The Stochastic Dynamic Journey Time Reliability Analysis by Considering the Spatial and Temporal Correlation

Tianlu, PAN

M. Phil

The Hong Kong
Polytechnic University

2012

The Hong Kong Polytechnic University

Department of Civil and Structural Engineering

The Stochastic Dynamic Journey Time Reliability Analysis by Considering the Spatial and Temporal Correlation

Tianlu, PAN

A Thesis Submitted in Partial Fulfillment

of the Requirements for the Degree of

Master of Philosophy

August, 2011

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis entitled "The Stochastic Dynamic Journey Time Reliability Analysis by Considering the Spatial and Temporal Correlation" is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

Signature:

Name of student: Tianlu PAN

Abstract

Travel time, which may be the most intuitive network performance index that can be perceived by the travelers and planners, is a fundamental measure in transportation systems. Because of their importance in traffic surveillance, management and control, path planning, and routing, travel time estimation and prediction have attracted significant research interests. On the other hand, traffic networks are fragile to the demand and supply uncertainties to which they exposed, especially under incident scenarios and adverse weather conditions. Due to the importance of transportation networks in economics and daily lives of citizens, the reliability of transportation networks cannot be overemphasized. Meanwhile, travel time reliability (TTR) has been widely recognized as one of the key performance measures that describe the reliability of a transportation network. To this end, this dissertation investigates two important topics: 1) estimating and predicting the distribution of stochastic dynamic travel time for short-term planning and intelligent transportation systems (ITSs) applications; 2) evaluating the travel time reliability index based on the stochastic dynamic travel time distribution obtained previously and analyzing the relationship between the TTR and the skewness of travel time distribution.

The thesis extends the definition of (deterministic) link travel time to a stochastic version by defining a kind of likelihood between the stochastic link inflow and outflow profiles. The physical meaning of the proposed likelihood is the probability that the difference between the cumulative link inflow and the cumulative link outflow be less than or equal to a prescribed bound, e.g one unit vehicle. Based on this likelihood definition, the probability mass function (PMF) of the link travel time is evaluated by defining some appropriate sampling interval. The dynamic link travel time distribution is evaluated by fitting this PMF with skew normal distribution. The dissertation then extends the deterministic nested delay operator to evaluate stochastic journey time distribution. The PMF of journey time is obtained by a series of “nested” conditional probabilities along the links on the route. By the same distribution fitting mechanism, the stochastic journey time distribution is deduced. Two empirical studies are conducted to verify the proposed method. The results prove a satisfactory performance of the proposed method for estimating and predicting stochastic dynamic travel time

and its distribution. It is also been validated that the shewness analysis is consistent with the empirical observations reported in the transportation literature. The results indicate the proposed methods are adaptive to abnormal traffic conditions.

To increase the accuracy of travel time prediction and to handle the abnormal traffic conditions such as traffic incidents, adverse weather conditions, the dissertation extends the SCTM to consider the temporal and spatial correlations of traffic flow for short-term traffic state prediction. Meanwhile, this dissertation also extends the methodology proposed previously to predict journey travel time distribution for TTR prediction analysis and real-time applications. To incorporate the correlations, the SCTM is expanded with a best linear predictor. This predictor is utilized to predict the inflow demand and supply functions. Historical traffic flow profiles are taken as inputs to the predictor to forecast the demand and supply functions. Meanwhile, the real-time measurement, as another input to the predictor, is utilized to correct the prediction. For real-time application, the prediction is conducted in a rolling horizon manner. The rolling horizon method is useful especially under abnormal traffic conditions, e.g. traffic incidents, adverse weather conditions. Finally, the traffic state and journey time predictions are verified by empirical studies, which prove that significant improvement can be achieved by incorporating the correlations into the SCTM framework.

In conclusion, this thesis contributes to the literature on estimation and prediction of stochastic dynamic traffic state and travel time distribution, as well as the travel time reliability analysis.

Acknowledgements

First of all, I would like to express my sincere gratitude to my supervisors: Dr. Agachai Sumalee and Prof. William H.K. Lam for their kind guidance and supervision. I have gained a lot of knowledge from them that are applicable for my future career. I would also like to thank Prof. Janny May Yee LEUNG and Prof. Zhiyou GAO for serving in my dissertation committee, and for their valuable suggestions and advices. Their valuable suggestions and criticisms have greatly enhanced the quality of this research and dissertation.

Many past and present members in transportation group of the Hong Kong Polytechnic University deserve my thanks for their friendship and technical expertise, including Dr. Renxin Zhong, Dr. Julio Ho, Dr. Paramet Luathep, and Dr. Karen Tam. I would also like to thank them for always being there to support me through the tough times as well as to share the great moments, for filling my last two years with so many unforgettable memories.

Finally, my love and thanks go to my parents. Mere words, in any of the languages that I know, are inadequate to express my gratitude for the unconditional love and support they have given to me.

Lucia Tianlu PAN

August 2011

Contents

List of Publications

Journal papers:

1. Sumalee, A., **Pan, T.**, Zhong, R., Uno, N., 2011. Dynamic stochastic journey time estimation and reliability analysis using stochastic cell transmission model: algorithm and case studies. Submitted to Transportation Research Part C, under revision.
2. Sumalee, A., Zhong, R., **Pan, T.**, Szeto, W., 2011. Stochastic cell transmission model (SCTM): A stochastic dynamic traffic model for traffic state surveillance and assignment. Transportation Research Part B, 45 (3), 507-533.
3. Zhong, R., Sumalee, A., **Pan, T.**, Lam, W.H.K, in press. Stochastic cell transmission model for traffic network with demand and supply uncertainties. Transportmetrica.

Conference proceedings:

1. Sumalee, A., **Pan, T.**, Zhong, R. (2011) Stochastic cell transmission model considering spatial and temporal correlations for traffic state prediction. Transportation Research Board 90th Annual Meeting, Washington, D.C., USA.
2. **Pan, T.**, Sumalee, A., Zhong, R. (2010) The stochastic cell transmission model considering spatial and temporal correlations for traffic state prediction. Proceedings of *the 15th Hong Kong Society for Transportation Studies (HKSTS) International Conference*, Hong Kong.
3. **Pan, T.**, Sumalee, A., Zhong, R., Uno, N. (2010) The Stochastic Cell Transmission Model Considering Spatial and Temporal Correlations for Traffic States Prediction. Proceedings of *the 3rd International Symposium of Dynamic Traffic Assignment (DTA)*, Takayama, Japan.

4. **Pan, T.**, Sumalee, A., Zhong, R., Uno, N. (2010) Dynamic stochastic journey time estimation and reliability analysis using stochastic cell transmission model: algorithm and case studies. Proceedings of *the* ***4th International Symposium on Transportation Network Reliability***, Minnesota, U.S.A.
5. **Pan, T.**, Sumalee, A., Zhong, R., Uno, N. (2010) Dynamic stochastic journey time estimation and reliability analysis using stochastic cell transmission model: algorithm and case studies, Proceedings of *the* 12th *World Conference on Transport Research Society (WCTR)*, Portugal, Lisbon.
6. **Pan, T.**, Sumalee, A. (2009) Dynamic Stochastic Journey Time Estimation and Reliability Analysis---the stochastic cell transmission model based approach. Proceedings of *the 14th Hong Kong Society for Transportation Studies (HKSTS) International Conference*, Hong Kong.
7. Sumalee, A., Zhong, R., **Pan, T.**, Uno, N. (2010) Stochastic cell transmission model for traffic network with demand and supply uncertainties. Proceedings of *the 3rd International Symposium of Dynamic Traffic Assignment (DTA)*, Takayama, Japan.
8. Zhong, R., Sumalee, A., **Pan, T.**, Varaiya, P. (2010) Stochastic cell transmission model for traffic network with demand and supply uncertainties. *Proceedings of the 4th International Symposium on Transportation Network Reliability*, Minnesota, U.S.A.
9. Sumalee, A., Zhong, R., Szeto, W.Y., **Pan, T.** (2009) Stochastic cell transmission model: traffic state modeling under uncertainties. Proceedings of *the 88th Transportation Research Board (TRB) Annual Meeting*, Washington, D.C., USA.
10. Sumalee A., Zhong, R., Szeto, W.Y., and **Pan T.** (2008) Stochastic cell transmission model under demand and supply uncertainties. Proceeding of *the 2nd Symposium of Dynamic Traffic Assignment (DTA)*, Leuven, Belgium.

List of Figures

List of Tables

Key notations

$n_i(k)$ number of vehicles in cell *i* at time step *k*;

$y_i(k)$ inflow volume from cell i-1 to i at time step *k*;

z number of cells in a SCTM segment;

$\rho_i(k)$ density of cell i at time step *k*;

$\rho_u(k)$ density of upstream boundary at time step *k;*

$\rho_d(k)$ density of downstream boundary at time step *k;*

$q_u(k)$ inflow rate into segment at time step *k;*

$q_d(k)$ outflow rate from segment at time step *k;*

$q_{in}^{l_m}(k)$ inflow rate into link l_m at time step k;

$q_{out}^{l_m}(k)$ inflow rate into link l_m at time step k;

$C_{in}^{l_m}(k)$ cumulative inflow into link l_m up to time step k;

$C_{out}^{l_m}(k)$ cumulative outflow from link l_m up to time step k;

$P_{k'|k}$ relative frequency of the exit time to be *k'* for vehicles entering the link at time step *k;*

$\tau_{l_m}(t)$ exit time from link l_m for vehicles entering link l_m at time *t*

$P_{m,nest}(\tau^m \mid k)$ relative frequency of τ^m to be the exit time step from link l_m for vehicles entering the origin of the path at time step *k*;

$\left[k_{lb}, k_{ub}\right]$ sampling region of link exit time;

α skewness parameter;

γ_{tm} temporal correlation coefficient;

γ_{sp} spatial correlation coefficient;

Chapter 1

Introduction

1.1 Motivation

Traffic flows on road networks are dynamic and stochastic. Dynamics reflects the evolutions of traffic states or traffic conditions over time for a certain location; and stochasticity originates from the uncertainties of travel demand and supply functions (such as the free-flow speed, capacity, and jam density) over time. Previous traffic empirical studies have revealed that the fundamental flow-density diagram admits large variations due to the weather conditions, vehicle types, driving behavior, *etc.* While the other kind of uncertainty is travel demand[1] variability, which is always regarded as recurrent uncertainty or disturbance in traffic flow. For O-D demand, the regular fluctuations may originate from human behaviors due to different days in a week or seasons in a year; while the abnormal weather conditions, traffic incident information and special events contribute to the irregular variations (the uncertainties of path flow demand and link flow demand are caused by similar reasons). The stochastic travel time is a result of these demand and supply uncertainties (Lint et al., 2008).

Traffic networks are fragile to the demand and supply uncertainties, especially under incident scenarios and adverse weather conditions. The importance of the reliability of transportation systems can never be overemphasized because of the importance of transportation systems in economics and daily lives of citizens. With the increase of

[1] For a traffic flow model, we may refer the inflow demand as travel demand in this thesis which should be distinguished from the concept of O-D demand. For example, a cell transmission model representation of a freeway segment, the upstream inflow and on-ramp flows to the segment are always taken as travel demand to the model in literature.

the value of travel time, unexpected schedule delay, including both early and late arrival, might incur loss to travelers. Therefore, travel time reliability (TTR) has been widely recognized as one of the key performance measures that describe the reliability of a transportation network. Journey time reliability, which describes the degree of stability of journey travel time for a certain route based on the travelers' experience or real-time prediction, significantly influences their route choice and departure time choice behavior. Moreover, the analysis and application of TTR heavily depend on the distribution of stochastic dynamic travel time. Therefore, to analyze the TTR we first need to provide accurate estimation and/or prediction of the distribution of stochastic dynamic travel time.

As an essential performance measure of the traffic network and an essential decision factor, travel time information is crucial for a variety of travel time reliability analysis and network planning applications. The estimation and prediction of travel time have attracted significant interests of researchers and other traffic engineering practitioners. By its very nature, travel time information adopted by the travel time reliability analysis and applications should possess several distinct characteristics: anticipative, dynamic, and furthermore, stochastic as mentioned previously. Conventionally, there are several main methods to estimate/predict travel (or journey) time: vehicle tracing technologies based approaches (e.g. Coifman and Cassidy, 2002), dynamic traffic assignment models (e.g. Carey and Ge, 2011), statistical and computational intelligence based approaches (e.g. Van Lint, 2004), and dynamic traffic flow propagation based matching algorithms (e.g. Lo and Szeto, 2002).

However, these existing methods either have high requirements for automatic vehicle identification (AVI), data mining algorithm or are only able to produce deterministic (or mean) travel (journey) time or mean travel time plus/minus its safety margins (e.g. variance of the travel time instead of the corresponding distribution). Also, these methods may not be suitable for current traffic measurement systems (*e.g.* the Performance Measurement System (PeMS) in California) which can just provide some

limited traffic information, such as flow counts, average speed, during fixed detecting time intervals with the detecting loops installed several kilometers away from the neighboring ones. Moreover, few of the existing methods can release the complete distribution of dynamic travel time. And, as it will be reviewed in Chapter 2, these methods are both analytically difficult and computationally expensive. Nevertheless, the evaluation process of route travel time distribution from link travel time is still not well developed. Therefore, in this dissertation, a practical method which is analytically easy and computationally efficient to estimate and predict dynamic travel time distribution for the purposes of TTR analysis and applications will be proposed.

Due to the similar environmental conditions and human behaviors, demand and supply profiles are correlated in space and time domain. However, most of the existing models and algorithms assume them to be independent. However, considering these spatial and temporal correlations along with the traffic dynamics might bring significant potential advantages for the development of efficient traffic state estimation/prediction (Chandra and Al-Deek, 2009; Min and Wynter, 2011). Therefore this dissertation further aims at estimating/predicting the distribution of dynamic travel time considering the spatial-temporal correlation of traffic flow. Detailed objectives of this dissertation and their interconnection will be discussed in the forthcoming sections.

1.2 Objectives

This thesis aims at developing a framework for assessing or predicting journey time reliability via integrating stochastic dynamic models and the dynamic traffic flow propagation based travel (journey) time estimation algorithm by considering temporal and spatial correlations of traffic flow. The following objectives will be pursued in this study:

i) To develop an algorithm to estimate/predict the distribution of stochastic dynamic

link travel time.

ii) To extend the algorithm to estimate/predict dynamic journey time distribution based on the link travel time distributions obtained in (i).

iii) To analyze the relationship between the skewness of travel time distribution and the level of congestion.

iv) To evaluate the travel time reliability index based on the dynamic journey time distribution generated in (ii) and to analyze the relationship between the reliability of travel time and the skewness of travel time distribution.

v) To extend the stochastic cell transmission model (SCTM) to consider spatial-temporal correlations of demand and supply functions to improve its accuracy for short-term traffic state prediction.

vi) To integrate (i)—(v) to develop a framework for journey time prediction and its reliability analysis.

1.3 Thesis contributions and organization

Chapter 2 carries out a brief literature review on related topics and provides background information on macroscopic traffic model, journey time estimation and prediction algorithm, and the implementation of temporal-spatial correlations of traffic flow for short-term traffic state prediction.

Chapter 3 reviews the development of stochastic cell transmission model (SCTM) (Sumalee et al., 2011) from the cell transmission model (CTM) (Daganzo 1994, 1995), the modified cell transmission model (MCTM) (Munoz et al., 2003). The chapter also discusses different representations and extensions of the SCTM with respect to different congestion levels of the underlying freeway segment and different

assumptions. Finally, several technical directions for the applications of SCTM in numerical simulation and empirical study are summarized.

Chapter 4 first introduces the definition of (deterministic) link travel time in terms of dynamic flow matching under first-in-first-out (FIFO) principle. The chapter then extends this definition to a stochastic version by defining a kind of likelihood between the stochastic link inflow and outflow profiles. Based on this likelihood definition, the probability mass function (PMF) of the link travel time is evaluated by defining some appropriate sampling interval. Then we can construct the dynamic link travel time distribution by fitting the PMF with skew normal distribution. After obtaining the link travel time distributions along a route, we extend the deterministic nested delay operator to evaluate stochastic journey time distribution. The PMF of journey time is obtained by a series of "nested" conditional probabilities along the links on the route. By the same distribution fitting mechanism, the stochastic journey time distribution is deduced. An empirical study is conducted to verify the proposed method. The results prove a satisfactory performance of the proposed algorithm for travel time distribution estimation. The shewness analysis is also consistent with the empirical observations argued in transportation literature (e.g.Van Lint *et al.*, 2008). Effects of different incidents are captured by the proposed model.

Chapter 5 extends the SCTM to consider the temporal and spatial correlations of traffic flow for short-term traffic state prediction and extends the methodology proposed in Chapter 4 to predict journey travel time distribution for TTR prediction analysis and real-time applications. Firstly, the SCTM is extended by a multivariate normal distribution based best linear predictor. This predictor is utilized to forecast the inflow demand and supply functions. Historical traffic flow profiles are taken as inputs to the predictor to forecast the demand and supply functions. The real-time measurements, to be more specific the "weighted" error between the statistics of the historical traffic flow profiles and the real-time measurements, are utilized to correct the prediction. The predicted demand and supply functions are taken as inputs to the

SCTM to perform short-term traffic state prediction. For real-time application, the prediction is conducted in a rolling horizon manner. The rolling horizon method is useful especially under abnormal traffic conditions, e.g. traffic incidents, adverse weather conditions. Finally, the traffic state and journey time predictions are verified by empirical studies. The results indicate that significant improvement over the origin SCTM can be achieved by incorporating the spatial-temporal correlations.

Chapter 6 gives a conclusion of this thesis. Some topics for the future research such as the application of journey time reliability prediction on route choice decision and traffic management as well as the issue of incident detection are also highlighted in this chapter.

An overview of the structure of the main body of this thesis is given in Figure 1.1.

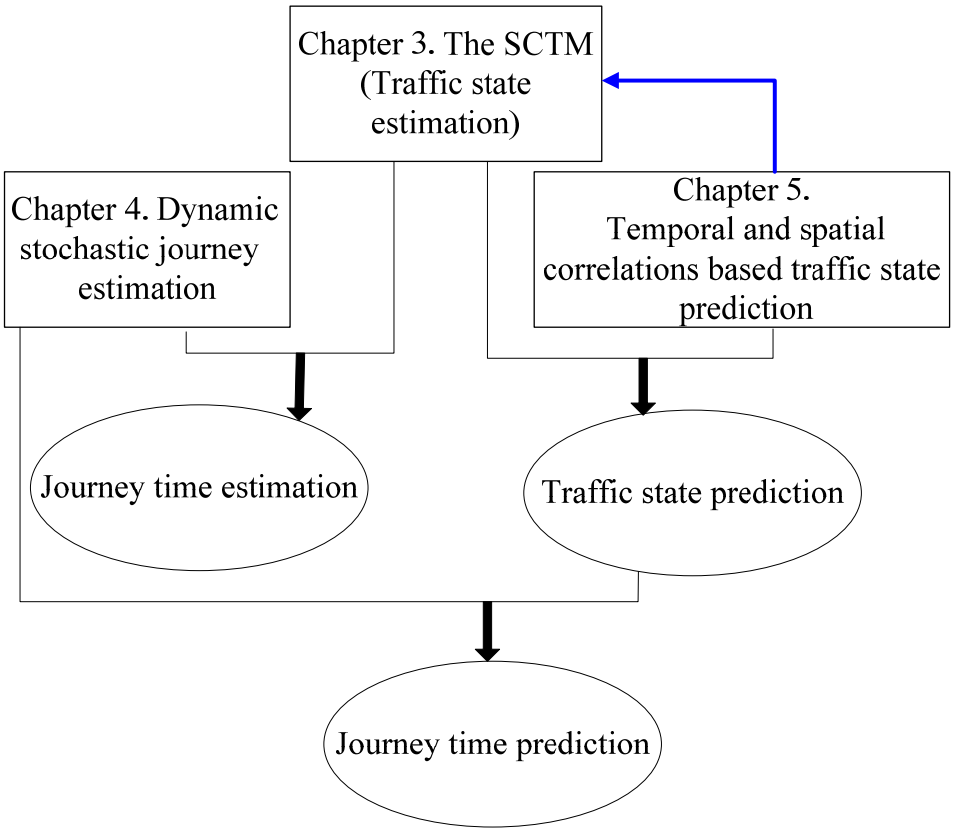


Figure 1.1. An overview of the structure of the thesis

Chapter 2

Background and literature review

2.1 Background

2.1.1 Traffic congestion and intelligent transportation systems

Traffic congestion on traffic networks has been recognized as a severe problem. Research has reached a consensus that building and expanding roadway facilities solely cannot handle the explosive growth in traffic volume and travel demand. This is because of the prohibitively high costs, as well as social, political, and environmental issues resulting from urban and suburban infrastructure construction. This problem motivates the idea of developing Intelligent Transportation Systems (ITS) to manage the road network infrastructure and to make efficient use of the existing capacity. Recently, the ITS has been recognized as an effective tool to alleviate traffic congestion by leading to more efficient travel demand and transportation network management, improving safety and efficiency, and hence reducing vehicle emissions.

