\tau}{v_f^2}} \tag{6.17}$$

We then get the expression of $\rho_{de}(\tau)$ as

$$\rho_{de}(\tau) = \rho_{max} K_{de}(\tau) \tag{6.18}$$

If the maximum duration of the deceleration area at location x is $T_{de}(x)$, then the average density ρ_{de} is

$$\rho_{de} = \frac{1}{T_{de}(x)} \int_0^{T_{de}(x)} \rho_{max} K_{de}(\tau) d\tau \tag{6.19}$$

Now we describe the derivation of speed distribution $g_{de}(v)$. At location x, assume a tiny area $[v, v + \Delta v]$, within which the ratio of sample points is $r_{[v,v+\Delta v]}$. Then $g_{de}(v)$ is given as

$$g_{de}(v) = \frac{r_{[v,v+\Delta v]}}{\Delta v} \tag{6.20}$$

Actually, the meaning of $r_{[v,v+\Delta v]}$ is similar to $\beta_i(x)$ in Equation (6.9), thus using Equation (11), we have

$$r_{[v,v+\Delta v]} = \frac{\rho_{de}(\tau)\Delta\tau}{\int_0^{T_{de}(x)} \rho_{de}(\tau) d\tau} \tag{6.21}$$

Using Equation (6.14) and replacing $v_{de}(\tau)$ with v for simplicity, we can calculate

$$\Delta\tau = |\tau(v + \Delta v) - \tau(v)| = \frac{(2u + 2v + \Delta v)\Delta v}{2au} \tag{6.22}$$

Finally, using Equations (6.16, 6.18, 6.20 and 6.22), we obtain

$$g_{de}(v) = \frac{(1 - \frac{v}{v_f})\frac{v+u}{au}}{\int_0^{T_{de}(x)} K_{de}(\tau) d\tau} \tag{6.23}$$

The above derivation process can also be applied in the case of acceleration area. We list the expressions of interest as follows:

$$v_{ac}(\tau) = -w + \sqrt{w^2 + 2bw\tau} \tag{6.24}$$

$$\rho_{ac}(\tau) = \rho_{max} K_{ac}(\tau) \tag{6.25}$$

$$\rho_{ac} = \frac{1}{T_{ac}(x)} \int_0^{T_{ac}(x)} \rho_{max} K_{ac}(\tau) d\tau \tag{6.26}$$

$$g_{ac}(v) = \frac{(1-\frac{v}{v_f})\frac{v+w}{bw}}{\int_0^{T_{ac}(x)} K_{ac}(\tau) d\tau} \tag{6.27}$$

$$K_{ac}(\tau) = 1 + \frac{w}{v_f} - \sqrt{\left(\frac{w}{v_f}\right)^2 + \frac{2bw\tau}{v_f^2}} \tag{6.28}$$

At this point, we have derived ρ_{de}, ρ_{ac} $g_{de}(v)$ and $g_{ac}(v)$. In Equation (6.13), we still need to compute $t_i(x)$ $i = 1, \dots, 5$, and then the joint PDF of vehicle location and speed can be obtained. From Figure 6.2, we can see that the expression of $t_i(x)$ might differ at different location intervals. Therefore, we first compute the location turning points, and then derive $t_i(x)$ for the locations between them.

As in Figure 6.2, we can easily compute l_1, l_2, l_{max} (the maximum queue length) and l_3 using the laws of motion and geometry. They are

$$l_1 = v_f^2 / 2a \tag{6.29}$$

$$l_2 = l_{max} - v_f^2 / 2b \tag{6.30}$$

$$l_{max} = (R - \frac{v_f}{2a}) \frac{wu}{w-u} \tag{6.31}$$

$$l_3 = \frac{w}{w-u} \left(\frac{v_f(u+v_f)}{2a} + uR \right) \tag{6.32}$$

We omit the details of computing $t_i(x)$ $i = 1, \dots, 5$, since this is trivial using the above formulas and the laws of motion. We thus summarize the derived PDF of vehicle location and speed in the under-saturated regime as follows,

$$p^{un}(v, x) = \frac{\sum_{i=1}^{5} \rho_i t_i(x) g_i(v)}{\int_0^L \sum_{i=1}^{5} t_i(x) \rho_i dx} \tag{6.33}$$

where,

$$\rho_1 = \rho_a, \ \rho_2 = \frac{1}{T_{de}(x)} \int_0^{T_{de}(x)} \rho_{max} K_{de}(\tau) d\tau, \ \rho_3 = \rho_{max},$$

$\rho_4 = \frac{1}{T_{ac}(x)} \int_0^{T_{ac}(x)} \rho_{max} K_{ac}(\tau) d\tau,\ \rho_5 = \rho_c;$

$g_1(v) = N(\mu_a, \sigma_a),\ \ g_2(v) = \frac{(1-\frac{v}{v_f})^{\frac{v+u}{au}}}{\int_0^{T_{de}(x)} K_{de}(\tau) d\tau},\ g_3(v) = Dir_{\{0\}}(v),$

$g_4(v) = \frac{(1-\frac{v}{v_f})^{\frac{v+w}{bw}}}{\int_0^{T_{ac}(x)} K_{ac}(\tau) d\tau},\ g_5(v) = N(\mu_d, \sigma_d);$

$K_{de}(\tau) = 1 + \frac{u}{v_f} - \sqrt{\left(\frac{u}{v_f} + 1\right)^2 - \frac{2au\tau}{v_f^2}},\ K_{ac}(\tau) = 1 + \frac{w}{v_f} - \sqrt{\left(\frac{w}{v_f}\right)^2 + \frac{2bw\tau}{v_f^2}};$

$T_{de}(x) = t_2(x),\ T_{ac}(x) = t_4(x);$

if $x \in [0, l_1]$,

$t_1(x) = C - \sum_{i=2}^5 t_i(x),\ t_2(x) = \left(\frac{1}{v_f} + \frac{1}{u}\right)x + \frac{v_f}{2a},\ t_3(x) = \left(R - \frac{v_f}{2a}\right)\left(1 - \frac{x}{l_{max}}\right),$

$t_4(x) = \frac{v_f}{b}\left(\frac{v_f}{2w} + 1\right),\ t_5(x) = (l_3 - x)\left(\frac{1}{v_f} + \frac{1}{w}\right) - \frac{v_f}{b}\left(\frac{v_f}{2w} + 1\right);$

if $x \in [l_1, l_2]$,

$t_1(x) = C - \sum_{i=2}^5 t_i(x)\ ,\ t_2(x) = \frac{v_f}{a}\left(\frac{v_f}{2u} + 1\right)\ ,\ \ t_3(x) = \left(R - \frac{v_f}{2a}\right)\left(1 - \frac{x}{l_{max}}\right)\ ,$

$t_4(x) = \frac{v_f}{b}\left(\frac{v_f}{2w} + 1\right),\ t_5(x) = (l_3 - x)\left(\frac{1}{v_f} + \frac{1}{w}\right) - \frac{v_f}{b}\left(\frac{v_f}{2w} + 1\right);$

if $x \in [l_2, l_{max}]$,

$t_1(x) = C - \sum_{i=2}^5 t_i(x)\ ,\ \ t_2(x) = \frac{v_f}{a}\left(\frac{v_f}{2u} + 1\right)\ ,\ \ t_3(x) = \left(R - \frac{v_f}{2a}\right)\left(1 - \frac{x}{l_{max}}\right)\ ,$

$t_4(x) = \frac{v_f}{b}\left(\frac{v_f}{2w} + 1\right)\frac{l_3 - x}{l_3 - l_2},\ t_5(x) = (l_3 - x)\left(\frac{1}{v_f} + \frac{1}{w} - \frac{v_f}{b(l_3 - l_2)}\left(\frac{v_f}{2w} + 1\right)\right);$

if $x \in [l_{max}, l_3]$,

$t_1(x) = C - \sum_{i=2}^5 t_i(x)\ \ ,\ t_2(x) = R + \left(\frac{1}{w} - \frac{1}{u}\right)x + \frac{v_f}{2a}\left(\frac{v_f}{u} + 1\right)\ \ ,\ \ \ t_3(x) = 0\ \ ,$

$t_4(x) = \frac{v_f}{b}\left(\frac{v_f}{2w} + 1\right)\frac{l_3 - x}{l_3 - l_2},\ t_5(x) = (l_3 - x)\left(\frac{1}{v_f} + \frac{1}{w} - \frac{v_f}{b(l_3 - l_2)}\left(\frac{v_f}{2w} + 1\right)\right);$

if $x \in [l_3, L]$,

$t_1(x) = C,\ t_2(x) = t_3(x) = t_4(x) = t_4(x) = 0.$

6.4.3 Congested regime

In the congested regime, there is a remaining queue l_r that is not cleared during the

green signal time in the current cycle. Vehicles in this queue are stuck ahead of the stop line and form the front of a new queue in the next cycle. As shown in Figure 6.3, the space-time diagram is again divided into five regions, as in the under-saturated regime (shown in Figure 6.2). In this diagram, the area of $[l_2, L]$ is almost the same as the area of $[l_1, L]$ in Figure 6.2. However, the deceleration area in $[0, l_2]$ is different from the case of under-saturated regime, because queue formation rate is w instead of u. Therefore, we should derive the expression of $t_2(x)$, ρ_{de} and $g_{de}(v)$ for this area. The details of derivation are neglected, since they are similar to the case of the under-saturated regime. Firstly we list the calculation of l_1, l_2, l_3 l_{max} and l_4 in Figure 6.3:

$$l_1 = v_f^2 /2a \tag{6.34}$$

$$l_2 = l_r + v_f^2 /2a \tag{6.35}$$

$$l_3 = \left(R - \frac{v_f}{2a}\right)\frac{wu}{w-u} + l_r - \frac{v_f^2}{2b} \tag{6.36}$$

$$l_{max} = \left(R - \frac{v_f}{2a}\right)\frac{wu}{w-u} + l_r \tag{6.37}$$

$$l_4 = l_r + \frac{w}{w-u}\left(\frac{v_f(u+v_f)}{2a} + uR\right) \tag{6.38}$$

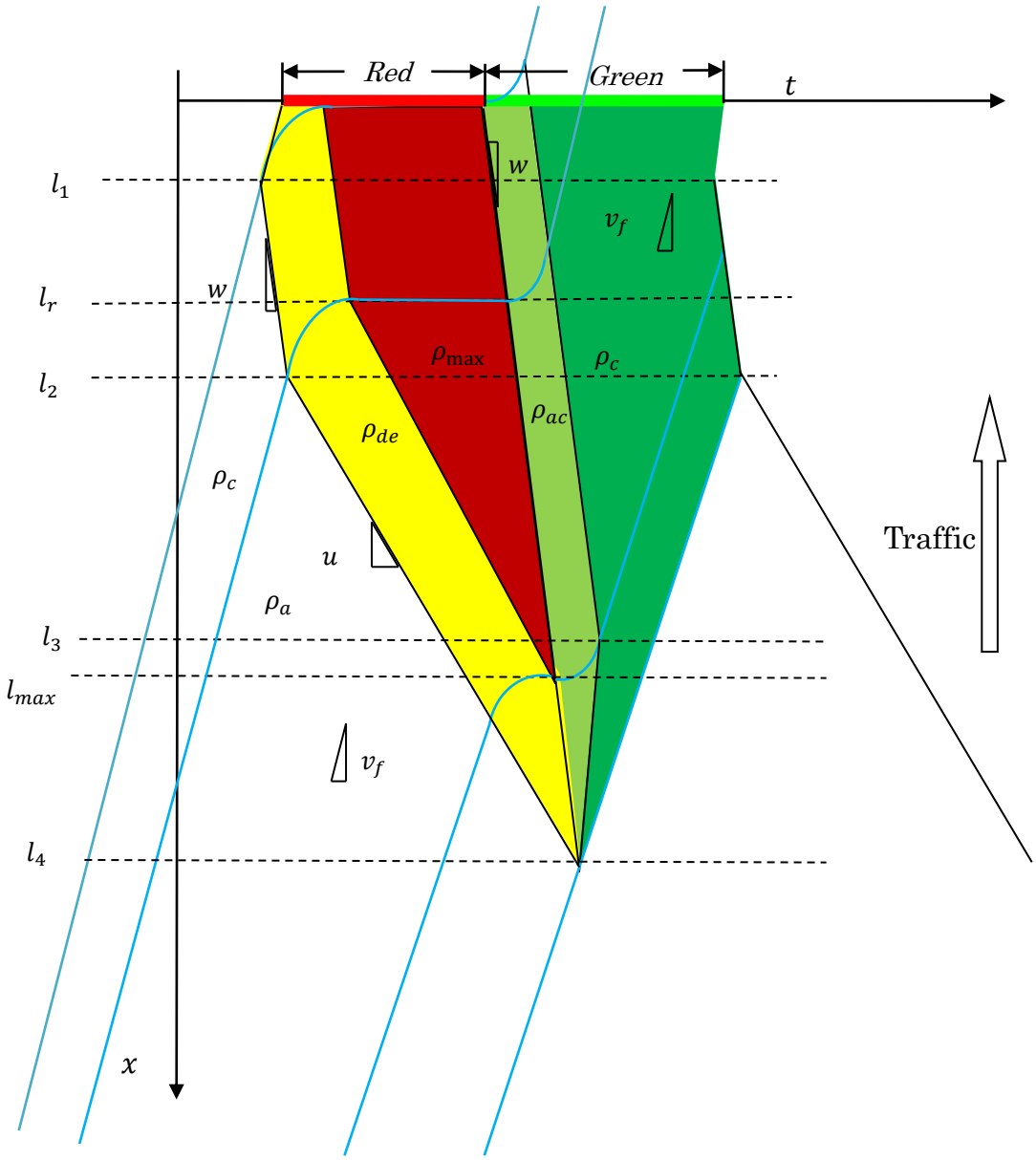


Figure 6.3 Space-time diagram of vehicle trajectories in congested traffic regime

The PDF of vehicle location and speed in the congested regime is summarized as follows:

$$p^{con}(v,x) = \frac{\sum_{i=1}^{5} \rho_i t_i(x) g_i(v)}{\int_0^L \sum_{i=1}^{5} t_i(x) \rho_i dx} \tag{6.39}$$

where

$$\rho_1 = \rho_a,\ \rho_3 = \rho_{max},\ \rho_4 = \frac{1}{T_{ac}(x)} \int_0^{T_{ac}(x)} \rho_{max} K_{ac}(\tau) d\tau,\ \rho_5 = \rho_c;$$

$g_1(v) = N(\mu_a, \sigma_a),\ g_3(v) = Dir_{\{0\}}(v),\ g_4(v) = \frac{(1-\frac{v}{v_f})^{\frac{v+w}{bw}}}{\int_0^{T_{ac}(x)} K_{ac}(\tau)d\tau}$,

$g_5(v) = N(\mu_d, \sigma_d)$;

$K_{de}^u(\tau) = 1 + \frac{u}{v_f} - \sqrt{\left(\frac{u}{v_f}+1\right)^2 - \frac{2au\tau}{v_f^2}}$, $K_{de}^w(\tau) = 1 + \frac{w}{v_f} - \sqrt{\left(\frac{w}{v_f}+1\right)^2 - \frac{2aw\tau}{v_f^2}}$,

$K_{ac}(\tau) = 1 + \frac{w}{v_f} - \sqrt{\left(\frac{w}{v_f}\right)^2 + \frac{2bw\tau}{v_f^2}}$;

$T_{de}(x) = t_2(x),\ T_{ac}(x) = t_4(x)$;

if $x \in [0, l_1]$,

$\rho_2 = \frac{1}{T_{de}(x)} \int_0^{T_{de}(x)} \rho_{max} K_{de}^w(\tau)d\tau,\ g_2(v) = \frac{(1-\frac{v}{v_f})^{\frac{v+w}{aw}}}{\int_0^{T_{de}(x)} K_{de}^w(\tau)d\tau}$,

$t_1(x) = 0,\ t_2(x) = \left(\frac{1}{v_f} + \frac{1}{w}\right)x + \frac{v_f}{2a},\ t_3(x) = \left(R - \frac{v_f}{2a}\right),\ t_4(x) = \frac{v_f}{b}\left(\frac{v_f}{2w}+1\right)$,

$t_5(x) = C - \sum_{i=1}^4 t_i(x)$;

if $x \in [l_1, l_r]$,

$\rho_2 = \frac{1}{T_{de}(x)} \int_0^{T_{de}(x)} \rho_{max} K_{de}^w(\tau)d\tau,\ g_2(v) = \frac{(1-\frac{v}{v_f})^{\frac{v+w}{aw}}}{\int_0^{T_{de}(x)} K_{de}^w(\tau)d\tau}$,

$t_1(x) = 0\ ,\ \ t_2(x) = \frac{v_f}{a}\left(\frac{v_f}{2w}+1\right)\ ,\ \ t_3(x) = \left(R - \frac{v_f}{2a}\right)\ ,\ \ t_4(x) = \frac{v_f}{b}\left(\frac{v_f}{2w}+1\right)\ ,$

$t_5(x) = C - \sum_{i=1}^4 t_i(x)$

if $x \in [l_r, l_2]$,

$\rho_2 = \frac{1}{T_{de}(x)} \int_{t=0}^{T_{de}(x)} \rho_{max} K_{de}^w(\tau)d\tau,\ g_2(v) = \frac{(1-\frac{v}{v_f})^{\frac{v+w}{aw}}}{\int_0^{T_{de}(x)} K_{de}^w(\tau)d\tau}$,

$t_1(x) = 0\quad ,\quad t_2(x) = (x - l_r)\left(\frac{1}{u} - \frac{1}{w}\right) + \frac{v_f}{a}\left(\frac{v_f}{2w}+1\right)\quad ,\quad t_3(x) = R - \frac{v_f}{2a} -$

$(x - l_r)\left(\frac{1}{u} - \frac{1}{w}\right),\ t_4(x) = \frac{v_f}{b}\left(\frac{v_f}{2w}+1\right),\ t_5(x) = C - \sum_{i=1}^4 t_i(x)$;

if $x \in [l_2, l_3]$,

$\rho_2 = \frac{1}{T_{de}(x)} \int_0^{T_{de}(x)} \rho_{max} K_{de}^u(\tau)d\tau,\ g_2(v) = \frac{(1-\frac{v}{v_f})^{\frac{v+u}{au}}}{\int_0^{T_{de}(x)} K_{de}^u(\tau)d\tau}$,

$t_1(x) = \left(\frac{1}{v_f} + \frac{1}{u}\right)(x - l_2)\quad ,\quad t_2(x) = \frac{v_f}{a}\left(\frac{v_f}{2u}+1\right)\quad ,\quad t_3(x) = R - \frac{v_f}{2a} - (x -$

$l_r)\left(\frac{1}{u}-\frac{1}{w}\right),\ t_4(x)=\frac{v_f}{b}\left(\frac{v_f}{2w}+1\right),\ t_5(x)=C-\sum_{i=1}^{4}t_i(x);$

if $x\in[l_3,l_{max}]$

$\rho_2=\frac{1}{T_{de}(x)}\int_0^{T_{de}(x)}\rho_{max}K_{de}^{u}(\tau)d\tau,\ g_2(v)=\frac{(1-\frac{v}{v_f})^{\frac{v+u}{au}}}{\int_0^{T_{de}(x)}K_{de}^{u}(\tau)d\tau},$