Among various subsystems of an ITS, the advanced traveler information system (ATIS) and the advanced transportation management system (ATMS) are key components. The basic objectives of the ATIS are to acquire, analyze, communicate, and present information to users to enhance personal mobility and hence the efficiency of travel, and safety. The ATIS provides pre-trip and/or en route travel information concerning traffic conditions through various information media. The information is broadcasted to support travelers' decision making. Meanwhile, the ATMS collects data from a variety of sources, such as loop detectors, probe vehicles, video cameras, and other communication systems. The ATMS aims at managing and adjusting the traffic control systems in the network to respond to dynamic traffic conditions released from the collected data. The ATMS is proven to be helpful in optimizing urban traffic

signals, ramp-metering control, variable speed limit (VSL) control, etc. Since the ATIS and ATMS are interrelated, they are also known as advanced traffic management information systems (ATMIS).

As an essential performance measure of the traffic network and a significant decision factor, travel time information is crucial for a variety of ATIS and ATMS applications. For instance, travel time is the key information among the information broadcasted to travelers by ATIS. The ATMS aids in providing route guidance to travelers (drivers) through the ATIS. Among the performance indexes, travel time is an essential performance index of the ATMS as it is the key decision variable for route choice as well as traffic control design (see e.g. Van Lint (2004) for an overview of ITS applications of travel time information).

Accurate travel time estimation and prediction are essential to access the underlying network traffic conditions for the purposes of path planning, routing. For example, under the rational assumption of the road users[1], providing travelers with accurate (estimated and/or predicted) travel time information allows them to make more reasonable decisions, which in turn yields not only cost-benefits for individual road users, but potentially results in more stable and less congested traffic conditions. Travel time estimation and prediction are also important inputs to traffic management and control systems (see e.g. Papageorgiou and Kotsialos, 2002; Papageorgiou et al., 2003). Travel time also provides an insight for the traffic surveillance (Chen, 2003). The estimation and prediction of travel time have attracted significant interests of researchers and other traffic engineering practitioners. By its very nature, travel time

[1] Rational road users are decision makers who make their choices (e.g. route and departure time) to minimize their expected costs (e.g. in terms of travel time and travel time reliability), subject to their personal preferences, attitudes (e.g. risk taking behaviors) and perception of the information on the traffic network to minimize their expected costs (e.g. in terms of travel time and travel time reliability), subject to their personal preferences, attitudes (e.g. risk taking behaviors) and perception of the information on the traffic network.

information adopted by the ATIS and ATMS applications should possess several distinct characteristics: anticipative, dynamic, and furthermore, stochastic.

2.1.2 Stochastic travel time and travel time reliability

Traffic networks are exposed to uncertainties from both demand and supply sides. Figure 2.1 depicts a typical calibration of fundamental flow-density diagram from empirical traffic data. From this figure, we note that the fundamental flow-density diagram admits large variations due to the extensive congestion, driving behavior, *etc.* The stochastic flow-density diagram can be regarded as supply uncertainty. In the microscopic modeling approach, stochastic flow-density diagram has been interpreted as the effects of anticipation, strong correlations in the vehicle motion on different lanes, delay in the driving adaptation or safe time-gap variations (Ngoduy, 2009). In the macroscopic modeling approach, stochasticity of the flow-density diagram has been modeled as a diffusion coefficient to reproduce significant elements of the synchronized traffic flow, the interactions between several vehicle classes (*e.g.* trucks and cars), randomness in driving behavior, traffic incidents, and adverse weather conditions, etc., (Chen et al., 2001; Ngoduy, 2009). The other kind of uncertainty is travel demand variability, which is always regarded as recurrent uncertainty or disturbance in traffic flow dynamics. Stochastic distribution of travel time is a result of these demand and supply uncertainties (Lint et al., 2008).

Traffic networks are fragile to these demand and supply uncertainties, especially under incident scenarios and adverse weather conditions. For instance, the incident on 9th May 2005 involved a heavy rainstorm and incidents on three roads in Kowloon area causing a severe gridlock problem. The rainfall on 23 June 2011 in Beijing caused a widely traffic congestion. Journey time reliability, which describes the degree of stability of journey travel time, plays an important role in travelers' route choice and departure time choice behavior. Therefore, the travel time uncertainty needs to be taken into account in the context of ATIS/ATMS to better enhance the performance of

transportation systems.

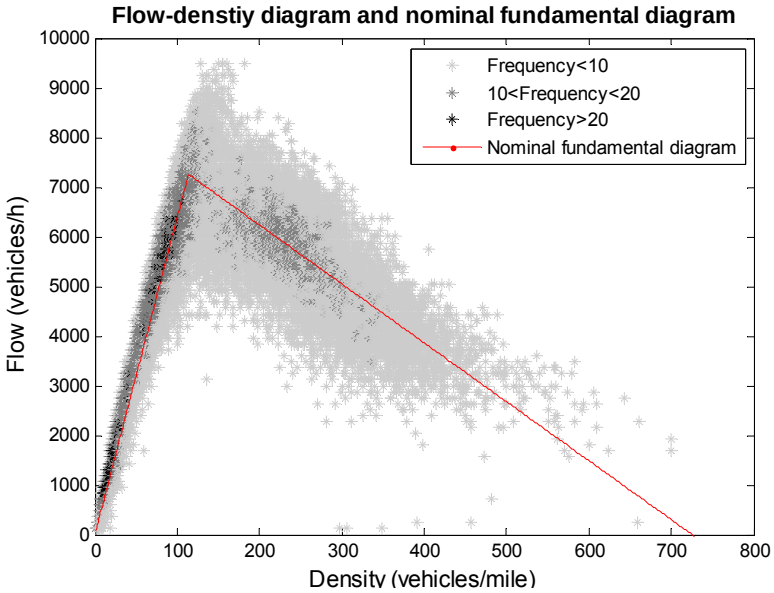


Figure 2.1. A fundamental flow–density diagram of traffic flow

As transportation network is a backbone of a city, the disruption or failure of the transport system may jeopardize the security and welfare of the population. For these reasons, travel time reliability (TTR) has been widely recognized as one of the key performance indicators of transportation networks (see e.g. Bell, 1999; Bell, and Cassir, 2000; Lo, 2002; Cassir et al., 2001). Existing studies on TTR primarily focus on devising indices to quantify the level of journey time uncertainty (or reliability) in a static network model (Bell and Cassir, 2000). Applications of travel time reliability for performance assessments of transportation network planning are investigated by many researchers (see, e.g. Plugurtha and Pasupuleti, 2010; Lyman and Bertini, 2008; Susilawati, et al., 2010). Travel time reliability based route choice/departure time choice modeling frameworks are also proposed as another application of travel time reliability (see, e.g. Bogers and Van Zuylen, 2004; Shao et al., 2006). Despite its importance, there is no uniform definition of TTR in sense that what should be precisely entailed by TTR or how it should be made operational (see e.g. Bell, 2000, 2002; Lo, 2002; Cassir et al., 2001). Nevertheless, the propagation of delay and uncertainty through the traffic network dynamically and spatially can be naturally observed. The level of travel time reliability and uncertainty, hence, should be

assessed dynamically.

2.1.3 Measures for travel time reliability

As previously explained, there are many different definitions for travel time reliability. As a subsequence, different measures for travel time reliability have been proposed in literature. However, what these measures have in common is that the longer-tailed the travel time distribution is on a particular time-of-day (TOD), day-of-the-week (DOW), the more unreliable travel time on a freeway network or corridor is considered (Lint et al., 2008). The above common measure yields a relative common definition of optimality in stochastic routing is to balance the risk of arriving late and travel time budgets rather than to minimize the expected travel time, i.e. travelers tend to prefer reliable over (on average) faster routes, in case the latter ones are very unreliable. A comprehensive overview of travel time reliability measures can be found in Lomax et al. (2003) and Lint et al. (2008). In this section, we summarize several measures for travel time reliability.

1. Statistical range methods, which is directly related to the shape of travel time distribution. For instance, Travel Time Window (e.g. mean travel time ± standard deviation). The underlying “plus” or “minus” operation indicates the possible spread of travel time around the expected value, wherein the distribution of travel time is implicitly assumed to be symmetric. Some other measures that can be categorized into this sort are the Percent Variation (i.e. the ratio of standard deviation and mean travel time), the Variability Index (the ratio between 95% travel time during peak hours and 95% during off peak).

2. Buffer time methods. The buffer time index represents the extra time that travelers must add to their average travel time when planning trips to ensure their on-time arrivals (or to have less than X% chance to miss an appointment) (Lomax *et al.,* 2003). The buffer time calculates the minutes of extra time needed to guarantee a statistically minimum number of arrivals within the preferred arrival time at destination (Pearce,

2001; Lomax et al., 2003). In this dissertation, we will apply buffer time index to characterize the travel time reliability.

3. Tardy-trip measures (e.g. misery index), which is defined as the gap between the average travel time of the 20% worst trips and the overall mean travel time (Pearce, 2001; Lomax et al., 2003).

4. Probabilistic measures: the probability that a trip would be made within the nominal travel time multiplied by a prescribed factor (Bell, 1999; Yang et al., 2000). The probabilistic measures are often used as measure for travel time unreliability in literature. Probabilistic measures utilize either a threshold for travel time or a predefined time window to distinguish between reliable and unreliable travel times. In this sense, choosing the parameters (e.g. prescribed factor) properly is essential for these probabilistic measures, which renders they are application and context specific.

5. Skew-width measures, wherein skewness of travel time is defined as the ratio of the distance between the 90^{th} and 50^{th} percentile travel time to the distance between the 50^{th} and 10^{th} percentile travel time. Generally, larger skewness of travel time means higher probability for extreme travel times (relative to the mean travel time) to occur.

2.2 Travel time estimation and prediction

2.2.1 Travel time estimation and prediction[2]

Much research work has been dedicated recently to the study of the way travelers react to the shape of the DDTT (see e.g. Bates et al. (2001); Pattanamekar et al. (2003); Chang et al. (2005); Gao and Chabini (2006)). However, less attention has been paid to developing tools for estimating the DDTT itself Sun et al. (2008); Yeon et al. (2008). In the context of optimal routing, most of the algorithms consider the expected value of link travel-time as a sufficient statistic for the problem and produce least expected travel-time paths, i.e. travelers are assumed to choose their routes with minimum expected costs (Fu and Rilett, 1998; Miller- Hooks and Mahmassani, 1998, 2000; Waller and Ziliaskopoulos, 2002). In reality, link travel times are heavily dependent on prevailing physical, traffic, and environmental conditions that cause the travel time to exhibit stochastic and time-variant behavior. In dynamic aspect, Bertsekas (2005) formulates the stochastic optimal routing problem as a stochastic optimal control problem. Static information about the network structure is combined with real-time information about actual travel-times to deduce an adaptive solution that is an optimal policy as opposed to an optimal path. An (adaptive) optimal policy generates a node-based decision rule that defines the optimal path from a given node to the destination conditioned on the realized travel-time. Gao and Chabini (2006); Gao et al. (2008) studied this optimal policy for optimal routing in stochastic time-dependent networks. Gao and Huang (2011) extended the model to consider the

[2] Travel time estimation is to "pertain" to reconstructing (mean) travel times of realized trips based on known speeds, flows, travel times, or other quantities which are mathematically related to travel times. On the other hand, travel time prediction refers to calculating travel times for unknown (future) traffic conditions. Indirect travel time prediction is to predict traffic quantities (e.g. speeds, densities) which are used to derive travel times, while direct travel time prediction is to predict travel times without the intermediate step of predicting other traffic quantities.

effect of real-time information on adaptive routing. However, all these methods ignore the issue of travel-time reliability which is an important decision factor in numerous practical applications. Several studies extend the dynamic optimal routing algorithms to consider the travel-time reliability, e.g. Gao et al. (2010); Gao and Huang (2011) adopted the prospect theory approach to capture travelers' strategic behavior when adapting to revealed traffic conditions en route in a stochastic network; Samaranayake et al. (in press) defined an optimality policy as: maximizing the probability of arriving on time at a destination given a departure time and a time budget which was validated by implementing the algorithm in the Mobile Millennium traffic information system. Nevertheless, recent research also suggests that even both the mean and variance of the travel time are obtained, still they should be used and interpreted with some reservations when telling the travel time unreliability, since they only account for a part of the costs of unreliability (van Lint et al., 2008; Cassir et al., 2001). The skewness of the travel time distribution plays an important role in travel time unreliability. As an essential performance measure of the traffic network and an essential decision factor, travel time information is crucial for a variety of travel time reliability analysis, network planning applications and optimal control and routing algorithms. For these reasons, it is instructive and important to develop sound evaluation methods for the distribution of stochastic dynamic link travel time (DDTT).

Travel times, taken as the reciprocal of speed, are shown to be roughly normally distributed. However, this conclusion is valid only under the assumptions that individual vehicles maintain a constant speed and that speeds are symmetric about their means (Kharoufeh and Gautam, 2004). Travel times are distributed with skewness under congested traffic condition.

Under deterministic environment, theoretical models have also been developed for estimating dynamic travel time from loop detector data based on traffic flow theory. To evaluate the travel time via the traffic flow propagation, the principle of

first-in-first-out (FIFO) of traffic flows and flow conservation is utilized, i.e. comparing the cumulative inflow and cumulative outflow profiles in which the exit time for traffic entering the link at time *t* is the time when the cumulative outflow of that link is equal to the cumulative inflow up to time *t*. Some of these methods perform well for normal-flow conditions only, while others are applicable for congested traffic flow conditions only. The results are reported as unsatisfactory during the transition periods from congested to un-congested conditions and vice versa. These models also tend to underestimate travel times under congested conditions because of the difficulties in capturing vehicle queue formations and dissipations under stochastic environment (traffic flow exhibits more randomness under congested condition) (Liu et al., 2010; Kharoufeh and Gautam, 2004; Soriguera et al., 2010).

Conventionally, there are several main methods to estimate/predict travel (or journey) time under stochastic environment: estimation from advanced surveillance systems: video image processing, automatic vehicle identification, smart phone/GPS tracking, probe vehicles (see e.g. Petty et al. (1998); Coifman (2002); Dion and Rakha (2006); Tam and Lam (2008); Haghani et al. (2010); Herrera et al. (2010); Sumalee and Wang (2012)), dynamic traffic assignment models (see e.g. Peeta and Ziliaskopoulos (2001); Wen (2008)), statistical and computational intelligence based approaches (see e.g. Wu et al. (2004); Yeon et al. (2008); van Lint et al. (2005); van Lint (2008); Jula et al. (2008); Hinsbergen et al. (2009); Ng and Waller (2010); Karlaftis and Vlahogianni (2011)), and dynamic traffic flow propagation based matching algorithms (see e.g. Lo and Szeto (2002); Carey and Ge (in press)). These methods can be regarded as fitting different types of models to observed data at individual sites. In this sense, major techniques for short-term travel-time estimation/prediction are of three major categories in transportation literature: parametric methods[3] (e.g. linear regression,

[3] The common assumption of the parametric methods is the model parameters of the system are known precisely. The parametric methods may produce larger prediction errors when the model parameters are biased.

time series models, dynamic traffic assignment models, Kalman[4] filtering techniques), nonparametric statistical methods (e.g. neural network models, simulation models, Bayesian models, support vector regression), and hybrid integration methods. The statistical approaches have more solid and widely accepted mathematical foundations than the CI based approaches (Karlaftis and Vlahogianni, 2011). On the other hand, the statistical approaches frequently fail when dealing with complex and highly nonlinear data and suffer from the curse of dimensionality. As highlighted by Karlaftis and Vlahogianni (2011), the CI based approaches, combining techniques from machine learning, adaptation, evolution and fuzzy logic to create intelligent models to emerge unstructured data by following some performance indexes, may outperform the statistical approaches. However, we may point out that abnormal traffic patterns caused by non-recurrent congestion or incidents may deteriorate the performance of these models (Fei et al., 2011). Some of these methods do not work well under demand and supply uncertainties. The DTA based approaches would suffer

[4] The Kalman filter (as a kind of state-space models) based approaches belong to the multivariate family of time series models that provide a good basis for modeling traffic data. In statistics and filtering literature, Kalman filter mainly refers to estimation of the state variables, which is an algorithm for efficiently doing exact Bayesian inference in a linear dynamical system. A commonly applied technique based on Kalman filtering for traffic state (including travel time) estimation is developed based on a predictor–corrector form of the Kalman filter wherein historical data is used for predicting the travel times, and real-time measurements are used to correct and update the prediction at each time instant (Vlahogianni et al., 2004; Jula et al., 2008). For the sake of simplicity and technical limitations of Kalman filter, most of the existing travel time estimators/predictors ignore possible spatial and temporal correlation between adjacent links and time instances. Under such kind of simplification, link travel times are implicitly assumed to be separable. In other words, to estimate/predict the travel time along a link, we utilize the historical and real-time data on the link and ignore possible network affects, e.g. spillback effect of congestion. As expected, these approaches have limited performance under congested traffic condition especially in presence of queue spillover. The readers can refer to an overview paper (Vlahogianni et al., 2004) and technical papers (Wang and Papageorgiou, 2005; Jula et al. 2008) for detailed discussions on Kalman filtering with applications in traffic engineering. The underlying assumptions of Kalman filtering are: the system is observable; the system has no strong nonlinearity (so that the system dynamics can be well approximated by its linearization).

from the disadvantages of the accuracy of network loading models and scalability. Nevertheless, most of these methods provide mean and variance of the dynamic travel time only, which may not sufficient for a broad range of applications.

As to the analytical aspect, only few works has dedicated to the stochastic travel time distribution (Kharoufeh and Gautam, 2004). The approach implicitly captures the time dependence of vehicle speed. The stochastic process for vehicle speed leads to a partial differential equation. When the environmental process is known to be a continuous-time Markov chain (CTMC), an explicit matrix equation for the stochastic dynamic link travel time distribution can be obtained. Although the technique presented therein provides an analytical result for the link travel-time distribution, some ingenuity will be required in determining the appropriate selection of transition rates for the CTMC and the speed function, no general solution is provided. Neither explicit expressions for the moments of the random travel time nor asymptotic approximations thereof are obtained in Kharoufeh and Gautam (2004). Since the method based on some partial differential equations (PDEs) and CTMC, it is very computationally demanding. Nevertheless, this method is not ready to be extended to capture route travel (or journal) time distributions.

2.2.2 Spatial-temporal correlations of traffic flow and its application to traffic forecasting

Short-term traffic prediction aims at estimating the anticipated traffic conditions at a future time given the historical traffic information in the "near future" and real-time detected traffic information (Lam et al., 2005; Sheu et al., 2009; Tam and Lam, 2008; Vlahogianni et al., 2004; 2005; 2007). Short-term traffic state prediction is one of the critical components of an ATIS and real-time implementation of surveillance and control tools. We adopt the table from Vlahogianni et al. (2004) to conclude the review on characteristics of the widely used methods for short-term traffic state forecasting in traffic engineering for completeness.

Most of the time series methods, e.g. the univariate Auto-Regressive Integrated Moving Average (ARIMA) time series models, are based on the autocorrelation function of the time series of traffic variable at a specific location, e.g. the location with measurement devices (Smith et al., 2002). Under this circumstance, the traffic state information of upstream and downstream locations has been largely neglected in these traffic prediction methods.

However, due to high traffic density and congestion in the network as well as the interaction of the demand and supply uncertainties along with the dynamic nature of traffic flow, the demand and supply uncertainties are correlated in both space and time domains. For example, the free-flow speeds are spatial correlated (cell-to-cell, lane-to-lane correlated) (Kwon et al., 2003); the demand profiles are temporal correlated (Yuan and Mills, 2005); the traffic state of a specific site is highly affected by the upstream and downstream traffic conditions (Chandra and Al-Deek, 2009; Min and Wynter, 2011). By considering these spatial and temporal correlations along with the traffic dynamics bring significant potential advantages for development of efficient traffic state estimation/prediction. The heuristic nearest neighbor methods, e.g. k-nearest neighbor algorithm (*KNN*) (Lam et al., 2005), incorporate temporal and spatial traffic characteristics into the forecasting process. However, performances of such kind of heuristic nearest neighbor methods are limited. As concluded by Smith et al. (2002) that the seasonal ARIMA model produces better accuracy than the heuristic nearest neighbor method at the price of highly computationally intensive.

Major existing approaches, regarding the treatment of temporal and spatial features of traffic flow for short-term forecasting first need to calibrate both temporal and spatial correlations among the available traffic data. One has to process an extremely large amount of data simultaneously in the calibrations of spatial and temporal correlations (Stathopoulos and Karlaftis, 2003). The calibrated results are then fed to the prediction algorithm externally (Vlahogianni et al., 2004). Besides the extremely large amount of data and arduous calibration, this kind of methods requires another preprocessing step.

In the preprocessing step, both autocorrelations and cross-correlations are calculated to obtain the appropriate time lag of each correlated series (Clark et al., 1993). These preprocessing steps, which are also computationally expensive, render the spatial and temporal features of traffic flow difficult to be incorporated in short-term traffic prediction. An alternative approach to consider the spatial and temporal correlation analysis is to utilize dynamics of the traffic model in conjunction with some extended dynamics, e.g. random walk (Wang and Papageorgiou, 2005), colored noise (Chui and Chen 2009). In this approach, the spatial and temporal correlations of the uncertainties are propagated by the extended dynamics, which in turn affects the traffic flow models.

Finally, we would like to point out that all of the above mentioned approaches for travel time estimation and/or prediction, except the analytical approach proposed by Kharoufeh and Gautam (2004), produce mean travel time or mean and variance of travel time rather than its distribution. Although the analytical approach proposed by Kharoufeh and Gautam (2004) provides approximation of dynamic link travel time distribution, it is very computationally expensive and is not ready to be extended to estimate/predict route travel time distribution. In this dissertation, we will propose a practical method that is analytically easy and computationally efficient to estimate and predict dynamic travel time distribution considering the spatial-temporal correlation of traffic flow. To furnish the development, we provide some background materials on macroscopic traffic flow modeling.

Table 2.1. Summary of characteristics for the widely used models in traffic forecasting (Vlahogianni et al., 2004)

		Parametric modeling			Non-parametric modeling	
		Smoothing	ARMA	Kalman filtering	Non-parametric regression	Neural networks
Hypothesis on the statistical nature of data		Deterministic	Stochastic	Stochastic Gaussian nature of initial conditions	Non-linear deterministic	Not required
Hypothesis on the temporal regularity (stationarity)		Regular	Weak stationary	Not required	Not required	Not required
Hypothesis on linearity or non-linearity		Linear	Input parameter*	Input parameter*	Input parameter*	Not required
Multivariate modeling		impossible	difficult	straightforward	difficult	Easily incorporated
DATA	Quantity	short series	extensive	extensive	extensive	extensive
	Quality	continuity	continuity	continuity	Not required	Not required
Results	Extraction	Explicit	Explicit	Explicit	Explicit	Implicit
	Accuracy	Low	Low but acceptable	Medium	High	Best
Nature of predictions	Static	Static	Recursive	Static	Updated through training	
Main advantages		Short series needed	Well-established theoretical background	Multivariate nature	Simple model structure	Non-stationary and non-linear environment, wide mapping capabilities
Main disadvantages		Relatively low accuracy	Weak stationarity, low accuracy in extreme values	Gaussian hypothesis	Non-linear deterministic nature, data intensive	Data intensive, complex internal structure

* Input parameter: the decision of linear or non-linear modeling must be predefined because it largely affects model structure

2.3 Macroscopic traffic flow model with physical queue

2.3.1 Basic definitions, quantities and requirements

The macroscopic dynamic traffic flow models attempt to replicate the aggregate response of a large number of vehicles while ignoring the behavior of the individual driver. These models represent traffic as a compressible fluid[5], in terms of traffic flow, density, and speed. Almost all the analytical dynamic traffic assignment (DTA) problems, nearly all model-based on-ramp metering control designs, and practical traffic engineering have adopted macroscopic modeling of traffic flow. As a result, dynamic traffic flow models are essential for dynamic traffic surveillance, assignment, and control. In particular, we rely on macroscopic dynamic traffic flow models to conduct (indirect) travel time estimation and prediction in this dissertation. We thus review some related fundamental quantities and a fundamentally important macroscopic traffic flow model. According to the HCM (2000), the following quantities are defined.

• Speed $v(y, t)$ is defined as a rate of motion expressed as distance per unit of time, where y, t represents position (measured in the direction of traffic flow) and time, respectively. Depending on how it is measured, $v(y, t)$ is referred to as either space mean speed or time mean speed (HCM, 2000). The other speed concept used in dynamic traffic flow models is the so-called free-flow speed, which is defined as the average speed of traffic measured under light conditions so that vehicles can move freely at their desired speed.

• Flow $f(y, t)$ is defined as the total number of vehicles that pass by the measure point y during a given time interval including t, divided by the length of the time interval.

• Density $\rho(y, t)$ is the number of vehicles occupying a (unit) length of roadway around

[5] The flow is incompressible when the density reaches the jam density .

the measure point y at time t. The term can be obtained by the flow-density relationship, i.e.

$$\rho(y,t) = \frac{f(y,t)}{v(y,t)}. \tag{2.1}$$

The flow-density relationship of a freeway segment is often referred to as the fundamental flow-density diagram (or fundamental diagram for short) of the segment. Depending on how we specify the speed-density relationship, the fundamental diagram can have different shapes. The Greenshields' quadratic fundamental diagram and the Newell's triangular fundamental diagram are two most common used fundamental diagrams.

Some restrictions are imposed upon the dynamic marcoscopic traffic flow models to describe the link traffic dynamics, travel time of the link, and flow propagation. Among these restrictions: non-negativity of traffic flow, flow conservation, flow propagation, the first in/first out (FIFO) principle, causality, continuity are the common requirements. Here, we provide a brief description on flow conservation, the FIFO principle and the travel time definition based on these restrictions. For the sake of clarity, we first define the flow conservation.

Flow Conservation: The conservation of traffic flow is one of the most important requirements that dynamic traffic flow models should possess. The conservation is enforced to prevent the situations that travelers enter the network vanish before reaching the destination during the planning horizon, or the total outflow exceeds the total inflow to a link at any time instance. As depicted in Figure 2.2, the flow conservation for a link a can be expressed as follows:

$$A_a(t) = D_a(t) + x_a(t), \tag{2.2}$$

where $A_a(t)$ denotes the cumulative arrivals up to time t and $D_a(t)$ is the cumulative departures up to time t, respectively. This equation states that cumulative traffic volume on link a at time t is equal to the difference between cumulative arrivals

and cumulative departures up to time t from the initial time t_0 with initial condition $x_a(t_0) = 0$. For the case that link a is not empty at the initial moment, we can revise (2.2) as

$$A_a(t) + x_a(t_0) = D_a(t) + x_a(t).$$

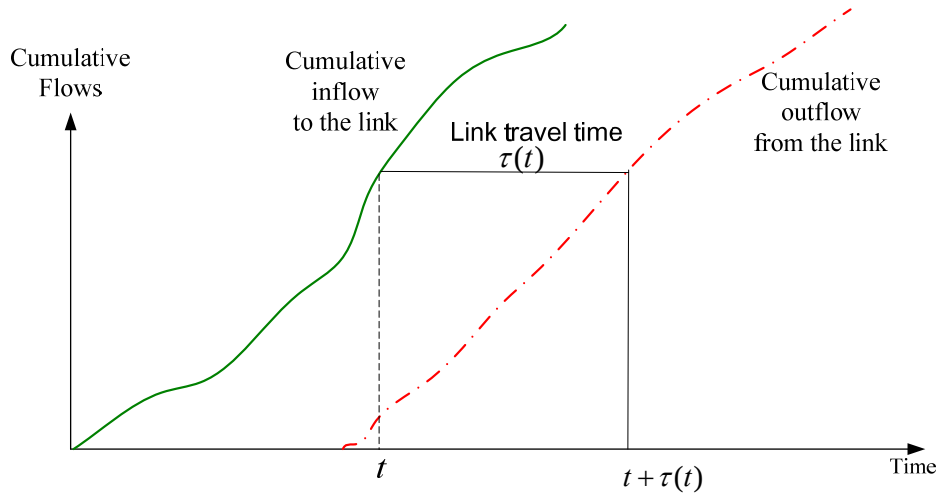


Figure 2.2. Illustration of travel time calculation under FIFO principle

Definition 2.1 First-In-First-Out (FIFO) Principle: Roughly speaking, the FIFO principle requires that vehicles entering a link first must also exit from the link first. Mathematically, this condition takes the form,

$$t' > t'' \Rightarrow t' + \tau(t') > t'' + \tau(t''), \tag{2.3}$$

where $\tau(t)$ is the link travel time for vehicles entering the link at time t.

Definition 2.2 Link travel time: Given that the FIFO principle hold, the travel time$\tau(t)$ for link a at time t can be deduced from the time-flow consistency equation as follows

$$A_a(t) = D_a\big(t + \tau(t)\big). \tag{2.4}$$

Nie and Zhang (2005) provided an overview on the role of FIFO principle in modeling network traffic dynamics. Note from Equation (2.2), the FIFO may imply no overtaking in practice. Therefore, as explained in Mun (2007), in traffic engineering practice the FIFO principle can not be always fulfilled. However, the voilation of

FIFO principle may be permitted in microscopic models through overtaking since we can specify different characteristics for different vehicles even though they entered a link at the same time. Different from microscopic models, in macroscopic (or aggregate) models vehicles are considered to take the same travel time to traverse a link if they enter it at the same time since macroscopic (or aggregate) models describe the average (statistical) behavior of vehicles (Mun, 2007). The FIFO principle is therefore should be enforced in dynamic assignment models for the issue of equity (Mun, 2007). If the FIFO principle is enforced for a traffic flow model, we have the following proposition.

Proposition 2.1 (Chow, 2007) If a traffic model satisfies the FIFO principle and the link travel time function $\tau(t)$ is differentiable, then the following condition will be satisfied

$$1+\frac{d\tau(t)}{dt}\geq 0, \tag{2.5}$$

for all times of entry t to the link.

2.3.2 The Lighthill-Whitham-Richards (LWR) model

Among the macroscopic traffic flow models, Lighthill-Whitham-Richards (LWR) model would be the most popular and most-cited one. In terms of fluid dynamics, the traffic dynamics modeled by the LWR model is governed by the following two equations.

$$\begin{aligned}&\frac{\partial\rho(y,t)}{\partial t}+\frac{\partial f(y,t)}{\partial y}=0,\\&f(y,t)=F\big(\rho(y,t)\big).\end{aligned} \tag{2.6}$$

The first equation of (2.6) is the principle of conservation of vehicles, which is followed from fluid mechanics. The second equation of (2.6) is a flow-density relationship which is also known as the “fundamental (flow-density) diagram”. As a “fluid-dynamic” traffic flow model, since the LWR model does not contain a

second-order derivative (such as a diffusion term), it is classified into the category of first-order model.

Detailed discussions on the LWR model can be found in Haberman (1977); Gomes (2004); Schönhof and Helbing (2007). The LWR model is capable of reproducing many important phenomena of freeway traffic. For instance, it captures the main difference between free-flow and congested traffic, which is that they propagate small disturbances in opposite directions, e.g. forward and backward waves, and at different speeds, e.g. free-flowing speed and backward wave speed. The LWR model also explains the formation and dissipation of queues upstream of a bottleneck, the dynamics of deceleration shock waves, and the absence of naturally forming acceleration shock waves (Gomes, 2004). However, in the meantime, the model is criticized for predicting some unrealistic traffic behavior, e.g. unable to describe unstable flow, unable to describe spontaneous breakdowns of traffic flow, steady (or static) flow-density fundamental diagram assumption (Gomes, 2004; Schönhof and Helbing, 2007) .

To allow source terms which may be due to ramp flows, the traffic dynamics of a freeway segment modeled by the LWR model is governed by

$$\begin{aligned} &\frac{\partial \rho(y,t)}{\partial t} + \frac{\partial f(y,t)}{\partial x} = \nu_{+}(y,t) - \nu_{-}(y,t), \\ &f(y,t) = F\big(\rho(y,t)\big), \end{aligned} \tag{2.7}$$

where $\nu_{\pm}(x,t)$ denotes the source terms which may be due to ramp flows with the plus sign denotes on-ramps and the minus sign denotes off-ramps. There are several ways to introduce stochastic elements to the LWR modeling framework, e.g.

1. stochastic initial and boundary conditions,
2. stochastic source terms, and
3. stochastic speed-density relationship or fundamental diagram.

We may recognize the forst two terms as demand uncertaity and the third term as supply uncertainty.

Chapter 3

The stochastic cell transmission model (SCTM) and its extensions

Firstly, this chapter reviews the previous works related to the development of the cell transmission model (CTM) from the deterministic nonlinear CTM to its extensions such as the Modified CTM (MCTM) and its piecewise linearized version---Switch Mode Model (SMM). Next, SCTM is introduced from the cell assignment with respect to the characteristics of freeways, to the evaluation stochastic dynamic traffic flow propagation. Finally, advantages/disadvantages of the model are listed with the cautions needed to be paid when implementing the SCTM for virtual traffic flow simulation and empirical study.

3.1 The cell transmission model and its extensions

3.1.1 The CTM and MCTM

The Cell Transmission Model (CTM), which was proposed by Daganzo (1994, 1995) is a deterministic dynamic traffic flow model which discretizes the LWR model in both time and space. As shown in Figure 3.1, the model devides a freeway segment into several homogeneous, consecutively numbered cells with length l_i, where i is a cell index. Idealy, the cell length should be equal to the distance travelled by free flowing vehicles during one simulation time increment, that is $l_i = v_f T_s$, where v_f is the free-flow speed of the freeway segment, and T_s is the simulation time increment. The model assumes that the traffic state is homogenous within one cell. The number of vehicles on cell i at time $k\Delta t$ which is taken as the state variable for the cell can be evaluated according to the following flow conservation equation:

$$n_i(k+1) = n_i(k) + \left(y_i(k) - y_{i+1}(k)\right) \tag{3.1}$$

where $y_i(k)$ is the inflow volume from cell *i*-1 to *i* during time interval $\left[kT_s, \left(k+1\right)T_s\right)$. The flow from cell *i* -1 into cell *i* is the minimum of three quantities:

$$y_i(k) = \min\left\{n_{i-1}(k), Q_i, (w_c / v_f)\left[N_i - n_i(k)\right]\right\} \tag{3.2}$$

where N_i is the maximum number of vehicles that can be present in cell *i with* $N_i = \rho_J l_i$; Q_i is the maximum number of vehicles that can flow into cell *i* during time interval $\left[kT_s, \left(k+1\right)T_s\right)$ with $Q_i = Q_M T_s$; w_c is the backward wave speed when traffic is congested (e.g. mile/hr); and v_f is the free-flow speed (e.g. mile/hr). $(w_c / v_f)\left[N_i - n_i(k)\right]$ is the jam-limited volume, which depends on the amount of available space of cell *i*. In this case, the CTM corresponds to a trapezoidal fundamental diagram shape, as shown in Figure 3.2., where Q_M is the maximum allowable flow rate (e.g. veh/hr), ρ_J is the jam density (e.g. veh/mile) and ρ_c is the critical density (e.g. veh/mile).

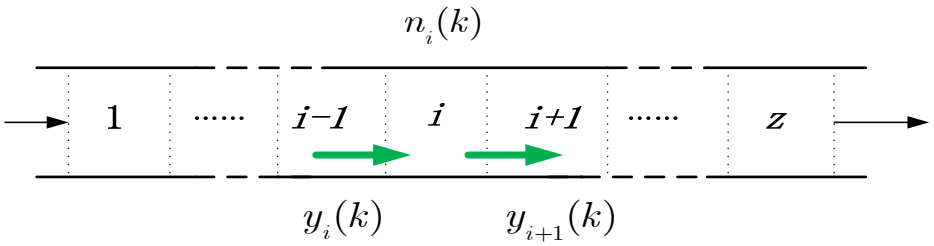


Figure 3.1. Cell partition of a freeway segment

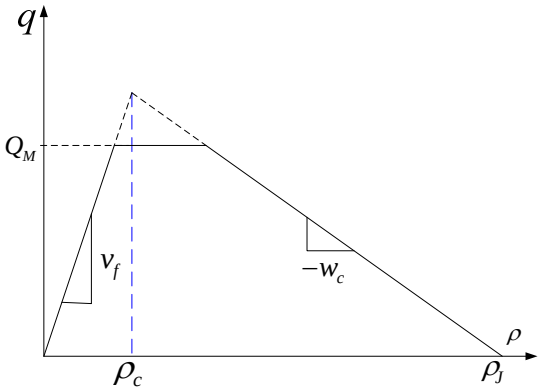


Figure 3.2. A trapezoidal fundamental diagram for CTM and MCTM

For traffic surveillance and control purpose, traffic density may be a more natural mearsure as it characterizes the level of congestion. For example, the mapping from traffic volume to traffic density is a multi-valued one as illustrated by the fundamental diagram, e.g. Figure 3.2. A flow rate of 5000 veh/hr may correspond to either free-flowing traffic condition (i.e. the left hand side of the fundamental diagram) or congested traffic condition (i.e. the right hand side of the fundamental diagram). On the other hand, the mapping from traffic density to traffic condition is unique. To this end, the modified cell transmission model (MCTM) was developed by Munoz et al. (2003) for traffic surveillance. Moreover, this model permits the CTM to adopt non-uniform cell lengths and leads to greater flexibility in partitioning freeways. In the MCTM, the density of cell *i* evolves as:

$$\rho_i(k+1) = \rho_i(k) + \frac{T_s}{l_i}\Big(q_{i,in}(k) - q_{i,out}(k)\Big), \tag{3.3}$$

where $\rho_i(k)$ is the vehicle density in cell *i* at time index *k*, $q_{i,in}(k)$ and $q_{i,out}(k)$ are the flow rates (in vehicles per unit time) entering and leaving cell *i* during the time interval $\left[kT_s, \left(k+1\right)T_s\right)$ respectively. The model defines some piecewise affine sending and receiving functions of traffic volumes to describe the interactions between neighboring cells as well as the shockwaves of a freeway segment. $q_i(k)$ is determined by taking the minimum of two quantities:

$$q_{i,in}(k) = \min(S_{i-1}(k), R_i(k)) \tag{3.4}$$

where $S_{i-1}(k)$ is the maximum flow that can be supplied by cell *i*-1 under free-flow conditions, over the interval $\left[kT_s, \left(k+1\right)T_s\right)$, and $R_i(k)$ is the maximum flow rate that can be received by cell *i* under congested conditions, over the same time interval:

$$S_{i-1}(k) = \min(v_{f,i-1}\rho_{i-1}(k), Q_{M,i-1}) \tag{3.5}$$

$$R_i(k) = \min(Q_{M,i}, w_{c,i}(\rho_{J,i} - \rho_i(k))) \tag{3.6}$$

From (3.4)-(3.6),

$$q_{i,in}(k) = \min(v_{f,i-1}\rho_{i-1}(k), Q_{M,i-1}, Q_{M,i}, w_{c,i}(\rho_{J,i} - \rho_i(k))) \tag{3.7}$$

The definitions of sending and receiving functions are useful when the model is extended to tackle general network topologies (Daganzo, 1995; Munoz et al., 2003). The CTM was further extended to track the path flows for the purpose of dynamic traffic assignment by Lo et al. (2001). As a discrete version of the LWR model, the CTM suffers from most of the drawbacks of the LWR model in describing traffic flow.

3.1.2 Switching Mode Model (SMM)

Although the MCTM is much simpler than many other higher order hydrodynamics based partial differential models, the nonlinear nature of the flow-density relationship due to (3.7) still makes it difficult to analyze and to be used as a basis for the design of traffic controllers. To avoid the nonlinearity, the Switching Mode Model (SMM) was proposed (Munoz et al. 2003). The SMM is a hybrid system (switched linear system) that switches among different sets of linear difference equations (representing different traffic states of the highway), and the mainline boundary data and the congestion status of the cells in a highway section detemine which equation is adopted (see Figure 3.4). The formulation of SMM avoids the nonlinearity of the CTM at the cost of introducing the "switching" condition and using triangular flow-density relationship. The triangular relatioship means $Q_{M,i}(\frac{1}{v_{f,i}}+\frac{1}{w_{c,i}})=\rho_{J,i}$ for each cell (i=1,2,….,z), as shown in Figure 3.3. As stated in Munoz *et al.* (2003), the following assumptions are made to simplify the problem:

Assumption 3.1 The traffic flow rates and densities of the segment upstream and downstream boundaries, as well as the flow rates of on-/off- ramps are known.

Assumption 3.2[1] There is at most one wave-front within the freeway segment.

[1] Some authors, e.g. Perakis and Roles (2006), refer this assumption as "At Most One Shock" which is stated as: there is at most one shock on the road, dividing an upstream uncongested region from a downstream congested region.