$t_1(x)=\left(\frac{1}{v_f}+\frac{1}{u}\right)(x-l_2)\quad,\quad t_2(x)=\frac{v_f}{a}\left(\frac{v_f}{2u}+1\right)\quad,\quad t_3(x)=R-\frac{v_f}{2a}-(x-$

$l_r)\left(\frac{1}{u}-\frac{1}{w}\right),\ t_4(x)=\frac{v_f}{b}\left(\frac{v_f}{2w}+1\right)\frac{l_4-x}{l_4-l_3},\ t_5(x)=C-\sum_{i=1}^{4}t_i(x);$

if $x\in[l_{max},l_4]$

$\rho_2=\frac{1}{T_{de}(x)}\int_0^{T_{de}(x)}\rho_{max}K_{de}^{u}(\tau)d\tau,\ g_2(v)=\frac{(1-\frac{v}{v_f})^{\frac{v+u}{au}}}{\int_0^{T_{de}(x)}K_{de}^{u}(\tau)d\tau},$

$t_1(x)=\left(\frac{1}{v_f}+\frac{1}{u}\right)(x-l_2)\quad,\quad t_2(x)=R+\left(\frac{1}{w}-\frac{1}{u}\right)(x-l_r)+\frac{v_f}{2a}\left(\frac{v_f}{u}+1\right)\quad,$

$t_3(x)=0,\ t_4(x)=\frac{v_f}{b}\left(\frac{v_f}{2w}+1\right)\frac{l_4-x}{l_4-l_3},\ t_5(x)=C-\sum_{i=1}^{4}t_i(x);$

if $x\in[l_4,L]$,

$t_1(x)=C,\ t_2(x)=t_3(x)=t_4(x)=t_4(x)=0.$

6.5 Learning model parameters from probe vehicle data

From traffic flow theory, we have derived joint probability distributions of vehicle location and speed on arterial roads for both under-saturated and congested regimes. These distributions are parameterized by network parameters (red signal time *R*, cycle time *C,* critical flow density ρ_c), driving behavior θ_{dr} ($\theta_{dr}=(a,b,\mu_a,\ \sigma_a,\mu_d,\ \sigma_d)$, deceleration rate *a* and acceleration rate *b*, average travel speed μ_a and variation σ_a of arrival flow and average travel speed μ_d and variation σ_d of arrival flow dissipation flow) and traffic state represented by arrival flow density ρ_a. We learn these parameters using reports of location and speed from probe vehicles.

In practice, the parameters above might change along with traffic conditions during the day. However, it is reasonable to assume constant parameters for certain time intervals of interest (for example, 7:00-9:00am, 3:00-5:00pm). During the

estimation time interval, let (x_j, v_j) $(j = 1, \dots, J,\ J$ is the number of reported probe points) represent the set of vehicle locations and speeds on the link of interest. We can learn the parameters by maximizing the likelihood (or more conveniently the log-likelihood) of the locations and speeds on this link with respect to these parameters. The estimation problem is given by:

$$\min_{R,C,\rho_c,\theta_{dr},\rho_a} \sum_{j=1}^{J} -ln(p(v_j, x_j)) \tag{6.40}$$

$$s.t. \begin{cases} R < C \\ \rho_a \leq \rho_c < \rho_{max} \\ u \leq w \end{cases}$$

Additional constraints and bounds may be added to limit the feasible set to physically acceptable values of the parameters and improve the estimation when little data is available. The model (6.40) is not convex but it is a small-scale optimization problem. Numerous optimization techniques can be used to solve this problem, including global optimization algorithms. Moreover, since the parameters should be physically meaningful, they can be bounded to limit the feasible set to a compact set. It is thus possible to do a grid search. The grid search defines a grid on the feasible set and evaluates the objective function for each set of parameters defined by the grid.

6.6 Data and results

Figure 6.4 The test arterial road in Toyota city, Japan

To validate the distributions of vehicle location and speed, we use data collected from private cars during the Green Project operated in Toyota City, Japan. More than 200 drivers participated in this exercise, which took place over six months from April 1 to September 30, 2011. On-board GPS equipment was installed in private cars and recorded points with time, location and speed every 1s. We filter the data to extract points recorded during the hours 3:00-5:00pm and 7:00-8:00am on weekdays for the test link (Figure 6.4). The traffic in 3:00-5:00pm is in the under-saturated regime, and the traffic in 7:00-8:00am is in the congested regime. The test link is an arterial road in the downtown area of Toyota City. It has two lanes and is 220 meters long, with traffic moving from east to west. For the test link, we compute the maximum likelihood estimates of the traffic parameters using the probe locations and speeds (Equation 6.40).

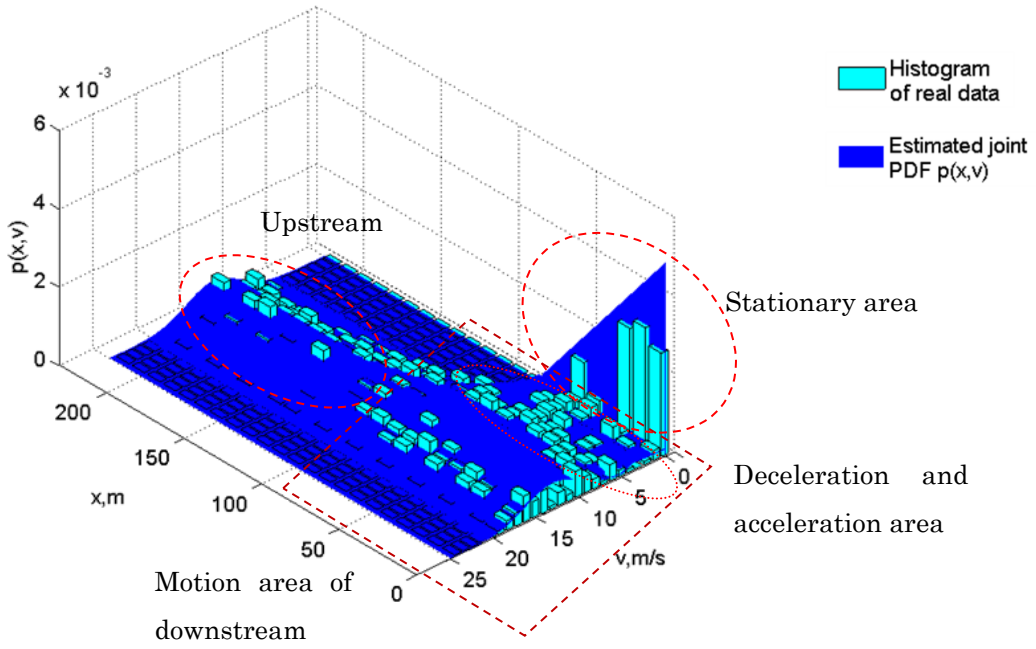


Figure 6.5 Histogram of data and estimated distribution for the under-saturated regime

The estimated joint PDF for under-saturated regime is showed in Figure 6.5. The estimated joint PDF fits real data well in upstream and motion area of downstream, but the fitting is not so good in stationary area of downstream. In particular, the

distribution characteristics in deceleration and acceleration area are captured by the joint PDF, which confirms the considerations of deceleration and acceleration in our model. On the whole, the proposed model captures most features of vehicle location and speed distributions.

In order to validate the joint PDF statistically, we perform the Kolmogrov-Smirnov (K-S) test (Massey, 1951). Because the derived distribution $p(v,x)$ is a two-dimensional joint distribution with random variables x and v, we validate its marginal distributions with respect to x and v instead of validating it directly. Thus, we derive the following marginal distributions:

$$f_x(x) = \int_0^{+\infty} p(v,x)dv \tag{6.41}$$

$$f_v(v) = \int_0^{L} p(v,x)dx \tag{6.42}$$

Actually, the marginal distribution of x is the location distribution $p(x)$ in Equation (6.7). Similarly, the marginal distribution $f_v(v)$ of v is denoted as $p(v)$. We test the hypothesis H_0: *the vehicle locations are distributed according to the distribution given by PDF* $p(x)$ *and the vehicle speeds are distributed according to the distribution given by PDF* $p(v)$ using the K-S test. The K-S test is a standard non-parametric test for stating whether samples are distributed according to a hypothetical distribution. The test is based on the K-S statistics which are computed as the maximum difference between the empirical and hypothetical cumulative distributions. The test provides a p-value, which informs the goodness of fit. Low p-values indicate that the data does not follow the hypothetical distribution. We reject hypothesis H_0 for p-values inferior to the significance level α. We set the parameter α 0.01 as common.

The test results validate the distributions of vehicle locations and speeds proposed in this chapter for both under-saturated and congested regimes (See Figure 6.6 and 6.7).

The p-value for the K-S test of vehicle location distribution is 0.1965 and 0.0833, while the p-value for the K-S test of vehicle speed distribution is 0.1574 and 0.0684, for under-saturated and congested regimes. All pass the test of $\alpha = 0.01$. Therefore, we accept the hypothesis H_0 with high confidence $1 - \alpha = 0.99$ and reach the conclusion that vehicle locations and speeds are distributed according to the proposed joint distribution. However, the p-values for both vehicle location and speed distributions indicate that the fitting in congested regime is worse than that in under-saturated regime. This is because traffic in congested regime is much unstable and its state may change from one to another during one hour (7:00-8:00am).