Based on the above assumptions, five modes are defined according to different traffic states for each segment(see Figure 3.4): (I) “Free flow-Free flow (FF)” (Figure 3.4a), (II) “Congestion - Congestion (CC)” (Figure 3.4b), (III) “Congestion - Free flow (CF)” (Figure 3.4c), (IV) “Free flow-Congestion 1 (FC1)” (Figure 3.4d), and (V) “Free flow - Congestion 2 (FC2)” (Figure 3.4e). The two “Free flow – Congestion” modes are determined by the moving direction of the wave front. If the wavefront is moving downstream, then the mode is FC1; while if the wave front is moving upstream, it is FC2 mode. The moving direction of the wave-front depends on the relative magnitudes of the supplied flow from the last uncongested cell (*cell l*-1) and the available space in the first congested cell downstream (*cell l*). As presumded in Assumption 3.2, there is at most one wave front in the freeway segment, modes (I) and (II) can be regarded as steady-state modes, while modes (III), (IV), and (V) can be viewed as transient modes with the wave front located on the boundary of cell *l*-1 and *l* (*l*=1,…, *z*-1).

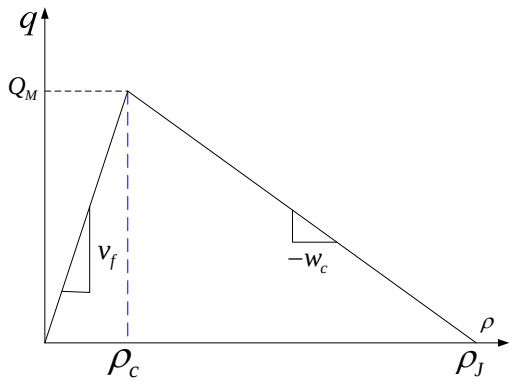


Figure 3.3. A triangular fundamental diagram for the SMM

In Figure 3.4, $\tilde{q}_u$ and $\tilde{\rho}_u$ are the inflow rate and density measured by the detectors at the upstream boundary, respectively， and $\tilde{q}_d$ and $\tilde{\rho}_d$ are the outflow rate and density measured at the downstream boundary, respectively. r_b denotes the on-ramp flow merges into the mainline at the upstream boundary of cell *b*, and f_e denotes the off-ramp flow diverts from the the mainline at the downstream boundary of cell *e*. In Munoz *et al*. (2003), $\tilde{\rho}_u(k)$, $\tilde{\rho}_d(k)$ in conjunction with the estimated densities

$\overline{\rho}_i(k)$ (if necessary) are applied to determine under which mode the freeway segment is operating and the location of wave-front with moving direction:

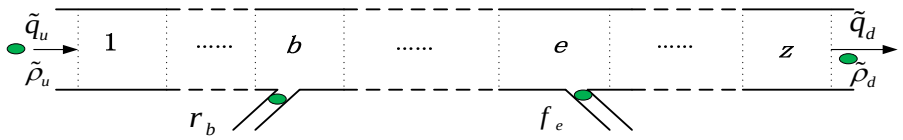
If $\tilde{\rho}_u(k) < \rho_{c,1}$ and $\tilde{\rho}_d(k) < \rho_{c,z}$, then mode FF is determined at time index *k*.

If $\tilde{\rho}_u(k) > \rho_{c,1}$ and $\tilde{\rho}_d(k) > \rho_{c,z}$, then mode CC is determined.

If $\tilde{\rho}_u(k) > \rho_{c,1}$ and $\tilde{\rho}_d(k) < \rho_{c,z}$, then mode CF is determined, and if furthermore $\overline{\rho}_j(k) > \rho_{c,j}$ and $\overline{\rho}_{j+1}(k) < \rho_{c,.j+1}$, then *l*=*j*, i.e. the wave-front is located at the boundary between cell *j* and cell *j*+1.

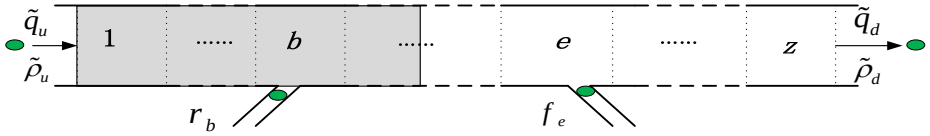
If $\tilde{\rho}_u(k) < \rho_{c,1}$ and $\tilde{\rho}_d(k) > \rho_{c,z}$, then mode FC is determined, and if furthermore $\overline{\rho}_j(k) < \rho_{c,j}$ and $\overline{\rho}_{j+1}(k) > \rho_{c,.j+1}$, then *l*=*j*;

Moreover, if $v_{f,j}\overline{\rho}_j(k) < w_{c,j+1}(\rho_{J,j+1} - \overline{\rho}_{j+1}(k))$, then FC1 is activated; else FC2 is activated.

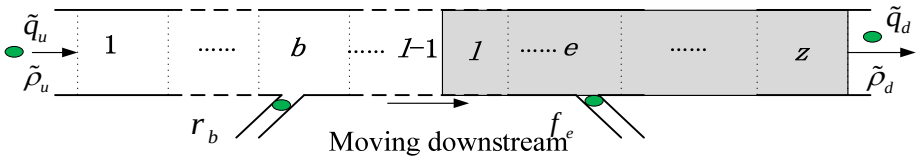


a). Free-flow-Free-flow (FF) mode

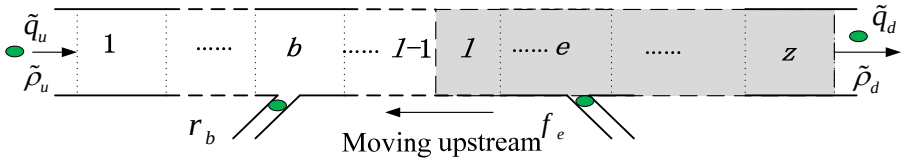
b). Congestion to Congestion (CC) mode



c). Congestion to Free-flow (CF) mode.



d). Free-flow to Congestion (FC 1) mode



e). Free-flow to Congestion (FC 2) mode

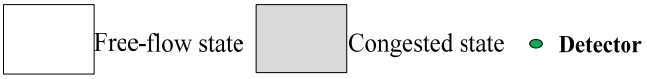


Figure 3.4. Freeway segment with *z* cells for SMM and SCTM

Because of the deterministic environment, there must be one mode and only one mode that represents the traffic state at a certain time. Each mode corresponds to a discrete-time linear system of the following form:

$$\overline{\rho}(k+1) = A_{s,l}\overline{\rho}(k) + B_{s,l}u(k) + B_{J,s,l}\rho_J + B_{Q,s,l}q_M \tag{3.8}$$

where *s* denotes the mode *FF, CC, CF, FC1* and *FC2*, and *l* denotes the location of wave-front, the vector of traffic densities of the freeway segment is denoted as $\overline{\rho}(k) = \left(\overline{\rho}_1(k),\ldots,\overline{\rho}_z(k)\right)^T$, and $u(k) = \left(q_u(k), r_b(k), f_e(k), q_d(k)\right)^T$ is comprised of inflow and outflow rates including the on-/off- ramp flows. The vectors ρ_J, q_M are jam densities and capacities of all the cells along the freeway segment, that is $\rho_J = \left[\rho_{J,1}, \rho_{J,2},\cdots \rho_{J,z}\right]^T$ and $q_M = \left[Q_1, Q_2,\cdots Q_z\right]^T$ .The matrices $A_{s,l}, B_{s,l}, B_{J,s,l},$ and $B_{Q,s,l}$ are system matrices of mode *s* with the subscript *l* denotes the wave-front location. For example, when *z*=4, *s*=*FC*2, *l*=2, *r*=2, *e*=3, the system matrices are thus defined:

$$A_{FC1,2} = \begin{bmatrix} -\frac{v_1 T_s}{l_1} & & & \\ \frac{v_1 T_s}{l_2} & -\frac{v_2 T_s}{l_2} & & \\ & \frac{v_2 T_s}{l_3} & 0 & -\frac{w_4 T_s}{l_3} \\ & & & -\frac{w_4 T_s}{l_4} \end{bmatrix}, B_{FC1,2} = \begin{bmatrix} \frac{T_s}{l_1} & & \\ & \frac{T_s}{l_2} & \\ & & -\frac{T_s}{l_3} \\ & & -\frac{T_s}{l_4} \end{bmatrix},$$

$$B_{J,FC1,2} = \begin{bmatrix} 0 & & \\ & 0 & \\ & & 0 \\ & & \frac{w_4 Ts}{l_4} \end{bmatrix}, B_{Q,FC1,2} = 0_{4\times 4}.$$

From the introductions of the CTM, the MCTM and the SMM, all of these models rely on the steady supply functions as well as the assumption that the inflow demand and supply functions are deterministic. However, as explained in the previous chapters, this assumption is not realistic due to the inherent demand and supply uncertainites. The Monte Carlo method is a simple way to extend the SMM to handle the demand and supply stochasticities. However, the potentially high computational cost and the techniques requiured to reduce the variances of the results may be the apparent disadvantages and difficulties. In order to model the effects of demand and supply uncertainties on traffic flow dynamics and its propagation, a stochastic dynamic traffic flow model that extends the SMM to consider stochastic parameters of the fundamental flow-density diagram as well as the stochastic travel demand was proposed by Sumalee *et al*. (2011). This model is titled as the stochastic cell transmission model, which is introduced and discussed in the following sections.

3.2 The stochastic cell transmission model (SCTM)

The stochastic cell transmission model (SCTM) proposed by Sumalee *et al.* (2011) is a stochastic dynamic traffic flow model for the traffic state estimation. The random demand profiles and stochastic supply functions, e.g. the variations of day-to-day inflow demand as illustrated in the left hand side of Figure 3.5 (inflow demand at non-rush hour on seven days) and uncertain parameters in the flow-density fundamental diagrams as illustrated in the right hand side of Figure 3.5 (flow-density data collected for 54 days), are taken as exogenous inputs to the SCTM. The SCTM evaluates the stochastic dynamic traffic states as demonstrated in Figure 3.6. We interpret it by the following three steps:

Step 1: The model defines five probabilistic events corresponding to different congestion status as shown in Figure 3.4 which have been described as operational modes and explained in details in previous SMM section. As the traffic states and

supply functions are both stochastic, therefore more than one traffic modes might exist simultaneously instead of just only one mode at each time in the deterministic case. Using real-time detection or historical data, the model evaluates the probabilities of the five modes.

Step 2: Evaluate the mean and auto-correlation of cell traffic densities for each mode based on the corresponding bilinear system formulation given the statistics of random inflow profiles, supply functions, and the traffic states evaluated in last iteration.

Step 3: Approximate the stochastic traffic densities (mean and standard deviation) by applying finite mixture to the five distributions and relevant probabilities obtained from Steps 1 and 2.

These three steps will be introduced in details in the following sections. In this chapter, demand and supply uncertainties are assumed to be independent Gaussian white noises in both time and space domains.

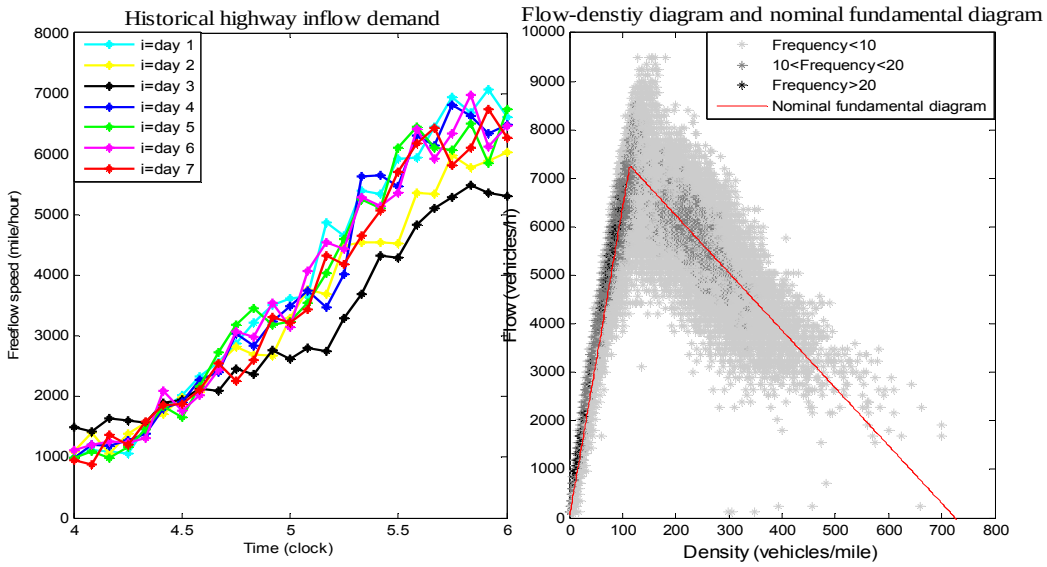


Figure 3.5. The stochastic inflow demand and fundamental diagram

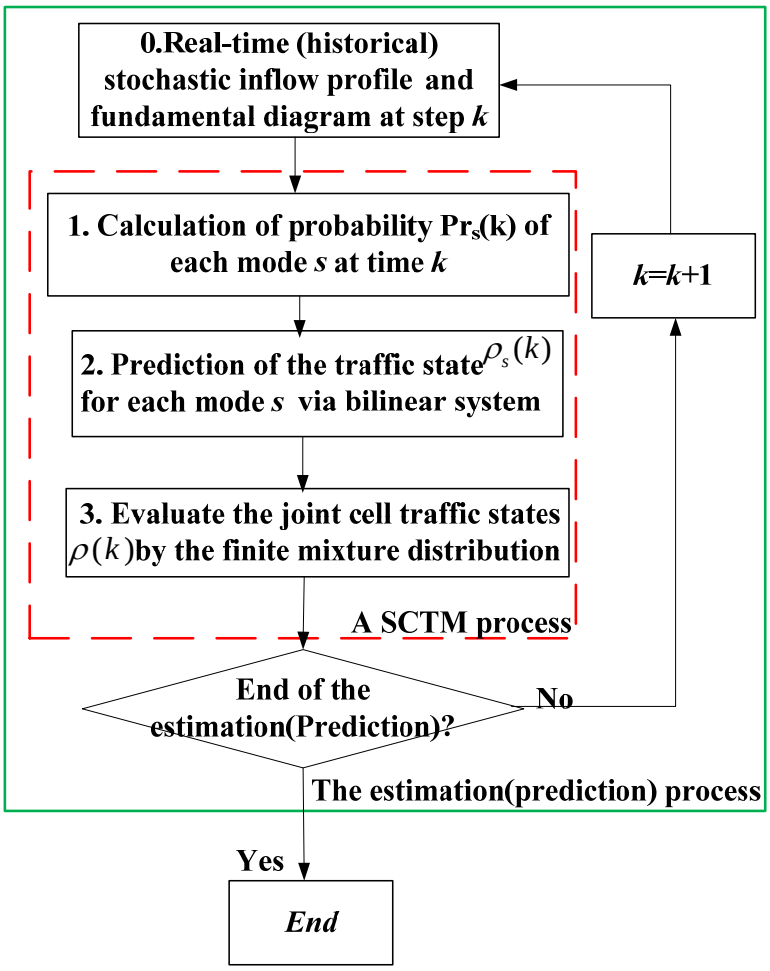


Figure 3.6. The implementation process of the SCTM

3.3 Probabilities of occurrence of modes

The SCTM adopts the five traffic operational modes introduced in the SMM to represent the traffic state under different congestion levels, i.e. FF, CC, CF, FC1 and FC2. However, due to the demand and supply uncertainties, we cannot determine under which exact mode the freeway segment is operating. Therefore, in the SCTM, we define several probabilities to describe the likehood of occurrence of these operational modes. In this chapter, the probabilities of occurrence of these five modes are denoted as $\Pr_s(k)$ with *s= FF, CC, CF, FC1, FC2.*

In Munoz (2003), the freeway between two neighboring detector stations is assigned

to be one SMM segement by enforcing Assumption 3.2. Actually, this assumption might be voilated especially when the length of the segment is long, e.g. 5 kilometers which may be divided into 10 cells. In Suamlee et al. (2011) and Zhong (2011), two kinds of assignments are proposed for choice to relax the one wave-front assumption. The first appropach is to assign multiple cells (generally more than 2 cells) between two neighboring detectors as one system as shown in Figure 3.4. The other one is to represent the freeway segment by cascaded sub-systems with each sub-system consisting of two cells only as shown in Figure 3.7.

3.3.1 Multiple cells case

The probabilities of occurrence of the four main operational modes can be defined as:

- FF mode: $P_{FF}(k) = \Pr\left(\tilde{\rho}_u(k) < \rho_{c,1} \cap \tilde{\rho}_d(k) < \rho_{c,z}\right),$ (3.9)
- CC mode: $P_{CC}(k) = \Pr\left(\tilde{\rho}_u(k) \geq \rho_{c,1} \cap \tilde{\rho}_d(k) \geq \rho_{c,z}\right),$
- CF mode: $P_{CF}(k) = \Pr\left(\tilde{\rho}_u(k) \geq \rho_{c,1} \cap \tilde{\rho}_d(k) < \rho_{c,z}\right),$
- FC mode: $P_{FC}(k) = 1 - (P_{FF}(k) + P_{CC}(k) + P_{CF}(k)),$

These definitions are proposed based on the assumption that there is no more than one wave-front in a SCTM segment. A SCTM segment is freeway section between adjacent detectors along the test site.The physical meannings of these probabilities can be thus explained, e.g. the probability of FF mode can be regarded as the likihood that the upstream detected density is less then the critical density of the first cell and the downstream detected density is less than the critical density of the last cell. The following step seperates the *CF* and *FC* modes by the location and moving direction of wave-front.

For CF mode, the conditional probability $P_{CF,l|CF}(k)$, that is the wave-front located at the boundary between cells l and l+1 (l=1,2,……, z-1) based on the CF mode as

$$P_{CF,l|CF,O}(k) = \Pr\left(\bar{\rho}_l(k) \geq \rho_{c,l} \cap \bar{\rho}_{l+1}(k) < \rho_{c,l+1}\right), \quad (3.10)$$

But the summation of conditional probabilieis $P_{CF,l|CF,O}(k)$ for l=1,……,z-1 might not be equal to 1. A weighting method is applied to overcome this disadvantage:

$$P_{CF,l|CF}(k)=\frac{P_{CF,l|CF,O}(k)}{\sum_{l=1}^{z-1}P_{CF,l|CF,O}(k)} \tag{3.11}$$

Then

$$P_{CF,l}(k)=P_{CF,l|CF}(k)\times P_{CF}(k). \tag{3.12}$$

FC mode involves both the location and wave-front moving direction problems. The conditional probability $P_{FC1,l|FC}(k)$, i.e. the wave-front located at the boundary between cells l and l+1 (l=1,2,……,z-1), with the wave-front moving downstream is defined as

$$P_{FC1,l|FC,O}(k)\triangleq \Pr\left(\overline{\rho}_l(k)<\rho_{c,l}\cap\overline{\rho}_{l+1}(k)\geq\rho_{c,l+1}\cap v_{f,l}\overline{\rho}_l(k)\leq w_{l+1}(\rho_{J,l+1}-\overline{\rho}_{l+1}(k))\right),$$

and the probability of the wave-front moving upstream as

$$P_{FC2,l|FC,O}(k)\triangleq \Pr\left(\overline{\rho}_l(k)<\rho_{c,l}\cap\overline{\rho}_{l+1}(k)\geq\rho_{c,l+1}\cap v_{f,l}\overline{\rho}_l(k)>w_{l+1}(\rho_{J,l+1}-\overline{\rho}_{l+1}(k))\right),$$

Similarly, the summation of $P_{FC1,l|FC,O}(k)$ and $P_{FC2,l|FC,O}(k)$ might not be equal to 1, so the weighting method is utilized:

$$P_{FCx,l|FC}(k)=\frac{P_{FCx,l|FC,O}(k)}{\sum_{l=1}^{z-1}\left(P_{FC1,l|FC,O}(k)+P_{FC2,l|FC,O}(k)\right)},\quad x=1,2 \tag{3.13}$$

Then, $$P_{FCx,l}(k)=P_{FCx,l|FC}(k)\times P_{FC}(k)$$

where $\rho_{c,i}$ is the critical density of cell i, v_f is the free flow speed, w_i is the backward congestion wave speed, $\rho_{J,i}$ is the jam density of cell i, $\overline{\rho}_i$ is the joint density of cell i, which is approximated by the finite mixture distribution of the five probabilistic events as it will be discussed in details in Section 3.3.3.

3.3.2 Two cells case

As previously mentioned, a freeway can be also divided into several subsystems with each subsystem consists of two cells as depicted in Figure 3.7. The essential advantage of this approach is that the description of each mode is simpler and Assumption 3.2 is automatically satisfied due to the homogeneous assumption within one cell, e.g. the wave-front location within each subsystem for transient modes (*CF*, *FC*1 and *FC*2) is not varying any more (it has to located at the boundary between the two cells). This in turn yields definitions of probabilities of occurrence of the five operational modes easy to be evaluated.

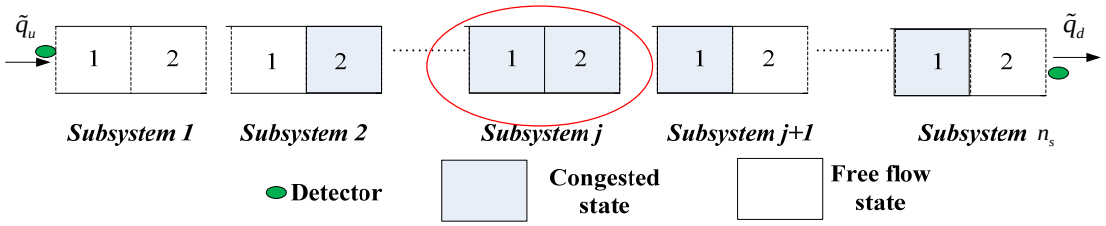


Figure 3.7. Subsystems and traffic states

Probabilities of occurrence of operational modes of subsystem *j* are given by:

FF: $P_{FF(j)}(k) = \Pr\left(\overline{\rho}_{j,1}(k) < \rho_{c,j,1} \cap \overline{\rho}_{j,2}(k) < \rho_{c,j,2}\right).$

CC: $P_{CC(j)}(k) = \Pr\left(\overline{\rho}_{j,1}(k) > \rho_{c,j,1} \cap \overline{\rho}_{j,2}(k) > \rho_{c,j,2}\right).$

CF: $P_{CF(j)}(k) = \Pr\left(\overline{\rho}_{j,1}(k) > \rho_{c,j,1} \cap \overline{\rho}_{j,2}(k) < \rho_{c,j,2}\right).$

*FC*1: $P_{FC1(j)}(k) = \Pr\left(\overline{\rho}_{j,1}(k) < \rho_{c,j,1} \cap \overline{\rho}_{j,2}(k) > \rho_{c,j,2} \cap\ v_{f,j,1}\overline{\rho}_{j,1}(k) < w_{c,j,2}(\rho_{J,1} - \overline{\rho}_{j,1}(k))\right)$

*FC*2: $P_{FC2(j)}(k) = \Pr\left(\overline{\rho}_{j,1}(k) < \rho_{c,j,1} \cap \overline{\rho}_{j,2}(k) > \rho_{c,j,2} \cap\ v_{f,j,1}\overline{\rho}_{j,1}(k) > w_{c,j,2}(\rho_{J,1} - \overline{\rho}_{j,1}(k))\right)$

3.3.3 Independent assumption and probability evaluation

The SCTM framework proposed by Sumalee et al. (2011) and Zhong (2011) has enforced the uncorrelated assumption. Take Equation (3.9) as an example, the joint

probability is evaluated as

$$P_{FF}(k) = \Pr\left(\tilde{\rho}_u(k) < \rho_{c,1} \cap \tilde{\rho}_d(k) < \rho_{c,z}\right) = \Pr\left(\tilde{\rho}_u(k) < \rho_{c,1}\right)\Pr\left(\tilde{\rho}_d(k) < \rho_{c,z}\right), \quad (3.14)$$

Equation (3.14) means that the event $\tilde{\rho}_u(k)$ is smaller than $\rho_{c,1}$ will not be influnced by the relationship between $\tilde{\rho}_d(k)$ and $\rho_{c,z}$. However, this assumption may not be true in practice which will be discussed in Chapter 5.

3.4 The bilinear system and finite mixture

The dynamics of each mode is represented by a bilinear system of the following form:

$$\rho_s(k+1) = \left(A_{s,0} + \sum_{i=1}^{p} A_{s,i}\omega_{s,i}(k)\right)\overline{\rho}(k) + \left(B_{s,0} + \sum_{i=1}^{p} B_{s,i}\omega_{s,i}(k)\right)\lambda_s(k) + B_s u(k), \quad (3.15)$$

where $A_{s,i}$ and $B_{s,i}$ $(i = 0,1,\cdots z)$ are the constant system matrices with the elements constituted by simulation time interval T_s and cell lengths l_i, $\omega_{s,i}(k)$ and $\lambda_s(k)$ denote the uncertain supply functions, and $u(k)$ is the vector of demand (or in-/out- flows of the segment).

Consider a freeway segment divided into four cells with one on-ramp connected to cell 2 and one off-ramp connected to cell 3, i.e. p=4, r=2, e=3. Denote the vector of cell traffic densities and vector of inflows as

$\overline{\rho}(k) = \left[\overline{\rho}_1(k), \overline{\rho}_2(k), \overline{\rho}_3(k), \overline{\rho}_4(k)\right]^T$, and $u(k) = \left[q_u(k), r_2(k), f_3(k), q_d(k)\right]^T$.

Under the *FF* mode $\omega_{FF,i}(k) = v_{f,i}$, $\lambda_{FF}(k) = 0_{4\times1}$, the supply uncertainty comes from variation of the free-flow speed. The system matrices for (3.15) under FF mode are

$$A_{FF,0}=I_{4\times4}\ ,\ A_{FF,1}=\begin{bmatrix}-\frac{T_s}{l_1} & & & \\ \frac{T_s}{l_2} & 0 & & \\ & 0 & 0 & \\ & & 0 & 0\end{bmatrix},\ A_{FF,2}=\begin{bmatrix}0 & & & \\ 0 & -\frac{T_s}{l_2} & & \\ & \frac{T_s}{l_3} & 0 & \\ & & 0 & 0\end{bmatrix},\ A_{FF,3}=\begin{bmatrix}0 & & & \\ 0 & 0 & & \\ & 0 & -\frac{T_s}{l_3} & \\ & & \frac{T_s}{l_4} & 0\end{bmatrix},$$

$$A_{FF,4}=\begin{bmatrix}0 & & & \\ 0 & 0 & & \\ & 0 & 0 & \\ & & 0 & -\frac{T_s}{l_4}\end{bmatrix},\ B_{FF}=\begin{bmatrix}\frac{T_s}{l_1} & & & \\ & \frac{T_s}{l_2} & & \\ & & 0 & \\ & & -\frac{T_s}{l_4} & 0\end{bmatrix}$$

Under *CC* mode, $\omega_{CC,i}(k)=w_{c,i}, \lambda_{CC}(k)=\left[\rho_{J,1},\ \rho_{J,2},\ \rho_{J,3},\ \rho_{J,4}\right]^T$, the backward wave speeds and the jam densities are the supply uncertainties. The system matrices for this mode are

$$A_{CC,0}=I_{4\times4}\ ,\ A_{CC,1}=\begin{bmatrix}-\frac{T_s}{l_1} & & & \\ 0 & 0 & & \\ & 0 & 0 & \\ & & 0 & 0\end{bmatrix},\ A_{CC,2}=\begin{bmatrix}0 & \frac{T_s}{l_1} & & \\ & -\frac{T_s}{l_2} & & \\ & & 0 & \\ & & & 0\end{bmatrix},$$

$$A_{CC,3}=\begin{bmatrix}0 & & & \\ & 0 & \frac{T_s}{l_2} & \\ & & -\frac{T_s}{l_3} & \\ & & & 0\end{bmatrix},\ A_{CC,4}=\begin{bmatrix}0 & & & \\ 0 & 0 & & \\ & 0 & 0 & \frac{T_s}{l_3} \\ & & 0 & -\frac{T_s}{l_4}\end{bmatrix}$$

$$B_{CC,0}=0_{4\times4}\ ,\ B_{CC,1}=-A_{CC,1}\ ,\ B_{CC,2}=-A_{CC,2}\ ,\ B_{CC,3}=-A_{CC,3}\ ,$$

$$B_{CC,4}=-A_{CC,4}\ ,B_{CC}=\begin{bmatrix}0 & \frac{T_s}{l_1} & & \\ & 0 & & \\ & & -\frac{T_s}{l_3} & \\ & & & -\frac{T_s}{l_4}\end{bmatrix}.$$

In *CF* mode, assume that the wave-front is located at the boundary of cell 2 and cell 3 ,

i.e. l=2. Then $\omega_{CF,2,i}(k) = w_{c,i}$, if i=1,2; $\omega_{CF,2,i}(k) = v_{f,i}$, if i=3,4, and $\lambda_{CF}(k) = \left[\rho_{J,1}, \rho_{J,2}, Q_{M,2}, 0\right]^T$

$$A_{CF,2,0} = I_{4\times4}\ ,\ A_{CF,2,1} = \begin{bmatrix} -\frac{T_s}{l_1} & & & \\ & 0 & & \\ & & 0 & \\ & & & 0 \end{bmatrix},\ A_{CF,2,2} = \begin{bmatrix} 0 & \frac{T_s}{l_1} & & \\ & -\frac{T_s}{l_2} & & \\ & & 0 & \\ & & & 0 \end{bmatrix},\ A_{CF,2,2} = \begin{bmatrix} 0 & \frac{T_s}{l_1} & & \\ & -\frac{T_s}{l_2} & & \\ & & 0 & \\ & & & 0 \end{bmatrix},$$

$$A_{CF,2,3} = \begin{bmatrix} 0 & & & \\ & 0 & & \\ & & -\frac{T_s}{l_3} & \\ & & \frac{T_s}{l_4} & 0 \end{bmatrix},\ A_{CF,2,4} = \begin{bmatrix} 0 & & & \\ & 0 & & \\ & & 0 & \\ & & & \frac{T_s}{l_4} \end{bmatrix}$$

$$B_{CF,2,0} = \begin{bmatrix} 0 & & & \\ & 0 & -\frac{T_s}{l_2} & \\ & & \frac{T_s}{l_3} & \\ & & & 0 \end{bmatrix},\ B_{CF,2,1} = -A_{CF,2,1}\ ,\ B_{CF,2,2} = -A_{CF,2,2}\ ,\ B_{CF,2} = \begin{bmatrix} 0 & -\frac{T_s}{l_1} & & \\ & 0 & & \\ & & 0 & \\ & & -\frac{T_s}{l_4} & 0 \end{bmatrix}$$

In *FC*1 mode, assume that the wave-front is located at the boundary of cell 2 and cell 3, i.e. l=2. In this case $\omega_{FC1,2,i}(k) = v_{f,i}$, *for* i=1,2; $\omega_{FC1,2,i}(k) = w_{c,i}$, *for* i=4; $\lambda_{FC1,2}(k) = \rho_{J,1}$.

$$A_{FC1,2,0} = I_{4\times4}\ ,\ A_{FC1,2,1} = \begin{bmatrix} -\frac{T_s}{l_1} & & & \\ \frac{T_s}{l_2} & 0 & & \\ & & 0 & \\ & & & 0 \end{bmatrix},\ A_{FC1,2,2} = \begin{bmatrix} 0 & & & \\ & -\frac{T_s}{l_2} & & \\ & \frac{T_s}{l_3} & 0 & \\ & & & 0 \end{bmatrix},$$

$$A_{FC1,2,4}=\begin{bmatrix}0 & & & \\ & 0 & & \\ & & 0 & \frac{T_s}{l_3} \\ & & & -\frac{T_s}{l_4}\end{bmatrix}, B_{FC1,2,4}=\begin{bmatrix}0\\0\\-\frac{T_s}{l_3}\\ \frac{T_s}{l_4}\end{bmatrix}, \quad B_{FC1,2}=\begin{bmatrix}\frac{T_s}{l_1} & & & \\ & \frac{T_s}{l_2} & & \\ & & 0 & \\ & & -\frac{T_s}{l_3} & \frac{T_s}{l_4}\end{bmatrix}$$

In *FC*2 mode, assume the wave-front is located at the boundary of cell 2 and cell 3, i.e. *l*=2. In this case $\omega_{FC1,2,i}(k)=v_{f,i}$, *for i=1;* $\omega_{FC1,2,i}(k)=w_{c,i}$, *for i=3,4;* $\lambda_{FC2,2}(k)=\left[\rho_{J,3},\rho_{J,4}\right]^T$.

$$A_{FC2,2,0}=I_{4\times4}\ , A_{FC2,2,0}=I_{4\times4}, A_{FC2,2,1}=\begin{bmatrix}-\frac{T_s}{l_1} & & & \\ \frac{T_s}{l_2} & 0 & & \\ & & 0 & \\ & & & 0\end{bmatrix},$$

$$A_{FC2,2,3}=\begin{bmatrix}0 & & & \\ & 0 & \frac{T_s}{l_2} & \\ & & -\frac{T_s}{l_3} & \\ & & & 0\end{bmatrix}, \quad A_{FC2,2,4}=\begin{bmatrix}0 & & & \\ & 0 & & \\ & & 0 & \frac{T_s}{l_3} \\ & & & -\frac{T_s}{l_4}\end{bmatrix},$$

$$B_{FC2,2,3}=\begin{bmatrix}0 & 0\\ -\frac{T_s}{l_2} & \\ \frac{T_s}{l_3} & \\ 0 & 0\end{bmatrix}, B_{FC2,2,4}=\begin{bmatrix}0 & 0\\ & 0\\ & -\frac{T_s}{l_3}\\ 0 & \frac{T_s}{l_4}\end{bmatrix}, B_{FC2,2}=\begin{bmatrix}\frac{T_s}{l_1} & & & \\ & \frac{T_s}{l_2} & & \\ & & 0 & \\ & & -\frac{T_s}{l_3} & \frac{T_s}{l_4}\end{bmatrix}$$

The mean of the density vector of mode *s*, i.e. $E\left(\rho_s(k+1)\right)$, can be evaluated as

$$E\left(\rho_s(k+1)\right)=\left(A_{s,0}+\sum_{i=1}^{p}A_{s,i}E\left(\omega_{s,i}(k)\right)\right)E\left(\overline{\rho}(k)\right)+\left(B_{s,0}+\sum_{i=1}^{p}B_{s,i}E\left(\omega_{s,i}(k)\right)\right)E\left(\lambda_s(k)\right)+B_sE\left(u(k)\right). \tag{3.16}$$

The auto-correlation matrix is:

$$\begin{aligned}Q_s(k+1)&=E\left(\rho_s(k+1)\rho_s(k+1)^T\right)\\&=\left[A_{s,0}+\sum_{i=1}^{p}A_{s,i}E\left(\omega_{s,i}(k)\right)\right]Q(k)\left[A_{s,0}+\sum_{i=1}^{p}A_{s,i}E\left(\omega_{s,i}(k)\right)\right]^T\\&\quad+\sum_{i=1}^{p}\gamma_{s,i}A_{s,i}Q(k)A_{s,i}^T+\sum_{\substack{i,j=1\\i\neq j}}^{z}\chi_{s,i,j}A_{s,i}Q(k)A_{s,j}^T+V_s(k),\end{aligned} \tag{3.17}$$

where

$$\begin{aligned}V_s(k)&=G_s(k)+G_s^T(k)+B_s\left(E\left(u(k)u^T(k)\right)\right)B_s^T\\&\quad+\left[B_{s,0}+\sum_{i=1}^{p}B_{s,i}E\left(\omega_{s,i}(k)\right)\right]\left(E\left(\lambda_s(k)\lambda_s^T(k)\right)\right)\left[B_{s,0}+\sum_{i=1}^{p}B_{s,i}E\left(\omega_{s,i}(k)\right)\right]^T\\&\quad+\sum_{i=1}^{p}\gamma_{s,i}B_{s,i}\left(E\left(\lambda_s(k)\lambda_s^T(k)\right)\right)B_{s,i}^T+\sum_{\substack{i,j=1\\i\neq j}}^{z}\chi_{s,i,j}B_{s,i}\left(E\left(\lambda_s(k)\lambda_s^T(k)\right)\right)B_{s,j}^T,\end{aligned} \tag{3.18}$$

$$\begin{aligned}G_s(k)&=G_{s,1}(k)+G_{s,2}(k)+G_{s,3}(k),\\G_{s,1}(k)&=\left[A_{s,0}E\left(\overline{\rho}(k)\right)\left(E\left(u(k)\right)\right)^T+\sum_{i=1}^{p}A_{s,i}E\left(\overline{\rho}(k)\right)\left(E\left(\omega_{s,i}(k)u(k)\right)\right)^T\right]B_s^T,\\G_{s,2}(k)&=\left[B_{s,0}+\sum_{i=1}^{p}B_{s,i}E\left(\omega_{s,i}(k)\right)\right]\left(E\left(\lambda_s(k)\right)\right)\left(E\left(u(k)\right)\right)^TB_s^T,\end{aligned} \tag{3.19}$$

$$\begin{aligned}G_{s,3}(k)&=\left[A_{s,0}E\left(\overline{\rho}(k)\right)+\sum_{i=1}^{p}A_{s,i}E\left(\omega_{s,i}(k)\right)E\left(\overline{\rho}(k)\right)\right]\left(E\left(\lambda_s(k)\right)\right)^T\\&\quad\left[B_{s,0}+\sum_{i=1}^{p}B_{s,i}E\left(\omega_{s,i}(k)\right)\right]^T+\sum_{i=1}^{p}\gamma_{s,i}A_{s,i}E\left(\overline{\rho}(k)\right)\left(E\left(\lambda_s(k)\right)\right)^TB_{s,i}^T\\&\quad+\sum_{\substack{i,j=1\\i\neq j}}^{z}\chi_{s,i,j}A_{s,i}E\left(\overline{\rho}(k)\right)\left(E\left(\lambda_s(k)\right)\right)^TB_{s,j}^T,\end{aligned}$$

$$\gamma_{s,i}=E\left(\omega_{s,i}(k)\omega_{s,i}(k)\right)-\left(E\left(\omega_{s,i}(k)\right)\right)^2=\operatorname{var}\left(\omega_{s,i}(k)\right),$$

$$\chi_{s,i,j}=E\left(\omega_{s,i}(k)\omega_{s,j}(k)\right)-E\left(\omega_{s,i}(k)\right)E\left(\omega_{s,j}(k)\right)=\operatorname{cov}\left(\omega_{s,i}(k),\omega_{s,j}(k)\right).$$

As it has been assumed that the supply functions are uncorrelated, so that $\chi_{s,i,j}=0$

whenever $i \neq j$ in (3.18) and (3.19).

To evaluate the mean and variance matrix of the joint traffic density vector $\overline{\rho}(k+1)$, a finite mixture distribution of $\rho_s(k+1)$, s = 1, 2, …, 5, is applied to approximate the overall effect of the five operational modes.

$$\mu(k)=E\left(\overline{\rho}(k+1)\right)=\sum\nolimits_s \Pr_s(k)E\left(\rho_s(k+1)\right). \tag{3.20}$$

Then the auto-correlation matrix $Q(k+1)$ and variance matrix $Var\left(\overline{\rho}(k)\right)$ can be evaluated as:

$$Q(k+1)=\sum\nolimits_s \Pr_s(k)Q_s(k+1)$$

$$Var\left(\overline{\rho}(k)\right)=\sum\nolimits_s \Pr_s(k)Q_s(k+1)-\mu(k)\mu^T(k). \tag{3.21}$$

3.5 Recommondations and summary of the SCTM

3.5.1 Recommendations for empirical studies and numerical simulations

The SCTM can be applied to both empirical studies and numerical simulations for traffic state estimation and prediction. For empirical studies, historical records of traffic conditions such as boundary densities and flow rates are available for simulation. However, for numerical simulations, one would like to test the model under extreme conditions such as sudden increase and decrease of the demand profiles, see e.g. Sumalee et al. (2011), Zhong et al. (2011). We may need to adjust the definitions of the probabilities of occurrence of the operational modes to adapt different environments and the available data. Indeed, as commented by Sumalee et al. (2011) that the definition of probabilities of occurrence of operational modes presented in the paper is only a feasible definition to the author, not a uniform definition of probabilities of occurrence of operational modes. One can revise the definition of probabilities of occurrence of operational modes to adapt to different

traffic conditions and the desired accuracy. For example, in Section 3, the probabilities of occurrence of the operational modes are defined in line with Assumption 3.1, i.e. according to the real measurements. However, for numerical simulations, they are unavailable. To this end, we may revise those probabilities of occurrence of operational modes by replacing the measurements with the estimated boundary densities $\overline{\rho}_1(k)$ and $\overline{\rho}_z(k)$ obtained from the SCTM. Furthermore, the inflow of the upstream $\tilde{q}_u(k)$, and outflow from the downstream $\tilde{q}_d(k)$ will be used to handle the extreme simulation conditions such as sudden increase and decrease of the demand profiles. Take the multiple cells scenario as an example, the probabilities of occurrence of operational modes can be restated as:

•FF mode:

$$P_{FF}(k)=\Pr\left(\begin{array}{l}\left(\left(\overline{\rho}_1(k)<\rho_{c,1}\right)\cap\left(\overline{\rho}_z(k)<\rho_{c,z}\right)\cap\left(v_{f,z}\overline{\rho}_z(k)<\tilde{q}_d(k)\right)\right)\cup\\ \left(\left(\overline{\rho}_1(k)\geq\rho_{c,1}\right)\cap\left(\overline{\rho}_z(k)<\rho_{c,z}\right)\cap\left(v_{f,z}\overline{\rho}_z(k)<\tilde{q}_d(k)\right)\cap\left(\tilde{q}_u(k)<w_{c,1}(\overline{\rho}_1(k)-\rho_{J,1})\right)\right)\end{array}\right)$$

•CC mode:

$$P_{CC}(k)=\Pr\left(\begin{array}{l}\left(\left(\overline{\rho}_1(k)\geq\rho_{c,1}\right)\cap\left(\overline{\rho}_z(k)\geq\rho_{c,z}\right)\cap\left(\tilde{q}_u(k)>w_{c,1}(\overline{\rho}_1(k)-\rho_{J,1})\right)\right)\cup\\ \left(\left(\overline{\rho}_1(k)\geq\rho_{c,1}\right)\cap\left(\overline{\rho}_z(k)<\rho_{c,z}\right)\cap\left(v_{f,z}\overline{\rho}_z(k)>q_d(k)\right)\cap\left(\tilde{q}_u(k)>w_{c,1}(\overline{\rho}_1(k)-\rho_{J,1})\right)\right)\end{array}\right)$$

•CFmode:

$$P_{CF}(k)=\Pr\left(\left(\overline{\rho}_1(k)\geq\rho_{c,1}\right)\cap\left(\overline{\rho}_z(k)<\rho_{c,z}\right)\cap\left(v_{f,z}\overline{\rho}_z(k)<\tilde{q}_d(k)\right)\cap\left(\tilde{q}_u(k)>w_{c,1}(\overline{\rho}_1(k)-\rho_{J,1})\right)\right),$$

• FC mode: $P_{FC}(k)=1-(P_{FF}(k)+P_{CC}(k)+P_{CF}(k)),$

3.5.2 Summary

This chapter reviewed the basic CTM proposed by Daganzo (1994), the modified CTM and its piecewise linearized version SMM., and then discussed the SCTM proposed by Sumalee *et al.* (2011), an extension of the CTM to the stochastic environment, from cell assignments, SCTM segments, evaluations of probabilities of occurrence of operational modes, stochastic bilinear system representations of different operational modes, to the finite mixture distribution approximation of the

overall effect of the five operational modes. This chapter also highlighted the advantages and disadvantages of the model and the technical issues regarding to its applications to different purposes.