More observations can be extracted from Figure 6.6 and 6.7. The histograms of vehicle locations show that a reported point is more likely to be located downstream rather than upstream and that upstream points are uniformly distributed on the road. These features are all captured by the derived PDF $p(x)$. The difference between under-saturated regime and congested regime lie in a longer downstream for the latter. This can also be observed in Figure 6.2 and 6.3. The histograms of vehicle speeds show there are two peaks of speed: one peak is at a speed of 0 (vehicles stationary in the queue) and the other is at free-flow speed (vehicles traveling freely before or after the queue). These features are all captured by the derived PDF $p(v)$, which is a mixture of the five components: distribution with mass distribution at 0, distribution in deceleration area, distribution in acceleration area, distribution approaching the queue and distribution dissipating the queue. However, similar with observation in Figure 6.5, 6.6 and 6.7 also show that the estimated distributions perform worse downstream (in the queuing area) than in upstream. This is because of the assumption of constant arrival flow density. Due to light synchronization, some links have arrivals with platoons with different flow densities (Bails, *et al.* 2012).

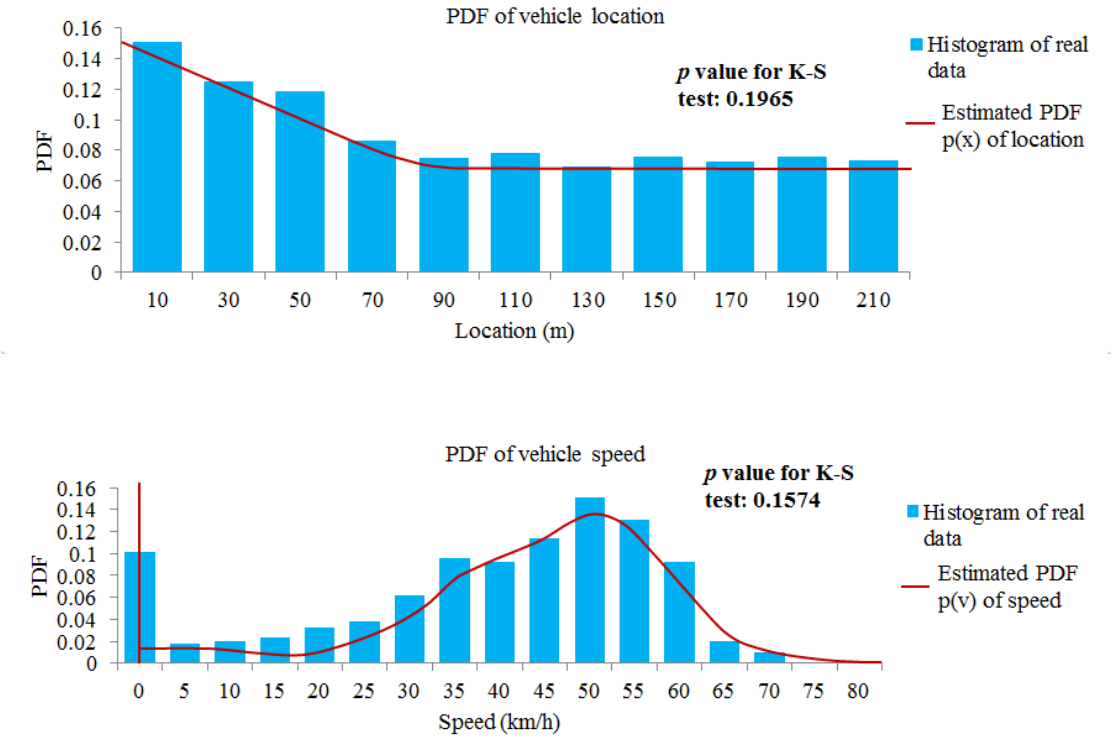


Figure 6.6 Histogram of data and estimated marginal distributions for the under-saturated regime

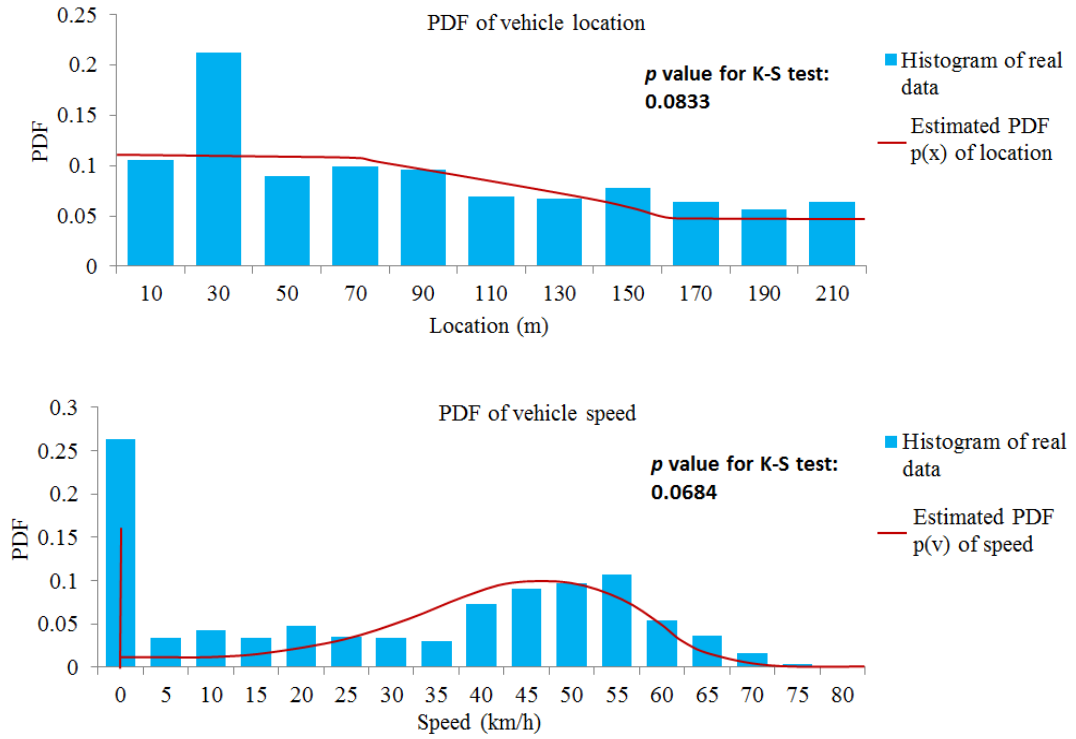


Figure 6.7 Histogram of data and estimated marginal distributions for the congested regime

6.7 Simplifying the joint PDF model

In order to capture the details of distribution of vehicle location and speed, the derived joint PDF models for both under-saturated regime and congested regime (section 6.4.2 and 6.4.3) are fine but complicate. The complexity of these models would limit the applications especially for road network where enormous computations are needed. The joint PDF models are complex, because we consider most details. In practice, some details are not significant comparing to the randomness of traffic. Therefore, it's necessary for us to simplify the joint PDF models while maintaining the major features of vehicle and speed distribution. In this section, we introduce our treatments for simplifying the joint PDF models for under-saturated regime and congested regime, respectively.

6.4.2 Under-saturated regime

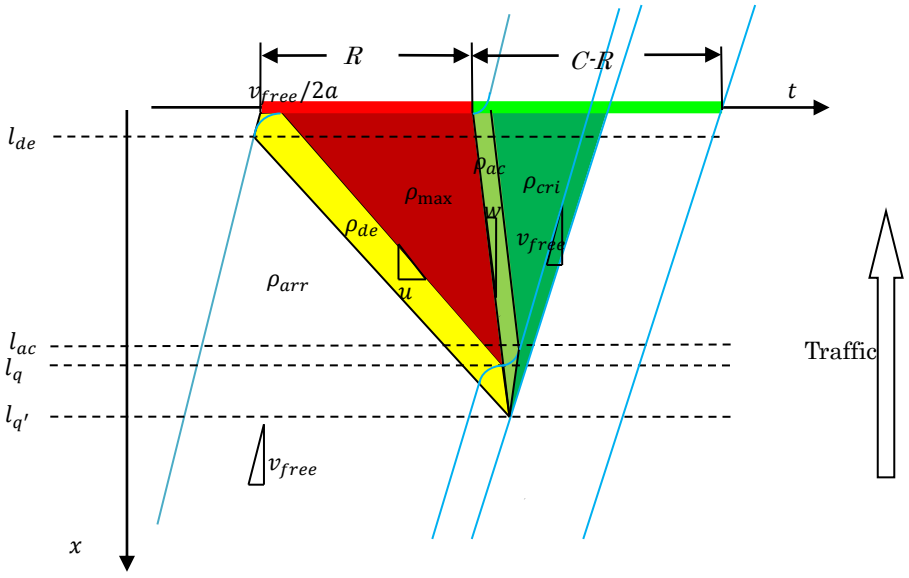


Figure 6.8 The original space-time diagram of under-saturated traffic regime

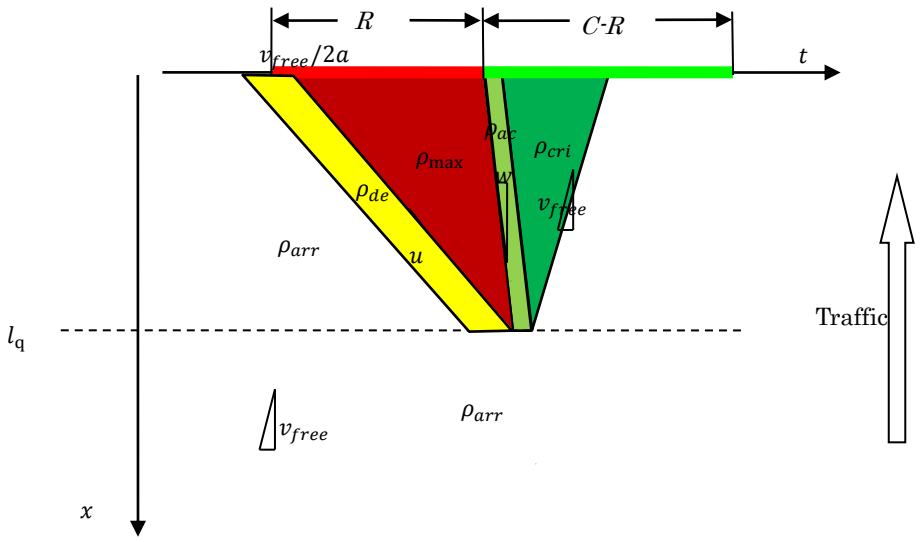


Figure 6.9 The simplified space-time diagram of under-saturated traffic regime

To clearly show our treatment, the original space-time diagram of under-saturated traffic regime is presented again in Figure 6.8. And the simplified one is presented in Figure 6.9. In Figure 6.9, we make three minor changes:

1) Extend the line that the deceleration begins to the stop line. Thus, the deceleration behavior in $[0,\ l_{de}]$ is regarded the same as that in $[l_{de},\ l_{ac}]$. The area added is tiny compared with the total deceleration area, since l_{de} is very small in reality.
2) Extend the line that the acceleration ends to the location l_q. In this way, he deceleration behavior in $[l_{ac},\ l_q]$ is regarded the same as that in $[l_{de},\ l_{ac}]$.
3) Delete the deceleration area, acceleration area and dissipating area in $[l_q,\ l_{q'}]$. The driving behavior is regarded as the same as in the free-flow area $[l_{q'},\ L]$.

In the original diagram (Figure 6.8), there are five possible cases for a probe point in the under-saturated traffic regime: $[0,\ l_{de}]$, $[l_{de},\ l_{ac}]$, $[l_{ac},\ l_q]$, $[l_q,\ l_{q'}]$, and $[l_{q'},\ L]$. After these treatments, we only need consider two cases for probe point: $[0, l_q]$ and $[l_q,\ L]$. Then, the joint PDF model for under-saturated traffic regime is

$$p(v,x) = \frac{\sum_{i=1}^{5} \rho_i t_i(x) g_i(v)}{\int_0^L \sum_{i=1}^{5} t_i(x) \rho_i dx} \tag{6.43}$$

where,

$l_q = (R - \frac{v_f}{2a})\frac{wu}{w-u}$

$\rho_1 = \rho_a,\ \rho_2 = \frac{1}{T_{de}(x)}\int_0^{T_{de}(x)} \rho_{max} K_{de}(\tau)d\tau,\ \rho_3 = \rho_{max},$

$\rho_4 = \frac{1}{T_{ac}(x)}\int_0^{T_{ac}(x)} \rho_{max} K_{ac}(\tau)d\tau,\ \rho_5 = \rho_c;$

$g_1(v) = N(\mu_a, \sigma_a),\ \ g_2(v) = \frac{(1-\frac{v}{v_f})^{\frac{v+u}{au}}}{\int_0^{T_{de}(x)} K_{de}(\tau)d\tau},\ g_3(v) = Dir_{\{0\}}(v),$

$g_4(v) = \frac{(1-\frac{v}{v_f})^{\frac{v+w}{bw}}}{\int_0^{T_{ac}(x)} K_{ac}(\tau)d\tau},\ g_5(v) = N(\mu_d, \sigma_d);$

$K_{de}(\tau) = 1 + \frac{u}{v_f} - \sqrt{\left(\frac{u}{v_f}+1\right)^2 - \frac{2au\tau}{v_f^2}},\ K_{ac}(\tau) = 1 + \frac{w}{v_f} - \sqrt{\left(\frac{w}{v_f}\right)^2 + \frac{2bw\tau}{v_f^2}};$

$T_{de}(x) = t_2(x),\ T_{ac}(x) = t_4(x);$

if $x \in [0, l_q]$,

$t_1(x) = C - \sum_{i=2}^{5} t_i(x)\ \ ,\ t_2(x) = \frac{v_f}{a}\left(\frac{v_f}{2u}+1\right)\ \ ,\ \ t_3(x) = \left(R - \frac{v_f}{2a}\right)\left(1 - \frac{x}{l_q}\right)\ \ ,$

$t_4(x) = \frac{v_f}{b}\left(\frac{v_f}{2w}+1\right),\ t_5(x) = \left(l_q - x\right)\left(\frac{1}{v_f}+\frac{1}{w}\right) - \frac{v_f}{b}\left(\frac{v_f}{2w}+1\right);$

if $x \in [l_q, L]$,

$t_1(x) = C,\ t_2(x) = t_3(x) = t_4(x) = t_4(x) = 0.$

We can further simplify model (6.43).

For $x \in [0, l_q]$, we have

$$\int_0^{T_{de}(x)} K_{de}(\tau)d\tau = \int_0^{T_{de}(x)} \left(1 + \frac{u}{v_f} - \sqrt{\left(\frac{u}{v_f}+1\right)^2 - \frac{2au\tau}{v_f^2}}\right)d\tau = \frac{v_f^2+3uv_f}{6au} \quad (6.44)$$

$$\int_0^{T_{ac}(x)} K_{ac}(\tau)d\tau = \int_0^{T_{ac}(x)} \left(1 + \frac{w}{v_f} - \sqrt{\left(\frac{w}{v_f}\right)^2 + \frac{2bw\tau}{v_f^2}}\right)d\tau = \frac{v_f^2+3wv_f}{6\mathrm{bw}} \quad (6.45)$$

Thus

$$g_2(v) = \frac{(1-\frac{v}{v_f})^{\frac{v+u}{au}}}{\int_0^{T_{de}(x)} K_{de}(\tau)d\tau} = \frac{6(1-\frac{v}{v_f})(v+u)}{v_f^2+3uv_f} \quad (6.46)$$

$$g_4(v) = \frac{(1-\frac{v}{v_f})\frac{v+w}{bw}}{\int_0^{Tac(x)} K_{ac}(\tau)d\tau} = \frac{6(1-\frac{v}{v_f})(v+w)}{v_f^2+3wv_f} \tag{6.47}$$

Moreover, we validate that $\int_0^{v_f} g_2(v)dv = 1$ and $\int_0^{v_f} g_4(v)dv = 1$. And then

$$\rho_2 = \frac{1}{T_{de}(x)}\int_0^{T_{de}(x)} \rho_{max}K_{de}(\tau)d\tau = \frac{\rho_{max}\left(v_f^2+3uv_f\right)}{3\left(v_f^2+2uv_f\right)} \tag{6.48}$$

$$\rho_4 = \frac{1}{T_{ac}(x)}\int_0^{T_{ac}(x)} \rho_{max}K_{ac}(\tau)d\tau = \frac{\rho_{max}\left(v_f^2+3wv_f\right)}{3\left(v_f^2+2wv_f\right)} \tag{6.49}$$

Then,

$$\rho_2 t_2(x)g_2(v) = \rho_{max}\left(1-\frac{v}{v_f}\right)\frac{v+u}{au} \tag{6.50}$$

$$\rho_4 t_4(x)g_4(v) = \rho_{max}(1-\frac{v}{v_f})\frac{v+w}{bw} \tag{6.51}$$

Assume the speed distributions in free-flow and dissipating traffic are the same as a normal distribution,

$$g_1(v) = g_5(v) = N(\mu, \sigma) \tag{6.52}$$

Then,

$$\sum_{i=1}^5 \rho_i t_i(x)g_i(v) = \rho_1 t_1 g_1 + \rho_2 t_2 g_2 + \rho_3 t_3 g_3 + \rho_4 t_4 g_4 + \rho_5 t_5 g_5 = (\rho_a t_1 + \rho_c t_5)N(\mu, \sigma) + \rho_{max}\left(1-\frac{v}{v_f}\right)\left(\frac{v+u}{au}+\frac{v+w}{bw}\right) + \rho_{max}t_3 Dir_{\{0\}}(v) \tag{6.53}$$

And,

$$\int_0^L \sum_{i=1}^5 t_i(x)\rho_i\, dx = C\rho_a L + l_q[t_2(\rho_2-\rho_a) + t_4(\rho_4-\rho_a)] + \frac{1}{2}l_q\left(R - \frac{v_f}{2a}\right)(\rho_{max}-\rho_a) + \frac{1}{2}l_q^2\left(\frac{1}{v_f}+\frac{1}{w}\right)(\rho_c-\rho_a) \tag{6.54}$$

Finally, we can obtain the simplified version of the joint PDF model for the under-saturated traffic regime

If $x \in [0, l_q]$,

$$p(v,x) = \frac{(\rho_a t_1+\rho_c t_5)N(\mu,\sigma)+\rho_{max}\left(1-\frac{v}{v_f}\right)\left(\frac{v+u}{au}+\frac{v+w}{bw}\right)+\rho_{max}t_3 Dir_{\{0\}}(v)}{C\rho_a L+l_q[t_2(\rho_2-\rho_a)+t_4(\rho_4-\rho_a)]+\frac{1}{2}l_q\left(R-\frac{v_f}{2a}\right)(\rho_{max}-\rho_a)+\frac{1}{2}l_q^2\left(\frac{1}{v_f}+\frac{1}{w}\right)(\rho_c-\rho_a)} \tag{6.55}$$

If $x \in [l_q, L]$,

$$p(v,x) = \frac{C\rho_a N(\mu,\sigma)}{C\rho_a L+l_q[t_2(\rho_2-\rho_a)+t_4(\rho_4-\rho_a)]+\frac{1}{2}l_q\left(R-\frac{v_f}{2a}\right)(\rho_{max}-\rho_a)+\frac{1}{2}l_q^2\left(\frac{1}{v_f}+\frac{1}{w}\right)(\rho_c-\rho_a)}$$