As mentioned previously, the uncorrelated assumption may be not reasonable in practice. To be more specific, the uncorrelated or independent assumptions enforced explicitly or implicitly in the SCTM framework cause several limitations to the definitions of probabilities of occurrence of operational modes and their evaluations. These drawbacks render the model may not perform very well when it is applied for short-term traffic flow prediction, especially under abnormal traffic conditions, e.g. incidents and adverse weather conditions due to the fact that the model does not consider the spatial-temporal correlations of the traffic flow fully and lacks of a proper prediction algorithm to forecast demand and supply profiles. This issue will be addressed in Chapter 5.

Chapter 4
Estimation of journey time distribution and reliability analysis

This chapter proposes an algorithm to estimate the stochastic dynamic journey time and indexes to quantify the stochastic dynamic travel time reliability. Section 4.1 clarifies the relationship between several definitions used in this chapter, i.e. cells, segments, links, and routes. Section 4.2 introduces the concept of likelihood and PMF (Probabilistic Mass Function) based on the stochastic cumulative link in-/out- flows for the construction of stochastic dynamic travel time distributions. Section 4.3 develops an algorithm to calculate the PMF of journey time of a route by extending the deterministic nested delay operator to a stochastic version based on a series of conditional probabilities. The distributions of dynamic link and route travel times are obtained by fitting their PMFs to a kind of skew normal distribution. Journey time reliability is analyzed in Section 4.4. Finally, the empirical study in Section 4.5 confirms the effectiveness and efficiency of the proposed algorithms by the consistency between the estimated journey time and the detected one as well as the consistency between the results obtained from the proposed skewness analysis and the statistical records reported in the transportation literature.

4.1 The definition of link, route and SCTM segment

This section clarifies the difference of SCTM segment and freeway link within a certain path. Consider a path on a long freeway corridor as depicted in Figure 4.1. Firstly, the path is naturally divided into links by the on-ramps and the off-ramps along the corridor without any interruption within each link. As illustrated in Figure 4.1(a), the path is partitioned into Nl links. As to segment, which is a concept frequently used in the SCTM (refer to Section 3.1.2), usually consists of several cells and always has detectors installed on its boundaries. The same freeway corridor represented by Ns segments is shown in Figure 4.1(b).

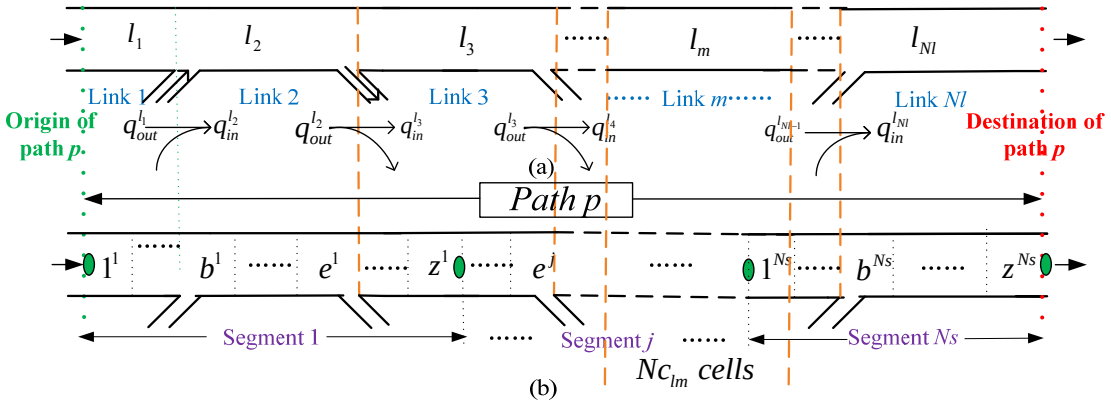


Figure 4.1. The relationship of the SCTM segment and the link used in travel time

With these concepts, we introduce the link travel time evaluation under a deterministic continuous-time macroscopic dynamic traffic flow model. The flow propagation equation for link l_m can be expressed as follows

$$C_{in}^{l_m}(t) = C_{out}^{l_m}(\tau_{l_m}(t)), \tag{4.1}$$

where $\tau_{l_m}(t)$ is the exit time from link l_m for a vehicle entering this link at time t. $C_{in}^{l_m}(t)$ and $C_{out}^{l_m}(\tau_{l_m}(t))$ are the cumulative inflow and outflow volumes at the entry and exit times of this link, respectively. The time-flow consistency equation in (4.1) is applicable on condition that the FIFO principle holds and there was no vehicles on the within link l_m at initial time. Under FIFO, the deterministic link travel time $\beta_{l_m}(t)$ can be defined as

$$\beta_{l_m}(t) = \tau_{l_m}(t) - t. \tag{4.2}$$

The journey time of route p is then defined by the nested delay operator with exit time $\tau_{l_m}(t), (\forall l_m \in p)$ of all the links on the path as:

$$\eta_p(t) = \tau_{l_{Nl}}(\cdots \tau_{l_m}(\cdots \tau_{l_1}(t))) - t. \tag{4.3}$$

where the exit time from link m is the entry time to link $m+1$.

Under the deterministic environment, Lo and Szeto (2002) extended this flow matching algorithm to approximate both link and route travel times by simultaneously tracing the link and route flows by the CTM and adopting the mean-value theorem in Calculus. This method is somehow tedious and difficult to be extended to handle the stochastic environment. In this research, we adopt the SCTM to model the stochastic traffic flow (or the SCTM is utilized as the network loading model). Rather than tracing the link and route flows simultaneously, we

first obtain the PMF of link travel times and then extend the nested delay operator to stochastic case to calculate the route travel time.

4.2 The PMF algorithm for link travel time estimation

4.2.1 Evaluation of cumulative flows from the SCTM

Because the SCTM propagates stochastic traffic density in discretized space and time domains, the above continuous cumulative flow matching algorithm for travel time calculation need to be transferred to a discrete version. Under the SCTM framework, the distributions of inflow rate $q_{in}^{l_m}(k)$ and outflow rate $q_{out}^{l_m}(k)$ of link l_m can be obtained by evaluating the probabilities of occurrence of operational modes and traffic states of the freeway segments corresponding to the entry and exit of the link. For example, in Figure 4.1 the entry of link l_3 is located at the e^1+1 cell within segment 1 while the exit of link 3 is located at cell e^j within segment j. The mean value of in-/out- flow rates (veh/hr) $q_{in}^{l_3}(k)$ and $q_{out}^{l_3}(k)$ can be evaluated by the following equations:

$$\begin{aligned} E\left(q_{in}^{l_3}(k)\right) &= \sum\nolimits_{s=FX,e^1} \Pr_s\left(k\right)\left(v_{f,e^1}\overline{\rho}_{e^1}(k) - q_{off,e^1}(k)\right) \\ &\quad + \sum\nolimits_{s=CX,e^1} \Pr_s\left(k\right)\left(w_{c,e^1+1}\left(\rho_{J,e^1+1} - \overline{\rho}_{e^1+1}(k)\right)\right), \\ E\left(q_{out}^{l_3}(k)\right) &= \sum\nolimits_{s=FX,e^j} \Pr_s\left(k\right)\left(v_{f,e^j}\overline{\rho}_{e^j}(k)\right) \\ &\quad + \sum\nolimits_{s=CX,e^j} \Pr_s\left(k\right)\left(w_{c,e^j+1}\left(\rho_{J,e^j+1} - \overline{\rho}_{,e^j+1}(k)\right) + q_{off,e^j}(k)\right) \end{aligned} \tag{4.4}$$

where $\Pr_{FX,e^j}\left(k\right)$ denotes the traffic modes that all the cells at the upstream of cell e^j within segment j is free-flow status, e.g. $FF, FC_{1,e^j}\left(FC_{1,e^j+1}, FC_{1,e^j+2}\cdots\right)$, $FC_{2,e^j-1}\left(FC_{2,e^j}, FC_{2,e^j+1}\cdots\right)$; while $\Pr_{CX,e^j}\left(k\right)$ denotes $CC, CF_{e^j}\left(CF_{e^j+1}, CF_{e^j+2}\cdots\right)$. $q_{in}^{l_m}(k)$ and $q_{out}^{l_m}(k)$ of other links can be similarly calculated. Equation (4.4) applies the same finite mixture algorithm as Equation (3.20). The standard deviation $\sigma_{q_{in}^{l_m}}\left(x\right)$ and $\sigma_{q_{out}^{l_m}}\left(x\right)$ can be evaluated via the same method as Equation (3.21).

Assume $q_{in}^{l_m}(k-1)$ and $q_{out}^{l_m}(k-1)$ are conducted during interval $[k-1,k)T_s$. Then mean value of cumulative flow volumes (vehicles) are:

$$\bar{C}_{in}^{l_m}(k) = \sum_{x=1}^{x=k-1} E\left(q_{in}^{l_m}(x)\right)T_s,$$
$$\bar{C}_{out}^{l_m}(k) = \sum_{x=1}^{x=k-1} E\left(q_{out}^{l_m}(x)\right)T_s. \tag{4.5}$$

The corresponding standard deviations are:

$$\sigma_{C_{in}^{l_m}}\left(k\right) = \sqrt{\sum_{x=1}^{x=k-1} \sigma_{q_{in}^{l_m}}^2\left(x\right)}T_s,$$
$$\sigma_{C_{out}^{l_m}}\left(k\right) = \sqrt{\sum_{x=1}^{x=k-1} \sigma_{q_{out}^{l_m}}^2\left(x\right)}T_s. \tag{4.6}$$

The cumulative inflow and outflow distributions can easily be represented as:

$$C_{in}^{l_m}(k) \cong f_{Cin}\left(\bar{C}_{in}^{l_m}(k), \sigma_{C_{in}^{l_m}}(k)\right),$$
$$C_{out}^{l_m}(k) \cong f_{Cout}\left(\bar{C}_{out}^{l_m}(k), \sigma_{C_{out}^{l_m}}(k)\right), \tag{4.7}$$

where $f_{Cin}(\cdot)$ and denotes a certain statistical distribution, $\bar{C}_{in}^{l_m}(k)$ denotes the mean value, and $\sigma_{C_{in}^{l_m}}(k)$ denotes the standard deviation of the cumulative link inflow. The cumulative outflow distribution can be similarly defined. In this research $f_{Cin}(\cdot)$ is assumed to be normally distributed.

4.2.2 The concept of likelihood and PMF

Since the cumulative inflow and outflow are both random processes, the FIFO principle based link travel time estimation method is not directly applicable[1]. Nevertheless, the FIFO

[1]If a macroscopic traffic flow model is adopted to simulate the traffic dynamics, the FIFO principle needs to be fulfilled by the underlying traffic flow model. In this chapter, SCTM is utilized as the network loading model. Under the deterministic CTM, the FIFO principle will be fulfilled if the cell length l_i is chosen such that $v_{f,i}T_s \le l_i$ where $v_{f,i}$ is the free flow speed of cell i, and T_s is the simulation time increment. This condition cannot always be satisfied in the SCTM framework since mathematically the free-flow speed $v_{f,i}$ can be anything along its distribution. Nevertheless, certain concept that is similar to the "almost sure (or with probability one)" in stochastic analysis can be adopted to redefine this condition. The probabilistic

principle is extended here to estimate the probability of travel time to be a certain value. For a given entry time, this extension estimates the probability of each of the future time interval to be the exit time of this entry time according to the likelihood that the difference between the cumulative inflow and the cumulative outflow at a future time step is less than a prescribed bound.

The probability of the time step k' to be the exit time for vehicles entering the link at time step k (i.e. the entry time $ET = k$) can be written as: $\Pr\left(C_{in}^{l_m}(k) = C_{out}^{l_m}(k') \middle| ET = k\right)$. However, this probability is not well defined. We, thus, propose the likelihood of time step k' to be the exit time (or $k' - k$ to be travel time) for vehicles entering the link at time k as the probability of the distribution of difference $\left(C_{out}^{l_m}(k') - C_{in}^{l_m}(k)\right)$ between the cumulative inflow $C_{in}^{l_m}(k)$ at a certain time step k and the cumulative outflow $C_{out}^{l_m}(k')$ at time step k' is less than a prescribed bound ε:

$$P'_{k'|k} = Pr\left(-\varepsilon \le C_{out}^{l_m}(k') - C_{in}^{l_m}(k) \le \varepsilon \mid ET = k\right) \tag{4.8}$$

where $\varepsilon \in R^+$ denotes a pre-defined positive small number. The definition of ε might influence the accuracy of result which will be discussed in Section 4.5.4.

Given an entry time k, all future time steps $[k+1, +\infty)$ could be the admissible exit time steps. However, this is not practical and unnecessary, e.g. it would be a waste of time and computational effort to trace all future time intervals. In this research, similar to Miller-Hooks and Mahmassani (1998, 2000), we evaluate the probability in Equation (4.8) against a certain subset of the possible exit time, which is defined by the lower-bound k_{lb} and upper-bound k_{ub} of the following sampling interval $[k_{lb}, k_{ub}]$ as:

---

version of the FIFO condition in the SCTM framework is roughly defined as $\Pr\left(v_{f,i}T_s \le l_i\right) \ge \chi$, where χ is a positive real number which satisfies $1-\varepsilon < \chi < 1$ for a small real number ε. We may choose χ very close to 1, such that $\Pr\left(v_{f,i}T_s \le l_i\right) \approx 1$, i.e. the event $v_{f,i}T_s \le l_i$ is almost sure. In the empirical studies, the cell lengths are defined according to this condition. In a word, the FIFO is well satisfied in the sense of "almost sure" in the SCTM formulation.

$$k_{lb} = \min\left\{k' : \left\|\left(\bar{C}^{l_m}_{in}(k) - i\sigma_{C^{l_m}_{in}}(k)\right) - \left(\bar{C}^{l_m}_{out}(k') + j\sigma_{C^{l_m}_{out}}(k')\right)\right\| \le \varepsilon_s\right\},$$
$$k_{ub} = \min\left\{k' : \left\|\left(\bar{C}^{l_m}_{in}(k) + i\sigma_{C^{l_m}_{in}}(k)\right) - \left(\bar{C}^{l_m}_{out}(k') - j\sigma_{C^{l_m}_{out}}(k')\right)\right\| \le \varepsilon_s\right\}, \quad (4.9)$$

where ε_s is a small positive number which is a prescribed error bound, and *i, j* are always selected as positive integers (non-integers are also acceptable) which can be adjusted as long as there is no overlapping between the two curves $\bar{C}^{l_m}_{in}(k) - i\sigma_{C^{l_m}_{in}}(k)$ and $\bar{C}^{l_m}_{out}(k') + j\sigma_{C^{l_m}_{out}}(k)$ (which in turn implies $k' > k$). The selection of *i* and *j* will also be discussed in Section 4.5.4. An illustration of this sampling region concept is given by Figure 4.2.

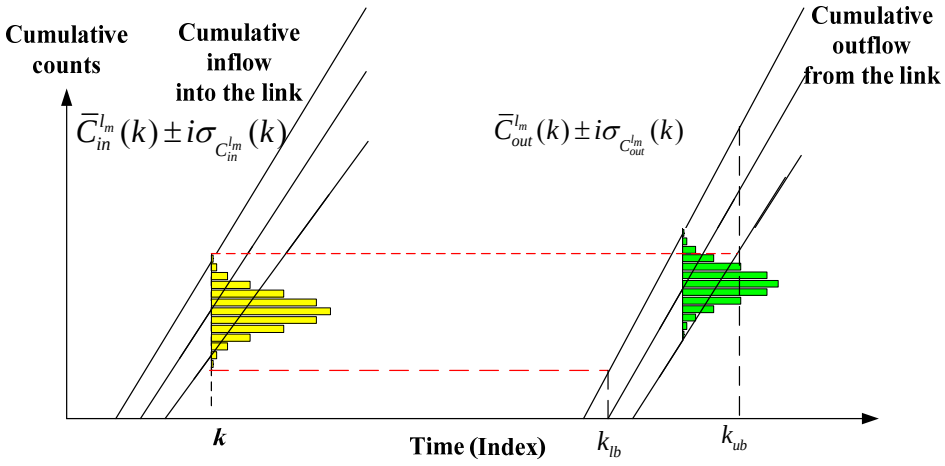


Figure 4.2. An illustration of the *sampling* region

Following the sampling technique previously described, we can obtain a series of $P'_{k'|k}$ which describes the likelihood of k' is the exit time index (link travel time to be equal to k'-k) for entry time k. Note that the summation of the probabilities $\sum_{k_{lb}}^{k_{ub}} P'_{k'|k}$ may not be equal to 1, so the relative frequency is introduced to normalize the probabilities. For a vehicle entering link l_m at time k, the relative frequency $P_{k'|k}$ is defined as:

$$P_{k'|k} = \frac{P'_{k'|k}}{\sum_{klb}^{kub} P'_{k'|k}}, \forall k' \in [k_{lb}, k_{ub}]. \quad (4.10)$$

From the normalized probabilities, we can construct the Probabilistic Mass Function (PMF) of the link travel time for traffic entering the link at time k.

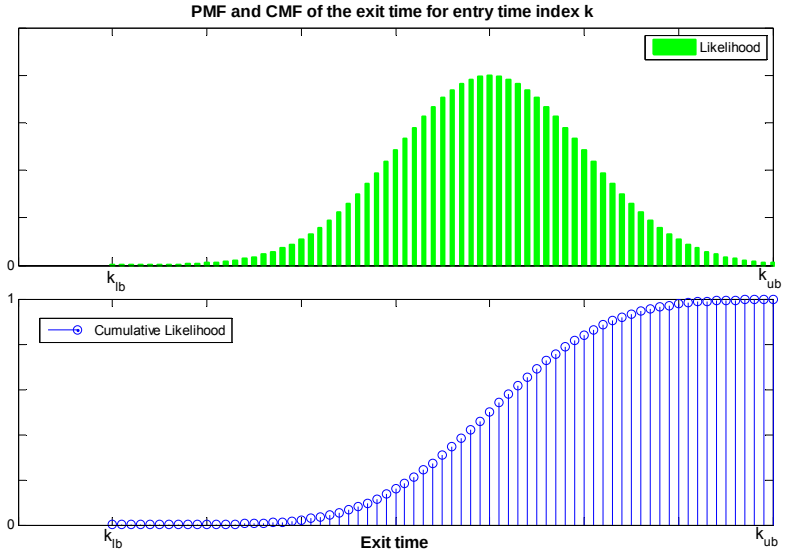


Figure 4.3 The PMF and the corresponding CMF respect to exit time index for entry time index *k*

An example of the PMF of the link travel time is shown in Figure 4.3. Each bar on the upper plot represents $P_{k'|k}$. The solid line with dots on the lower plot denotes the corresponding cumulative mass function (CMF).

4.2.3 Interpretation of the likelihood and simplification of cumulative flow calculations

To interpret and simplify the likelihood, $e_k(k') = C_{out}^{l_m}(k') - C_{in}^{l_m}(k)$ is defined as the matching error which measures the difference between the cumulative inflow and outflow distributions. Equation (4.8) is then equivalent to

$$P'_{k'|k} = \Pr\left(-\varepsilon \le e_k(k') \le \varepsilon \mid ET = k\right).$$

Simplify the calculation of $e_k(k')$ as:

$$\begin{aligned} e_k(k') &= C_{out}^{l_m}(k') - C_{in}^{l_m}(k) \\ &= \sum_{x=1}^{k'-1} q_{out}^{l_m}(x) T_s - \sum_{x=1}^{k-1} q_{in}^{l_m}(x) T_s \\ &= \sum_{x=k}^{k'-1} q_{out}^{l_m}(x) T_s - \left(-\sum_{x=1}^{k-1} q_{out}^{l_m}(x) + \sum_{x=1}^{k-1} q_{in}^{l_m}(x) \right) T_s \\ &= \sum_{x=k}^{k'-1} q_{out}^{l_m}(x) T_s - \sum_{i=1}^{NC_{l_m}} \rho_i^{l_m}(k) s_i^{l_m} \end{aligned} \tag{4.11}$$

where NC_{l_m} is the number of cells on link l_m, $\rho_i^{l_m}(k)$ is the density of the i^{th} cell on link

l_m at time *k*, and $s_i^{l_m}$ is the length of the cell i. Note that $s_i^{l_m}$ and T_s are constants, meanwhile the link outflow profile and cell densities are normally distributed by assumption, and the matching error is normally distributed[2]. Thus, it is sufficient for us to obtain the mean and variance of $e_k(k')$ so as to evaluate the probability defined by Equation (4.8). The expectation of Equation $e_k(k')$ is

$$E\left(e_k(k')\right) = \sum_{x=k+1}^{k'} T_s \cdot E\left(q_{out}^{l_m}(x)\right) - \sum_{i=1}^{NC_{l_m}} s_i^{l_m} \cdot E\left(\rho_i^{l_m}(k)\right).$$

If the above random variables are uncorrelated, we can obtained the variance of $e_k(k')$ easily by the additive of independent normal variables. However, the variance is not additive in general due to the correlation. Nevertheless, we can make use of the covariance matrix[3] of the corresponding vector (augmented by the above random variables) obtained from the SCTM to proceed the variance evaluation which is discussed in details in the Appendix of the chapter.

4.3 The extended nested delay operator and distribution fitting

Consider a path which starts at link l_1 and ends at l_{Nl} as depicted in Figure 4.4(a). Tracing the path travel time distribution for vehicles entering the origin at time step *k*, we extend the nested delay operator to stochastic case. For vehicles entering the origin at time step *k*, the relative frequency of time-step τ^{Nl} to be the exit time from the destination of the path for is evaluated as the following progress:

$$P_{1,nest}(\tau^1 \mid k) = P_1(\tau^1 \mid k), \quad \tau^1 \in \left[k_{lb}^{l_1}, k_{ub}^{l_1}\right]$$

$$P_{2,nest}(\tau^2 \mid k) = \sum_{\tau^1 = k_{lb}^{l_1}}^{k_{ub}^{l_1}} P_2(\tau^2 \mid \tau^1) P_{1,nest}(\tau^1 \mid k), \quad \tau^2 \in \left[\left(k_{lb}^{l_1}\right)_{lb}^{l_2}, \left(k_{ub}^{l_1}\right)_{ub}^{l_2}\right]$$

…

[2] Please see the appendix for details.

[3] If *X* is a random vector with n components, the matrix $Var(X) = E\left((X - E(X))(X - E(X))^T\right)$ is the (theoretical) covariance matrix.

$$P_{m,nest}(\tau^m \mid k) = \sum_{\tau^{m-1}=\left\lfloor\left\lfloor k^{l_1}_{lb}\right\rfloor\cdots\right\rfloor^{l_{m-1}}_{lb}}^{\left\lfloor\left\lfloor k^{l_1}_{ub}\right\rfloor\cdots\right\rfloor^{l_{m-1}}_{ub}} P_m(\tau^m \mid \tau^{m-1}) P_{m-1,nest}(\tau^{m-1} \mid k), \quad \tau^m \in \left[\left\lfloor\left\lfloor k^{l_1}_{lb}\right\rfloor\cdots\right\rfloor^{l_m}_{lb}, \left\lfloor\left\lfloor k^{l_1}_{ub}\right\rfloor\cdots\right\rfloor^{l_m}_{ub}\right]$$

$$\cdots$$

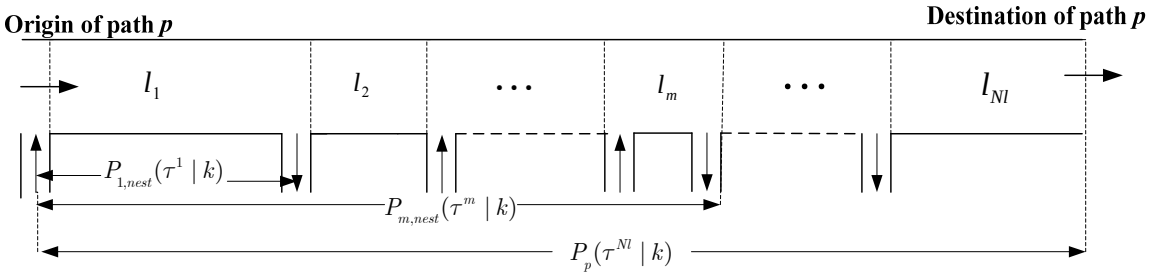
$$P_p(\tau^{Nl} \mid k) = \sum_{\tau^{Nl-1}=\left\lfloor\left\lfloor k^{l_1}_{ub}\right\rfloor\cdots\right\rfloor^{l_{Nl-1}}_{lb}}^{\left\lfloor\left\lfloor k^{l_1}_{ub}\right\rfloor\cdots\right\rfloor^{l_{Nl-1}}_{ub}} P_{Nl}(\tau^{Nl} \mid \tau^{Nl-1}) P_{Nl-1,nest}(k^{Nl-1} \mid k), \tag{4.12}$$

where $k^{l_1}_{lb}, k^{l_1}_{ub}$ can be defined according to (4.9), while the superscript l_m is used to denote the link number. Other bounds can be defined recursively as:

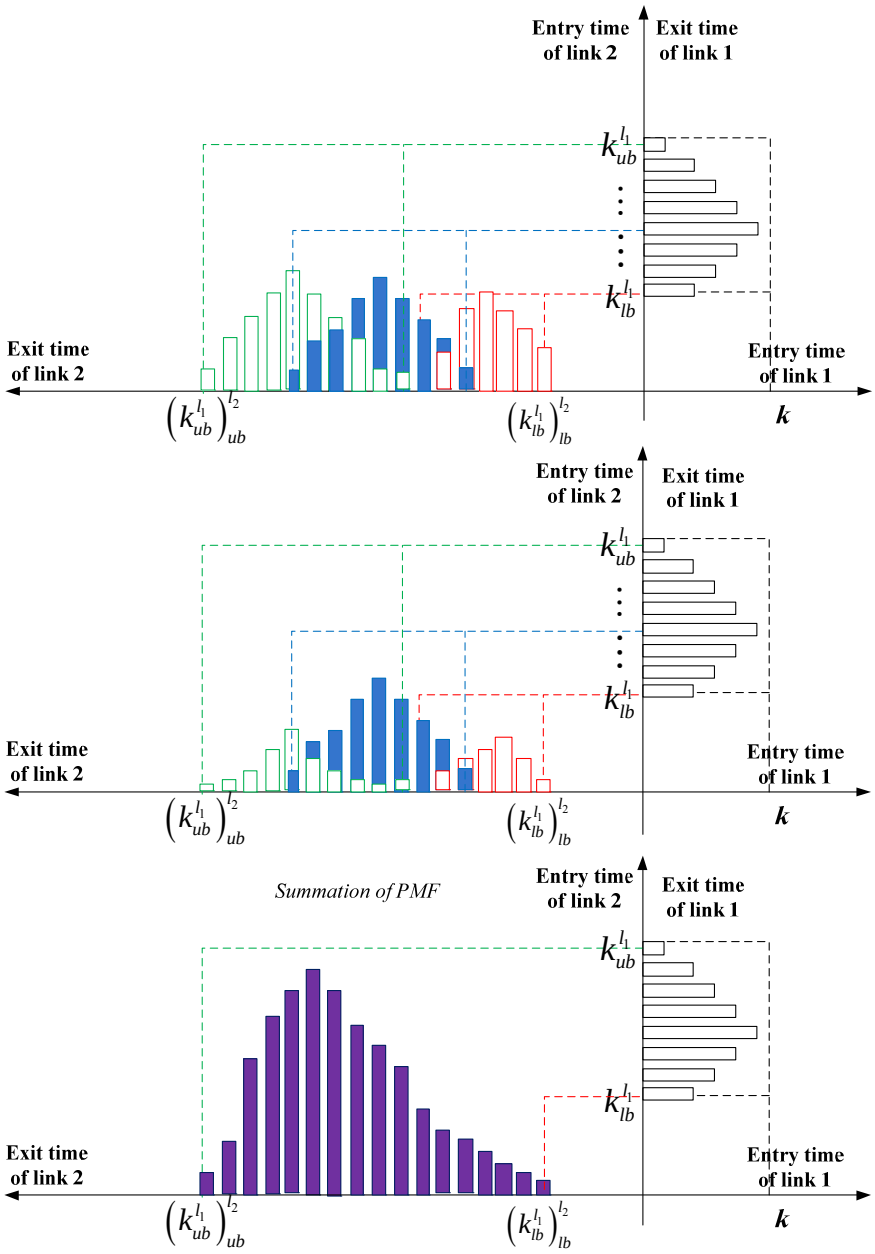
$$\begin{aligned}
&\left(k^{l_1}_{lb}\right)^{l_2}_{lb} = \min\left\{k' : \left\|\left(\overline{C}^{l_2}_{in}(k^{l_1}_{lb}) - i\sigma_{C^{l_2}_{in}}(k^{l_1}_{lb})\right) - \left(\overline{C}^{l_2}_{out}(k') + j\sigma_{C^{l_2}_{out}}(k')\right)\right\| \le \varepsilon_s\right\},\\
&\left(k^{l_1}_{ub}\right)^{l_2}_{ub} = \min\left\{k' : \left\|\left(\overline{C}^{l_2}_{in}(k^{l_1}_{ub}) + i\sigma_{C^{l_2}_{in}}(k^{l_1}_{ub})\right) - \left(\overline{C}^{l_2}_{out}(k') - j\sigma_{C^{l_2}_{out}}(k')\right)\right\| \le \varepsilon_s\right\},\\
&\vdots\\
&\left(\left(k^{l_1}_{lb}\right)\cdots\right)^{l_m}_{lb} = \min\left\{k' : \left\|\left(\overline{C}^{l_m}_{in}\left(\left(\left(k^{l_1}_{lb}\right)\cdots\right)^{l_{m-1}}_{lb}\right) - i\sigma_{C^{l_m}_{in}}\left(\left(\left(k^{l_1}_{lb}\right)\cdots\right)^{l_{m-1}}_{lb}\right)\right) - \left(\overline{C}^{l_m}_{out}(k') + j\sigma_{C^{l_m}_{out}}(k')\right)\right\| \le \varepsilon_s\right\},\\
&\left(\left(k^{l_1}_{ub}\right)\cdots\right)^{l_m}_{ub} = \min\left\{k' : \left\|\left(\overline{C}^{l_m}_{in}\left(\left(\left(k^{l_1}_{ub}\right)\cdots\right)^{l_{m-1}}_{ub}\right) + i\sigma_{C^{l_m}_{in}}\left(\left(\left(k^{l_1}_{ub}\right)\cdots\right)^{l_{m-1}}_{ub}\right)\right) - \left(\overline{C}^{l_m}_{out}(k') - j\sigma_{C^{l_m}_{out}}(k')\right)\right\| \le \varepsilon_s\right\}.
\end{aligned} \tag{4.13}$$

Figure 4.4(b) demonstrates the evaluation of stochastic nested delay operator over a path with 2 links. Each time step within the sampling region $\left[k^{l_1}_{lb}, k^{l_1}_{ub}\right]$ both correspond to a probability of exit time for link 1 and a PMF over link 2. Propose a weighed summation to each exit time of link 2, the PMF of path travel time can be evaluated. Meanwhile, the sampling region is extended to $\left[\left(k^{l_1}_{lb}\right)^{l_2}_{lb}, \left(k^{l_1}_{ub}\right)^{l_2}_{ub}\right]$.

van Lint et al. (2008) found that the stochastic journey time may not follow the normal distribution in which different patterns of skewness can be observed under different traffic conditions. The travel time under the free-flow condition usually follows a normal distribution. On the other hand, under the congested condition or the congestion onset/dissolve process, the link travel time may follow a skewed distribution (van Lint et al., 2008; Kharoufeh and Gautam, 2004). Thus, fitting the PMF to a distribution with skewness should better describe the uncertainty of journey time (or link time). The shape of the journey time distribution will also affect the evaluation of the travel time reliability.



(a) Schematic diagram for path travel time



(b) Stochastic nested delay operator for a path with *2* links

Figure 4.4. Nested delay operator

After obtaining the PMF of the path journey time, various random distribution fitting techniques can be applied to obtain the distribution that best fits the estimated PMF. $T_{rs}(k) = g\left(\xi(k), \omega(k), \alpha(k)\right)$ is the distribution of the dynamic journey time at time *k*, if $g\left(\overline{\beta}(k), \sigma_{\beta(k)}, \alpha(k)\right)$ best fits the PMF and/or the corresponding CMF with mean $\xi + \omega\delta\sqrt{\frac{2}{\pi}}$, standard deviation $\omega\sqrt{1-\frac{2\delta^2}{\pi}}$ and shape parameter $\alpha(k)$ which reflects the skewness of the PMF, where $\delta = \frac{\alpha}{\sqrt{1+\alpha^2}}$. In this chapter, the skew normal distribution is used to describe the characteristic of the travel time distribution with the PDF (Azzalini and Capitanio, 1999):

$$g(x) = 2\frac{1}{\sqrt{2\pi\omega}}e^{-\frac{(x-\xi)^2}{2\omega^2}}\int_{-\infty}^{\alpha\left(\frac{x-\xi}{\omega}\right)}\frac{1}{\sqrt{2\pi\omega}}e^{-\frac{t^2}{2}}dt \tag{4.14}$$

where the distribution is called positive skew if $\alpha > 0$ with a longer right tail and concentrating on the left side, while negative skew if $\alpha < 0$ with a longer left tail and concentrating on the right side (see e.g. Figure 4.5).

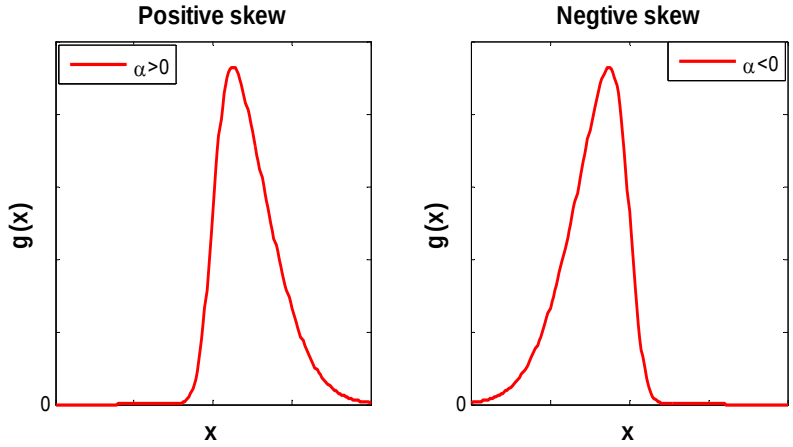


Figure 4.5. The skew normal distribution

4.4 Index of travel time reliability and conclusion of PMF based journey time estimation algorithm

Several indicators of travel time reliability were proposed in literature. For illustrative purpose we adopt the buffer time index as the measure of travel time reliability for the case studies in

this chapter. The buffer time index (BTI) is one of the most widely used travel time reliability indicators, in which the buffer time represents the extra time that travelers must add to their average travel time when planning trips to ensure his on-time arrival (Lomax *et al.,* 2003). This concept is similar to the safety margin measure proposed in Hall (1983) which was recently adopted in Lam *et al.* (2008). In this chapter, the probability of arriving on-time is set at the 95th percentile which can be translated as "can be late for work one day per month without getting into too much trouble" or 95% of travel time observations can be found under this criterion. Buffer time index can then be calculated by Equation (4.15):

$$Buffer\ Time\ Index = \left[\frac{95th\ Pencentile\ Travel\ Time - Average\ Travel\ Time}{Average\ Travel\ Time} \times 100\% \right] \quad (4.15)$$

From Equation (4.15), it is obvious that adequate information about the journey time distribution, including mean, standard deviation and the shape parameter, are required to evaluate the BTI properly. The skew normal distribution satisfies this requirement.

4.5 Empirical study 1: Journey time estimation on a segment of Hanshin Expressway

This section presents an empirical study to illustrate and validate the proposed algorithm for estimating the dynamic stochastic journey time and calculating the buffer time index. This empirical study is tested on a long corridor within the Hanshin expressway in Osaka, Japan, and focuses on estimating the influence of accidents to the journey time reliability.

4.5.1 Description of the test site and data preparation

The empirical study is conducted on a 12 km segment on No.11 Hanshin Expressway Ikeda corridor from Toyonaka city to the CBD of Osaka as illustrated in Figure 4.6. This expressway, which includes 10 sections, is composed of 9 links determined by the on-ramps and off-ramps as shown in Figure 4.7.

Table 4.1 presents the callibration results of the parameters in fundamental diagrams based on the historical data under normal scenario (i.e. no incident), which is provided by detectors

equipped within each section. The historical data, which will be used in simulation, includes flow and speed, which is provided by all the 11 detectors from 0:00 AM of May 11 (Sunday), 2008 to 23:55 PM of May 17 (Saturday), 2008 with a detecting interval as 5 minutes. Table 4.2 depicts the assignment of cells along the whole expressway corridor, where the 38 cells mentioned in the table constitute the 10 SCTM segments. The simulation step is chosen as 10 seconds. The objective of this empirical study is to simulate the effects of traffic incidents to the dynamic journey time and traffic states, e.g. how would different incidential locations and time of occurrence and duration generate different kinds of traffic jams. Table 4.3 depeicts the a sample of incident record which indicates 3 traffic accidents and 3 traffic jams happened on the main road during 7:00 AM-21:00 PM on May 12, 2008: Jam 1898 during 7:16 AM-7:28 AM starts from Link 5 and Link 6 is not caused by accident. Jam 1909 and Jam 1910 during 8:38 AM-10:35 AM on Link 10 and its upstream segments are induced by Incident 1599 and Incident 1419 during 8:30 AM-10:35 AM on Link10, while Incident 1417 does not lead to any congestion. Note that the information on the jams are not complete, e.g. the spillback effects of congestion are not recorded in the database.