6.4.3 Congested regime

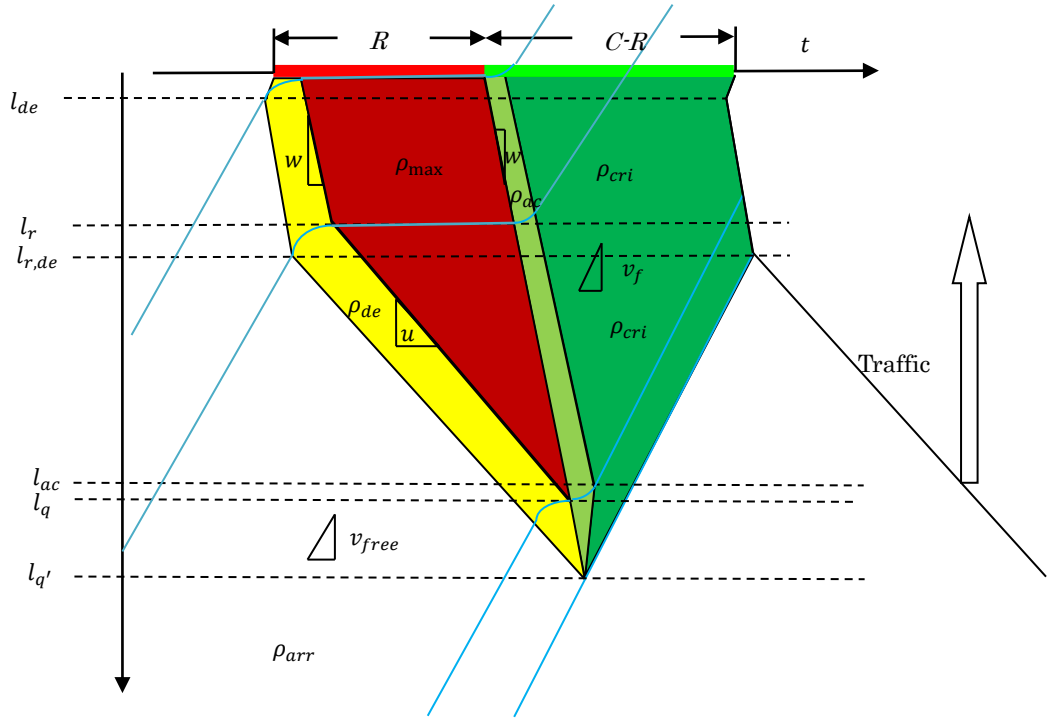


Figure 6.10 The original space-time diagram of congested traffic regime

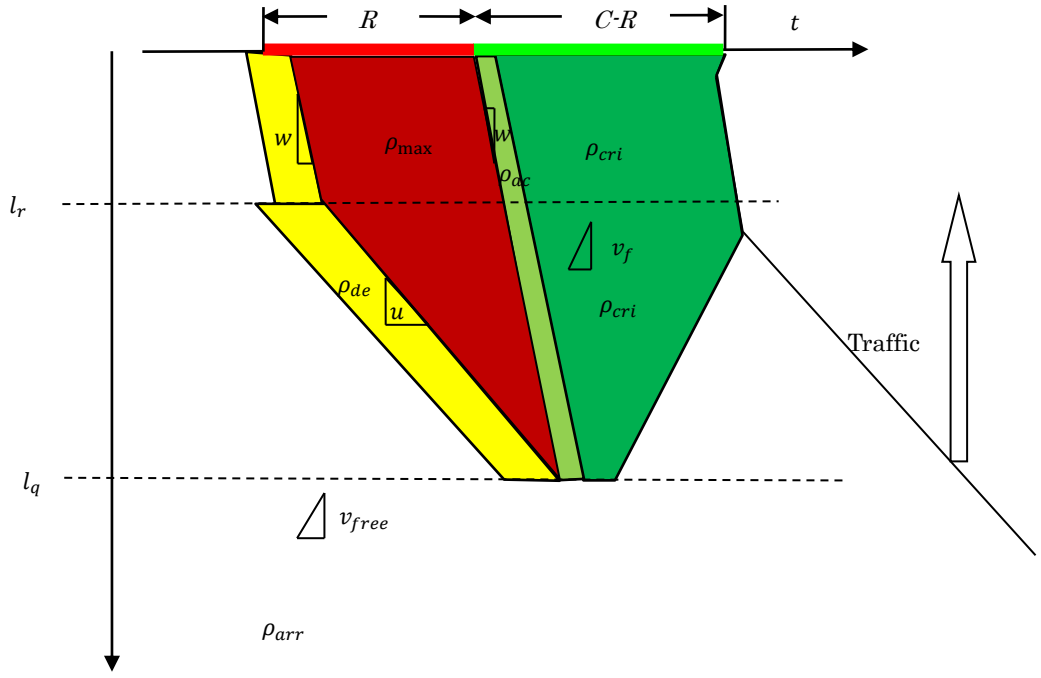


Figure 6.11 The simplified space-time diagram of congested traffic regime

Basically, the process of simplifying the joint PDF model for congested traffic regime is similar with that for under-saturated traffic regime. Compared with the

changes made for under-saturated traffic regime, one more is made for saturated traffic regime(see Figure 6.10 and 6.11):

- Extend the line that the deceleration from density ρ_{arr} begins to location l_r. Thus, the deceleration behavior in $[l_r,\ l_{r,de}]$ is regarded the same as that in $[l_{r,de},\ l_{ac}]$.

In the original diagram (Figure 6.10), there are seven possible cases for a probe point in the congested traffic regime: $[0,\ l_{de}]$, $[l_{de},\ l_r]$, $[l_r,\ l_{r,de}]$, $[l_{r,de},\ l_{ac}]$, $[l_{ac}, l_q]$, $[l_q, l_{q'}]$, and $[\ l_{q'},\ L]$. Among them, $[0,\ l_{de}]$, $[l_r,\ l_{r,de}]$, $[l_{ac},\ l_q]$, and $[l_q, l_{q'}]$ are determined by the deceleration and acceleration. In real situation, compared with the other situations, the length of them are very small. Accurately modelling them brings us very complex models of joint distribution and heavy computation burden. In the simplified traffic regime (Figure 6.11), there are three possible cases in the congested traffic regime: $[0,\ l_r]$, $[l_r,\ l_q]$, and $[l_q,\ L]$. Then, the joint PDF model for congested traffic regime is

$$p(v,x) = \frac{\sum_{i=1}^{5} \rho_i t_i(x) g_i(v)}{\int_0^L \sum_{i=1}^{5} t_i(x)\rho_i dx} \tag{6.56}$$

where

$$l_q = \left(R - \frac{v_f}{2a}\right)\frac{wu}{w-u} + l_r$$

$$\rho_1 = \rho_a,\ \rho_3 = \rho_{max},\ \rho_4 = \frac{1}{T_{ac}(x)}\int_0^{T_{ac}(x)} \rho_{max} K_{ac}(\tau)d\tau,\ \rho_5 = \rho_c;$$

$$g_1(v) = N(\mu_a, \sigma_a),\ g_3(v) = Dir_{\{0\}}(v),\ g_4(v) = \frac{(1-\frac{v}{v_f})\frac{v+w}{bw}}{\int_0^{T_{ac}(x)} K_{ac}(\tau)d\tau},$$

$$g_5(v) = N(\mu_d, \sigma_d);$$

$$K_{de}^{u}(\tau) = 1 + \frac{u}{v_f} - \sqrt{\left(\frac{u}{v_f}+1\right)^2 - \frac{2au\tau}{v_f^2}}\ ,\ K_{de}^{w}(\tau) = 1 + \frac{w}{v_f} - \sqrt{\left(\frac{w}{v_f}+1\right)^2 - \frac{2aw\tau}{v_f^2}},$$

$$K_{ac}(\tau) = 1 + \frac{w}{v_f} - \sqrt{\left(\frac{w}{v_f}\right)^2 + \frac{2bw\tau}{v_f^2}};$$

$T_{de}(x) = t_2(x),\ T_{ac}(x) = t_4(x);$

if $x \in [0, l_r]$,

$\rho_2 = \frac{1}{T_{de}(x)} \int_0^{T_{de}(x)} \rho_{max} K_{de}^w(\tau) d\tau,\ g_2(v) = \frac{(1-\frac{v}{v_f})\frac{v+w}{aw}}{\int_0^{T_{de}(x)} K_{de}^w(\tau) d\tau},$

$t_1(x) = 0\ ,\ \ t_2(x) = \frac{v_f}{a}\left(\frac{v_f}{2w} + 1\right)\ ,\ \ t_3(x) = \left(R - \frac{v_f}{2a}\right)\ ,\ \ t_4(x) = \frac{v_f}{b}\left(\frac{v_f}{2w} + 1\right)\ ,$

$t_5(x) = C - \sum_{i=1}^4 t_i(x);$

if $x \in [l_r, l_q]$,

$\rho_2 = \frac{1}{T_{de}(x)} \int_0^{T_{de}(x)} \rho_{max} K_{de}^u(\tau) d\tau,\ g_2(v) = \frac{(1-\frac{v}{v_f})\frac{v+u}{au}}{\int_0^{T_{de}(x)} K_{de}^u(\tau) d\tau},$

$t_1(x) = \left(\frac{1}{v_f} + \frac{1}{u}\right)(x - l_r)\ ,\ \ t_2(x) = \frac{v_f}{a}\left(\frac{v_f}{2u} + 1\right),\ \ t_3(x) = \left(R - \frac{v_f}{2a}\right)\left(1 - \frac{x - l_r}{l_q - l_r}\right)$

$t_4(x) = \frac{v_f}{b}\left(\frac{v_f}{2w} + 1\right),\ t_5(x) = C - \sum_{i=1}^4 t_i(x);$

if $x \in [l_q, L]$,

$t_1(x) = C,\ t_2(x) = t_3(x) = t_4(x) = t_4(x) = 0.$

Further, similar with the calculation in the case of under-saturated traffic regime, we can calculate ρ_4 and $g_4(v)$ for $[0, l_q]$,

$$\rho_4 = \frac{1}{T_{ac}(x)} \int_0^{T_{ac}(x)} \rho_{max} K_{ac}(\tau) d\tau = \frac{\rho_{max}\left(v_f^2 + 3wv_f\right)}{3\left(v_f^2 + 2wv_f\right)} \tag{6.57}$$

$$g_4(v) = \frac{(1-\frac{v}{v_f})\frac{v+w}{bw}}{\int_0^{T_{ac}(x)} K_{ac}(\tau) d\tau} = \frac{6(1-\frac{v}{v_f})(v+w)}{v_f^2 + 3wv_f} \tag{6.58}$$

Then, if $x \in [0, l_r]$, we have

$$\int_0^{T_{de}(x)} K_{de}^w(\tau) d\tau = \int_0^{T_{de}(x)} \left(1 + \frac{w}{v_f} - \sqrt{\left(\frac{w}{v_f} + 1\right)^2 - \frac{2aw\tau}{v_f^2}}\right) d\tau = \frac{v_f^2 + 3wv_f}{6aw} \tag{6.60}$$

Using Equation (6.60), we can easily obtain

$$\rho_2 = \frac{1}{T_{de}(x)} \int_0^{T_{de}(x)} \rho_{max} K_{de}^w(\tau) d\tau = \frac{\rho_{max}\left(v_f^2 + 3wv_f\right)}{3\left(v_f^2 + 2wv_f\right)} \tag{6.61}$$

$$g_2(v) = \frac{(1-\frac{v}{v_f})\frac{v+w}{aw}}{\int_0^{T_{de}(x)} K_{de}^w(\tau) d\tau} = \frac{6(1-\frac{v}{v_f})(v+w)}{v_f^2 + 3wv_f} \tag{6.62}$$

if $x \in [l_r, l_q]$,

$$\int_0^{T_{de}(x)} K_{de}^u(\tau)d\tau = \int_0^{T_{de}(x)} \left(1 + \frac{u}{v_f} - \sqrt{\left(\frac{u}{v_f}+1\right)^2 - \frac{2au\tau}{v_f^2}}\right) d\tau = \frac{v_f^2+3uv_f}{6au} \quad (6.63)$$

Using Equation (6.63), we can obtain

$$\rho_2 = \frac{1}{T_{de}(x)} \int_0^{T_{de}(x)} \rho_{max} K_{de}^u(\tau)d\tau = \frac{\rho_{max}\left(v_f^2+3uv_f\right)}{3\left(v_f^2+2uv_f\right)} \quad (6.64)$$

$$g_2(v) = \frac{(1-\frac{v}{v_f})\frac{v+u}{au}}{\int_0^{T_{de}(x)} K_{de}^u(\tau)d\tau} = \frac{6(1-\frac{v}{v_f})(v+u)}{v_f^2+3uv_f} \quad (6.65)$$

Therefore, the simplified joint PDF model for congested traffic regime is

$$p(v,x) = \frac{\sum_{i=1}^5 \rho_i t_i(x) g_i(v)}{\int_0^L \sum_{i=1}^5 t_i(x)\rho_i dx} \quad (6.66)$$

where,

if $x \in [0, l_r]$	$t_1(x) = 0,$ $t_2(x) = \frac{v_f}{a}\left(\frac{v_f}{2w}+1\right),$ $t_3(x) = \left(R - \frac{v_f}{2a}\right),$ $t_4(x) = \frac{v_f}{b}\left(\frac{v_f}{2w}+1\right),$ $t_5(x) = C - \sum_{i=1}^4 t_i(x).$	$\rho_1 = 0,$ $\rho_2 = \frac{\rho_{max}\left(v_f^2+3wv_f\right)}{3\left(v_f^2+2wv_f\right)},$ $\rho_3 = \rho_{max},$ $\rho_4 = \frac{\rho_{max}\left(v_f^2+3wv_f\right)}{3\left(v_f^2+2wv_f\right)},$ $\rho_5 = \rho_c.$	$g_1(v) = N(\mu, \sigma),$ $g_2(v) = \frac{6(1-\frac{v}{v_f})(v+w)}{v_f^2+3wv_f},$ $g_3(v) = Dir_{\{0\}}(v),$ $g_4(v) = \frac{6(1-\frac{v}{v_f})(v+w)}{v_f^2+3wv_f},$ $g_5(v) = N(\mu, \sigma).$
if $x \in [l_r, l_q]$	$t_1(x) = \left(\frac{1}{v_f}+\frac{1}{u}\right)(x - l_r),$ $t_2(x) = \frac{v_f}{a}\left(\frac{v_f}{2u}+1\right),$ $t_3(x) = \left(R - \frac{v_f}{2a}\right)\left(1 - \frac{x-l_r}{l_q-l_r}\right)$ $t_4(x) = \frac{v_f}{b}\left(\frac{v_f}{2w}+1\right),$ $t_5(x) = C - \sum_{i=1}^4 t_i(x).$	$\rho_1 = \rho_a,$ $\rho_2 = \frac{\rho_{max}\left(v_f^2+3uv_f\right)}{3\left(v_f^2+2uv_f\right)},$ $\rho_3 = \rho_{max},$ $\rho_4 = \frac{\rho_{max}\left(v_f^2+3wv_f\right)}{3\left(v_f^2+2wv_f\right)},$ $\rho_5 = \rho_c.$	$g_1(v) = N(\mu, \sigma),$ $g_2(v) = \frac{6(1-\frac{v}{v_f})(v+u)}{v_f^2+3uv_f},$ $g_3(v) = Dir_{\{0\}}(v),$ $g_4(v) = \frac{6(1-\frac{v}{v_f})(v+w)}{v_f^2+3wv_f},$ $g_5(v) = N(\mu, \sigma).$
if $x \in [l_q, L]$	$t_1(x) = C.$	$\rho_1 = \rho_a.$	$g_1(v) = N(\mu, \sigma).$

6.7 Summary

In this chapter, we have derived an analytical joint probability distribution function (PDF) of vehicle locations and speeds on an arterial link. To validate these distributions, a Kolmogorov-Smirnov test is performed using probe data collected during a field test in Toyota City. The numerical results show that vehicle locations and speeds are distributed according to the proposed model with confidence of 0.99 for both under-saturated and congested regimes. Additionally, the proposed model can capture most features of vehicle location and speed distributions.

It should be noted that, the driving behaviors are assumed the same for all vehicles in deceleration and acceleration in the proposed model. We make this treatment based on three points. (1), unlike in the situation 1 that vehicles travel freely, the driving behavior is mainly controlled by the signal in acceleration and deceleration and the differences of driver behavior among drivers are not large. (2), compared with other areas, the acceleration area and deceleration area are relatively small; the ignorance of differences of driving behavior will not bring great negative effects. (3), the expression of joint PDF will become more complex if the differences of driving behavior are considered. Additionally, the driving behaviors are assumed different in arriving and dissipating in the proposed model. In practical application, we can assume the same driving behaviors and the free-flow speed as their mean for simplicity.

Although the joint PDF is derived from extending a location distribution proposed by Hofleitner *et al.*, it is a brand new model which models location and speed distributions simultaneously. The derived joint PDF not only makes it possible to learn macroscopic traffic parameters and link parameters from location and speed data of probe vehicles, but also provides a way that describes vehicle speed

distribution in space and time for arterial roads.

References

Bar-Gera, H. (2007). Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: A case study from Israel. *Transportation Research Part C: Emerging Technologies*, 15(6) , 380–391.

Bails, C., Hofleitner, A. Xuan, Y. and Bayen A. (2012). A Three-Stream Model for Arterial Traffic. Transportation Research Board CD-ROM Preprints (TRB Paper No. 12-1212). National Research Council, Washington, D. C.

Ban, X., Hao, P. and Sun, Z. (2011). Real time queue length estimation for signalized intersections using travel times from mobile sensors. *Transportation Research Part C: Emerging Technologies*, 19(6), 1133–1156.

Ban, X., Herring, R. Hao, P. and Bayen, A. (2009). Delay pattern estimation for signalized intersections using sampled travel times. *Transportation Research Record: Journal of the Transportation Research Board*, 2130(1), 109–119.

Bierlaire, M., Chen, J. and Newman, J. (2013). A probabilistic map matching method for smartphone GPS data. *Transportation Research Part C: Emerging Technologies*, 26, 78–98.

Cao, P., Miwa, T., Yamamoto, T. and Morikawa, T. (2013). Bilevel Generalized Least-Square Estimation of Dynamic Origin-Destination Matrix for Urban Network Using Probe Vehicle Data. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 2333*, 66–73.

Evans, L. C. (1998). Partial Differential Equations. Graduate Studies in Mathematics, Vol. 19. American Mathematical Society, Providence, RI.

GPS.gov, (2013). http://www.gps.gov/systems/gps/performance/accuracy/. Accessed on June 20th, 2013.

Hao, P., Ban, X. Bennett, K. P. Ji, Q. and Sun, Z. (2012). Signal Timing Estimation Using Sample Intersection Travel Times. *IEEE Transactions on Intelligent Transportation Systems*, 13(2), 792–804.

Hellinga, B., Izadpanah, P. Takada, H. and Fu, L. (2008). Decomposing travel times measured by probe-based traffic monitoring systems to individual road segments. *Transportation Research Part C: Emerging Technologies*, 16(6), 768–782.

Hofleitner, A., Herring, R. and Bayen, A. (2012a). Arterial travel time forecast with streaming data: A hybrid approach of flow modeling and machine learning. Transportation *Research Part B: Methodological*, 46(9), 1097–1122.

Hofleitner, A., Herring, R., Abbeel, P. and Bayen, A. (2012b). Learning the dynamics of arterial traffic from probe data using a dynamic Bayesian network. *IEEE Transaction on Intelligent Transportation System*, 13(4), 1–15.

Hofleitner, A., Herring, R. and Bayen, A. (2011). A hydrodynamic theory based statistical model of arterial traffic. CCIT Research Report, UCB-ITS-CWP-2011-2.

Hunter, T., Abbeel, P. and Bayen, A. (2013). The path inference filter: model-based low-latency map matching of probe vehicle data. In *Algorithmic Foundations of Robotics X*, Springer Berlin Heidelberg, 591–607.

Jenelius, E. and Koutsopoulos, H. N. (2013). Travel time estimation for urban road networks using low frequency probe vehicle data. *Transportation Research Part B: Methodological*, 53, 64–81.

Lighthill, M. J. and Whitham, G. B. (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.

Liu, H. X., Wu, X., Ma, W. and Hu, H. (2009). Real-time queue length estimation for congested signalized intersections. *Transportation Research Part C: Emerging Technologies*, 17(4), 412–427.

Lou, Y., Zhang, C., Zheng, Y., Xie, X. Wang, W. and Huang, Y. (2009). Map-matching for low-sampling-rate GPS trajectories. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 352–361.

Massey, F.J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253), 68–78.

Mehran, B., Kuwahara, M. and Naznin, F. (2012). Implementing kinematic wave theory to reconstruct vehicle trajectories from fixed and probe sensor data. *Transportation Research Part C: Emerging Technologies*, 20(1), 144–163.

Miwa, T., Sakai, T. and Morikawa, T. (2004). Route Identification and Travel Time Prediction Using Probe-Car Data. *International Journal of ITS Research*, 2(1), 2004.

Skabardonis, A. and N. Geroliminis. (2005). Real-time estimation of travel times on signalized arterials. In *Proceedings of the 16th International Symposium on Transportation and Traffic Theory*, 1–21.

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

Van Aerde, M., Hellinga, B., Yu, L. and Rakha, H. (1993). Vehicle Probes as Real-Time ATMS Sources of Dynamic OD and Travel Time Data. Large Urban Systems, *Proceedings of the Advanced Traffic Management Conference*.

Witte, T. H. and Wilson, A. M. (2004). Accuracy of non-differential GPS for the determination of speed over ground. *Journal of Biomechanics*, 37(12), 1891–1898.

Zheng, F. and Van Zuylen, H. (2012). Urban link travel time estimation based on sparse probe vehicle data. *Transportation Research Part C: Emerging Technologies*, 31, 145–157.

Chapter 7

Conclusion and future work

In this thesis, we have explored methodologies of estimating urban traffic conditions using probe vehicle data for Intelligent Transportation Systems. In order to provide comprehensive knowledge of traffic conditions, we measure traffic conditions using three fundamental macroscopic variables: flow, speed and density. Considering the characteristics of probe vehicle data, travel times from probe vehicle data are applied in modeling flow and link travel time, and scatted probe points are applied in the density-based model for traffic monitoring. We first summarize the proposed methods in this thesis, and then describe the conclusions made from implementations or experiments. Furthermore, the limitations of this research are pointed out. Finally, several subsequent researches are indicated for future work.

7.1 Conclusion

7.1.1 Proposed methods

1) We proposed a three-step methodology of estimating dynamic link flow including: travel time allocation, link performance function fitting and dynamic link flow estimation. In the first step, method of proportional allocation is used to decompose probe travel time onto individual links. In the second step, link performance function is obtained from a derived speed-density function. The speed–density function is derived from Gazis's nonlinear follow-the-leader model. In order to estimate link flow, a Bayesian method (BM) that incorporates prior distribution of link speed and an ordinary method (OM) are applied in third step. In addition to the average value of link flow, Both the BM and OM estimate the variance of estimated link flow at the same time.
2) A two-stage methodology of estimating dynamic origin-destination (OD) matrices

using probe vehicle data is presented. In the first stage, dynamic link flows are estimated using the aforementioned three-step methodology. In the second stage, A DTA-based bi-level generalized least-square (GLS) estimator considering the distance between the estimated and target OD matrices as well as the distance between the calculated and observed link flows is formulated so as to estimate dynamic OD matrices from estimated link flows. In the iterative solution procedure, the upper level is solved using the extended Bell algorithm, while the commercial DTA system VISSIM is applied to produce the assignment matrix in the lower level.

3) An analytical model incorporated with a truncated distribution is proposed to model the probability distribution of travel time in a signalized road section. Link travel time is decomposed into time-in-motion (the time a vehicle is actually moving) and time-in-queue. Time-in-motion is modeled using a truncated distribution, while time-in-queue is modeled as a mixed distribution consisting of a mass distribution and a uniform distribution. A probability density function for travel time is then derived and parameterized by fraction of queuing vehicles, red signal time, and motion behavior parameters including truncation points. These parameters are obtained from sample link travel times using a maximum likelihood estimator.

4) The joint probability distribution function (PDF) of vehicle location and speed on an arterial road is proposed using hydrodynamic theory and horizontal queuing theory. Specifically, the joint PDF models for both under-saturated traffic regime and congested traffic regime are derived. The proposed models are parameterized by link parameters (red signal time, cycle time, and critical flow density), driving behavior (average deceleration and acceleration, average speed and variation of arrival flow and dissipating flow) and traffic state (arrival flow density) that are

learned from historical probe data (location and speed) by maximum likelihood estimation.

5) The simplified versions for the proposed joint PDF models are presented by making some small modifications on those intervals determined by the deceleration and acceleration. These simplifications make the proposed joint distributions be applicable in urban network with acceptable computation efficiency.

7.1.2 Performance of proposed methods

1) Proposed methodology of estimating dynamic link flows and O-D flows can provide acceptable estimates even for low polling frequency or low penetration probe vehicle data, it is recommended for applications in practice.
2) The Bayesian method can effectively utilize the prior information of travel time distribution from archived probe vehicle data, and can even give estimates for links no probe vehicle is observed.
3) The bi-level GLS estimator for dynamic O-D flow estimation is validated, and the convergence of extended Bell algorithm is validated. Additionally, the DTA module in VISSIM can be applied for solving the lower level of the bi-level model.
4) The truncation of link travel time distribution is significant and has impacts on the variability and reliability of travel time, and the proposed link travel time distribution can estimated from probe data that polling interval is less than 10s when the significance level is set to 0.1. It is recommended that truncation should be considered for many of the quantities (e.g. travel time, speed and demand) used in traffic engineering.
5) The proposed joint PDF model for vehicle location and speed is a brand new analytical model which model which models location and speed distributions

simultaneously. This joint PDF model can capture most features of empirical distribution of probe points. The derived joint PDF not only makes it possible to learn macroscopic traffic parameters and link parameters from location and speed data of probe vehicles, but also provides a way that describes vehicle speed distribution in space and time for arterial roads.

6) The simplified joint PDF model is more concise than the original one. And it is recommended that the simplified joint PDF model is applied for network-level traffic monitoring where huge computations are needed.

7.1.3 Limitations of this research

1) The polling frequency is assumed to be constant in one experiment. This may be not completely consistent with the real probe vehicle data in practice due to data missing, communication congestion, invalid record, etc.
2) Identical vehicles are assumed in all experiments, since the real composition of vehicles is unknown.
3) In the link flow estimation, prior distribution of link mean speed is necessary for each link in each time interval. To this end, large amounts of archived probe vehicle data are needed.
4) The arrival traffic for an intersection is assumed to be constant and stable in the time interval interested. Particularly, this is suitable for the cases of isolated intersections or non-isolated intersections with similar arrival traffic from different directions in the upstream. If the arrival traffic varies a lot, the performance of proposed models would decrease.
5) The proposed methods haven't been implemented and validated in practice, where more complex factors are needed to be considered.

7.2 Future work

This research makes a start for estimating comprehensive urban traffic conditions using probe vehicle data. Based on this research, there are many possible extension studies in future work.

1) The VISSIM commercial system works well as the traffic simulator in this research, but other commercial DTA systems may also be suitable. It is necessary to make a comparison of the effects that different DTA systems have on DTA-based O-D estimation.
2) In the O-D flows estimation, only link travel times from probe vehicle data are applied. More information provided by probe vehicles such as trip origin-destination points and trajectories may be useful to improve the estimation results.
3) Based on the derived model of link travel time distribution, it is possible to assess the variability and reliability of link travel time. And the derived model can also be applied for link/route travel time estimation.
4) Stochastic Frontier method can be applied to estimate the truncation points in the derived model of link travel time distribution.
5) The proposed joint PDF model of vehicle location and speed can be applied for traffic estimation and prediction. A possible program could be: An extended urban traffic network model (Lin and Xi, 2008; Lin *et al.*, 2012) is applied to model the spatial-temporal evolution of traffic; a dynamic Bayesian network (DBN) is used to model the stochastic dynamics of the traffic states and the dependencies of observations (probe data) and traffic states; and then the expectation-maximization (EM) algorithm (Dempster *et al*., 1977) can be used to estimate the maximization likelihood parameters of proposed model; finally, traffic states will be estimated in

real-time using the estimated parameters and real-time probe data.

6) The joint PDF may also be applied in traffic control. A rough idea is deriving an analytical expression of expected total delay, and then minimizing the expected total delay with respect to traffic signal parameters.
7) The proposed models should be implemented in real network with real probe vehicle data.

References

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1–38.

Lin, S. and Xi, Y. (2008). An efficient model for urban traffic network control. In *Proc. 17th IFAC World Congress*, 4066–4071.

Lin, S., De Schutter, B. Xi, Y. and Hellendoorn, H. (2012). Efficient network-wide model-based predictive control for urban traffic networks. *Transportation Research Part C: Emerging Technologies*, 24, 122–140.