Figure 4.6. The map of the test segment from Toyonaka to Osaka

Link ID (Length/km)	Section ID	$\hat{v}_f$	$\sigma_{\hat{v}_f}$	$\hat{w}_c$	$\sigma_{\hat{w}_c}$	$\hat{\rho}_c$	$\hat{\rho}_J$	$\sigma_{\hat{\rho}_J}$	$\hat{Q}_m$	$\sigma_{\hat{Q}_m}$
Link 1 (2.2)	1	82	6.2	13	5.0	39	279	98	3300	248
Link 2(1.0)	2	106	16.2	14.5	4.9	28.1	235	78.5	3000	456
Link 3(0.3)	3	78	9.2	13.6	3.7	44.3	302	76.3	3500	407
Link 4(3.1)	4	73	6.5	23.8	6.6	54.2	222	50.6	4000	350
Link 5-6 (1.5)	5	78	18.6	22.3	9.9	51	231	100	4000	953
	6	83	7.4	18.8	5.2	48	261	63.3	4000	356
Link 7(1.8)	7	71	5.2	19.7	6.2	55.7	259	70.6	4000	288
Link 8(0.9)	8	72	5.1	23.8	4.3	56.6	225	31.0	4000	289
Link 9(0.3)	9	73	6.5	16.2	5.2	52.0	286	84.0	3800	340
						$\hat{\rho}_a$ $\hat{\rho}_b$				
Link 1(1.0)	10	69	6.3	13.1	2.7	58 150	450	64.0	4000	367

Table 4.1 The parameters of fundamental diagrams along the route under normal case

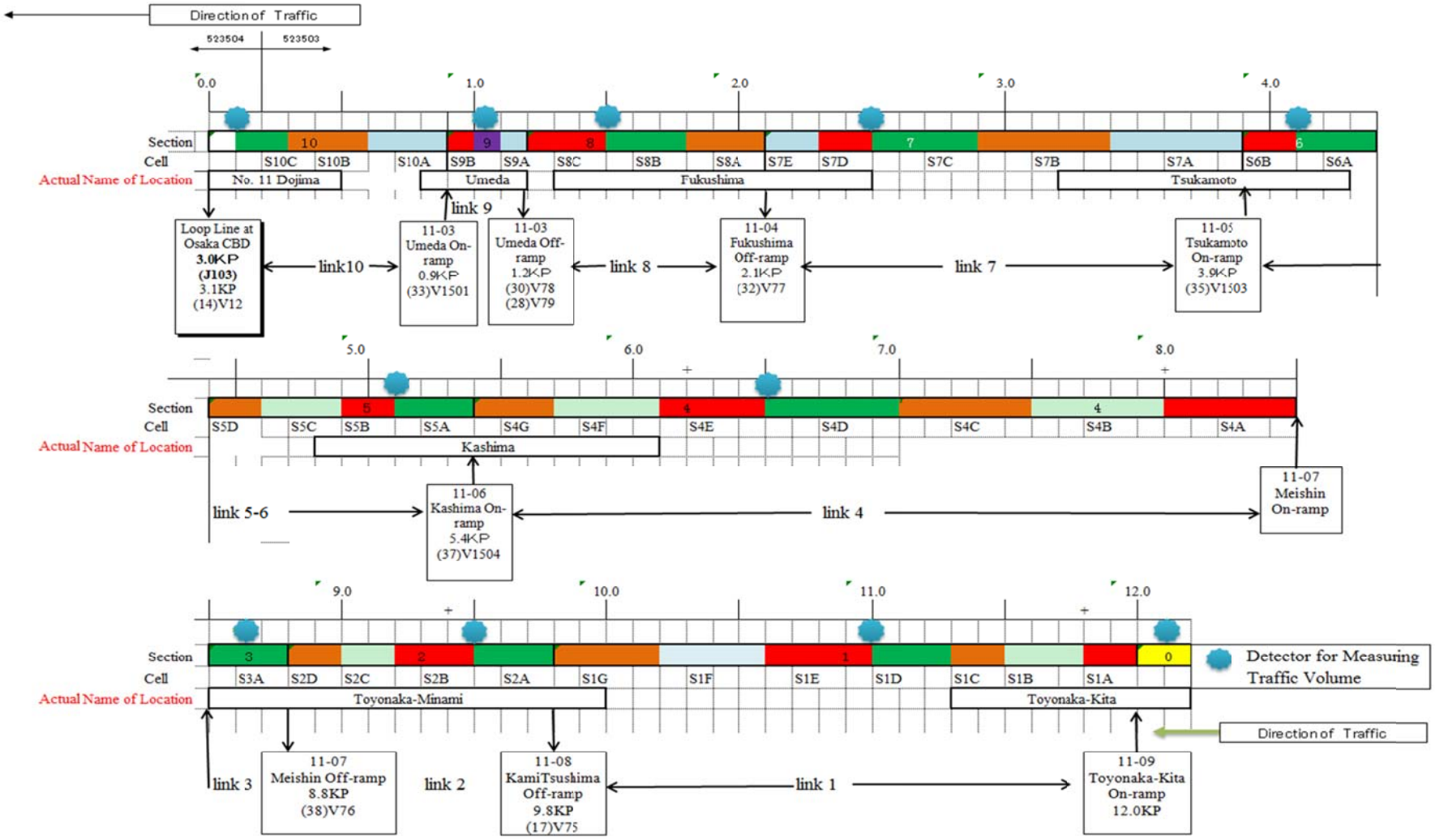


Figure 4.7 The assignment of cells and the locations of the detectors

Link ID and length (km)	Section ID	Cell ID and Length(km)	SCTM segment ID
Link 1(2.2) Toyonaka-Kita On-ramp -Kami Tsushima Off-ramp	Section 1	*Cell S1A(0.2) Cell S1B(0.3) Cell1S1 C(0.2) Cell S1D(0.3)*	Segment S1A-S1D
		*Cell S1E(0.4) Cell S1 F(0.4) Cell S1 G(0.4)	Segment S1E-S2A
Link 2(1.0) Kami Tsushima Off-ramp -Meishin Off-ramp	Section 2	Cell S2A(0.3)*	
		*Cell S2B(0.3) Cell S2C(0.2) Cell S2D(0.2)	Segment S2B-S3A
Link 3(0.3) Meishin Off-ram -Meishin On-ramp	Section 3	Cell S3A (0.3))*	
Link 4(3.1) Meishin On-ramp -Kashima On-ramp	Section 4	*Cell S4A(0.5) Cell S4B(0.5) Cell S4C(0.5) Cell S4 D(0.5)*	Segment S4A-S4D
		*Cell S4 E(0.4) Cell S4F(0.4) Cell S4G(0.3)	Segment S4E-S5A
Link 5-6 (1.5) Kashima On-ramp -Tsukamoto On-ramp	Section 5	Cell S5A(0.3)*	
		*Cell S5B(0.2) Cell S5C(0.3) Cell S5D(0.2)	Segment S5B-S6A
	Section 6	Cell S6A(0.3)*	
		*Cell S6B(0.2)	Segment S6B-S7C
Link 7(1.8) Tsukamoto On-ramp -Fukushima Off-ramp	Section 7	Cell S7A(0.5) Cell S7B(0.5) Cell S7C(0.4)*	
		*Cell S7D(0.2) Cell S7E(0.2)	Segment S7D-S8B
Link 8(0.9) Fukushima Off-ramp -Umeda Off-ramp	Section 8	Cell S8A(0.3) Cell S8B(0.3)*	
		*Cell S8C(0.2)	Segment S8C-S9A
Link 9(0.3) Umeda Off-ramp -Umeda On-ramp	Section 9	Cell S9A(0.2)*	
		*Cell S9B(0.2)	Segment S9B-S10C
Link 10(1.0) Umeda On-ramp -Osaka CBD	Section 10	Cell S10A(0.3) Cell S10B(0.3) Cell S10C(0.2)*	

*The location of detector

Table 4.2. The assignment of cells and segments for SCTM

Incident ID	Incident Occurrence Date	Incident Occurrence Time	Incident Ending Date	Incident Ending Time	Cause	Severity	Location	Cells involved
1898	5/12/2008	7:16:00 AM	5/12/2008	7:28:00 AM	Jam		4.0 - 5.0	
1599	5/12/2008	8:30:00 AM	5/12/2008	10:35:00 AM	Traffic Accident	One lane closure	0.7-0.9	Cell10 A
1909	5/12/2008	8:38:00 AM	5/12/2008	10:35:00 AM	Jam		0.7 - 0.9	
1910	5/12/2008	8:43:00 AM	5/12/2008	10:36:00 AM	Jam		3.0 - 3.0	
1415	5/12/2008	10:02:00 AM	5/12/2008	10:14:00 AM	Obstacle dropped from veh.	One lane closure	0.2	Cell10 C
1417	5/12/2008	10:41:00 AM	5/12/2008	10:50:00 AM	Mulfunction of Veh.	One lane closure	1.0 - 3.0	Cell7 B-Cell9B

Table 4.3 Incident record of the main road between 7:00 and 21:00 on Monday

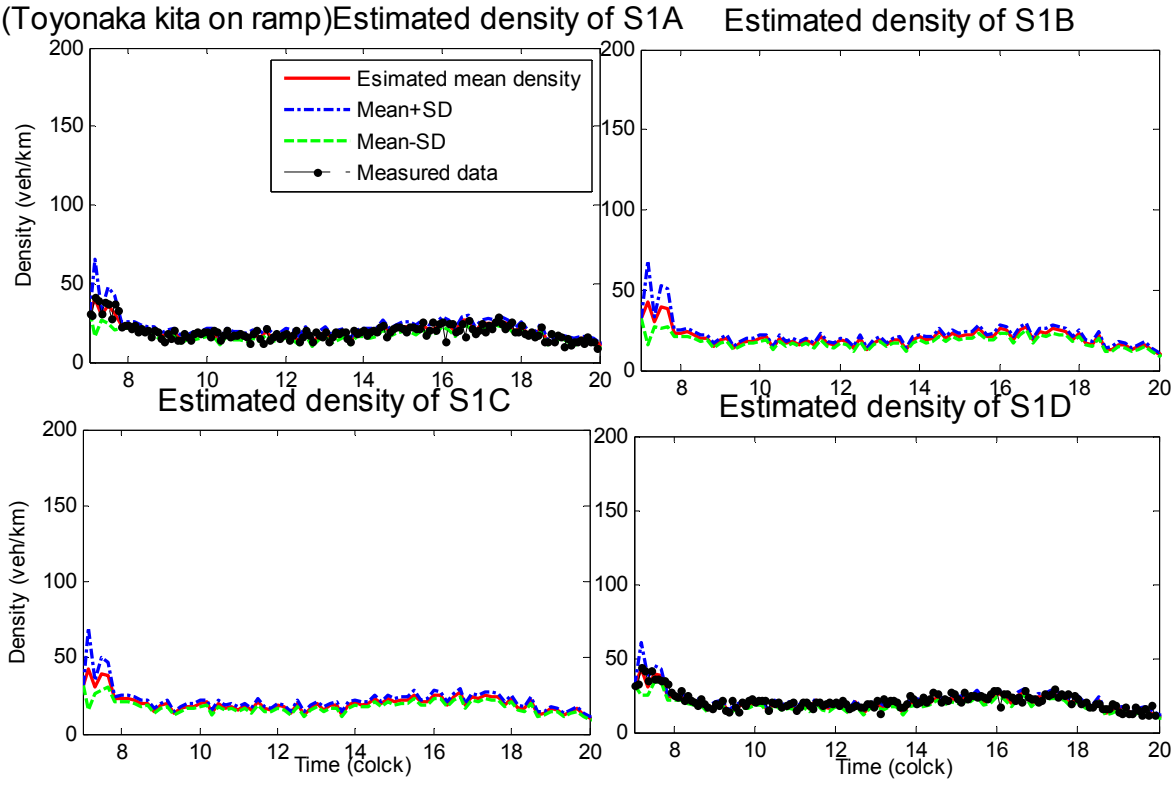
4.5.2 Traffic density estimation under incident scenarios

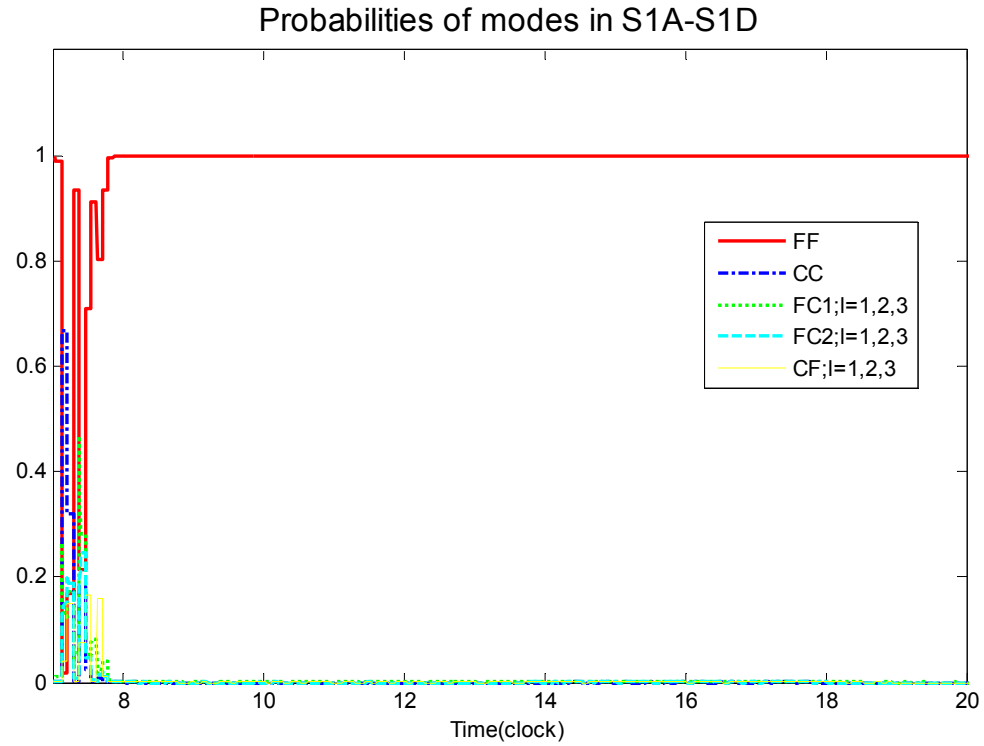
Figures 4.8 (a-i) present the simulation results generated by the SCTM given the calibrated parameters listed in Table 4.1, the measured segment inflow, outflow and on-/off- ramp flow rates on May 12, 2008 from 7:00 AM to 20:00 PM, the detected boundary densities every 5 minutes and the incident records listed in Table 4.3. The fundamental diagram parameters are adjusted in accordance with the lane closure conditions under traffic accidents (as long as the incident information is recorded). In the figures, the black lines denote the detected traffic densities along the route, and the colored lines show the estimated traffic densities with plus or minus one standard deviation.

As shown in Figure 4.8 (a)-(b), cells S1A-S1E maintain free flowing condition throughout the whole simulation period. This implies that this segment is not influenced by any spillback effect of the congestion caused by incidents. However, from cell S1F and onward (from upstream to downstream), all the downstream links suffer from congestion induced by the spillback effect caused by Jam 1898 originated from cells S5B-S6B (i.e. post kilometer 4 km to 5 km in the incident record) at 7:16 AM. The congestion wave-front spends about 25 minutes to reach cell S1F. We may deduce the average speed of this backward wave to be 15-16 km/hr, which is close to the calibrated values.

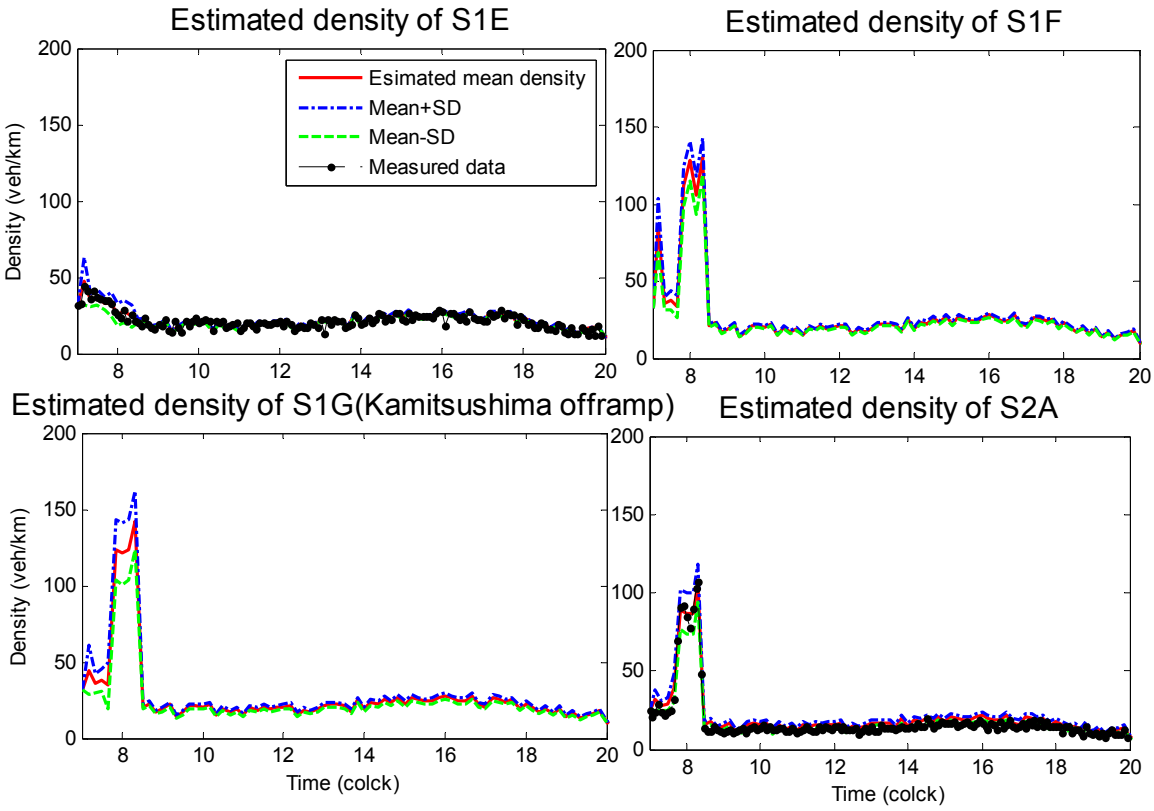
Jam 1909 and Jam 1910 are induced by Incident 1599 at cell S10A. The congestion wave propagates to upstream until it reaches cell S4D. During this spillback process, the congestion wave caused by Jam 1909 and Jam 1910 merges with that caused by Jam 1898 at S7A. The sudden increase of traffic density of cell S1C around 10:10 AM reflects the effect of Incident 1415. However, no obvious congestion is caused by this incident. Similarly, Incident 1417 does not lead to any significant disturbance to the traffic flow.

The estimated traffic density against the detected one presented in Figure 4.8 confirms that the SCTM performs well in stochastic traffic state estimation even under incident scenarios. Moreover, the model is also able to capture congestion formation, spillback and dissolve. However, there are still some discrepancy between simulation result and measurement because the related fundamental diagrams cannot be adjusted timely due to the lack of traffic data under incident. Finally, these stochastic densities are used to evaluate dynamic stochastic journey time and the subsequent journey time reliability analysis.

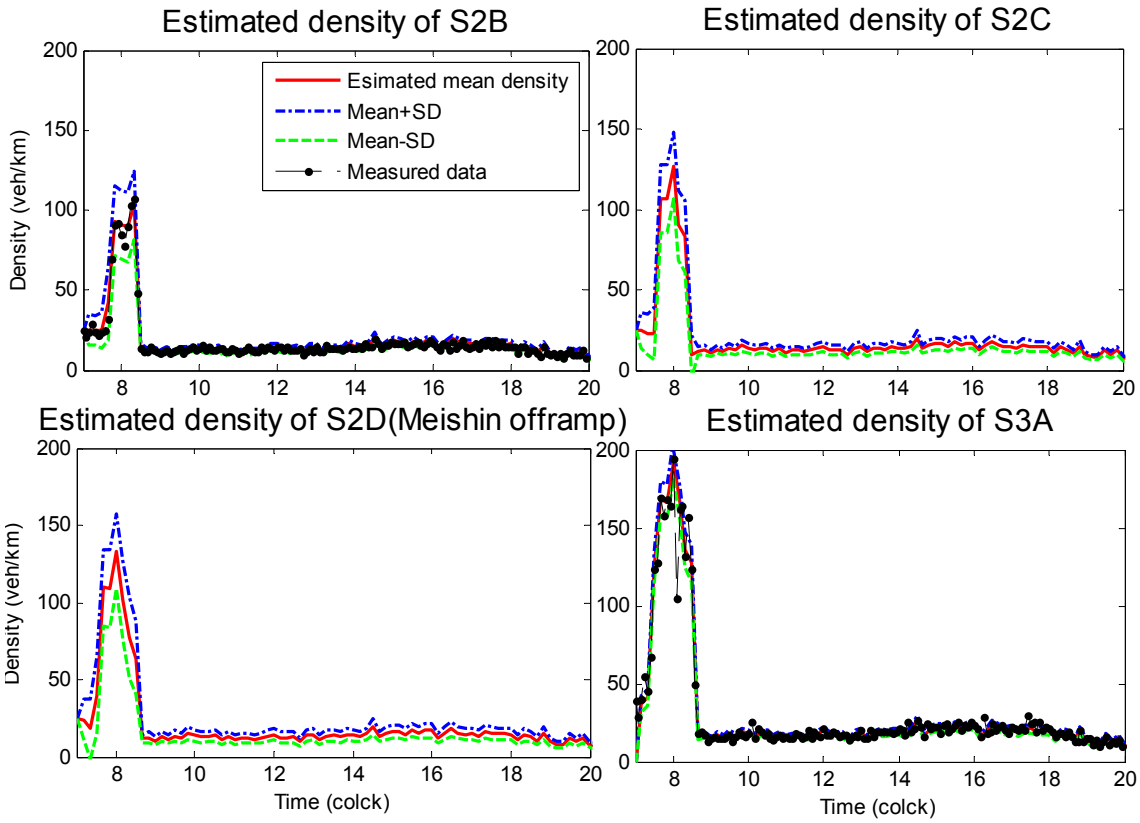




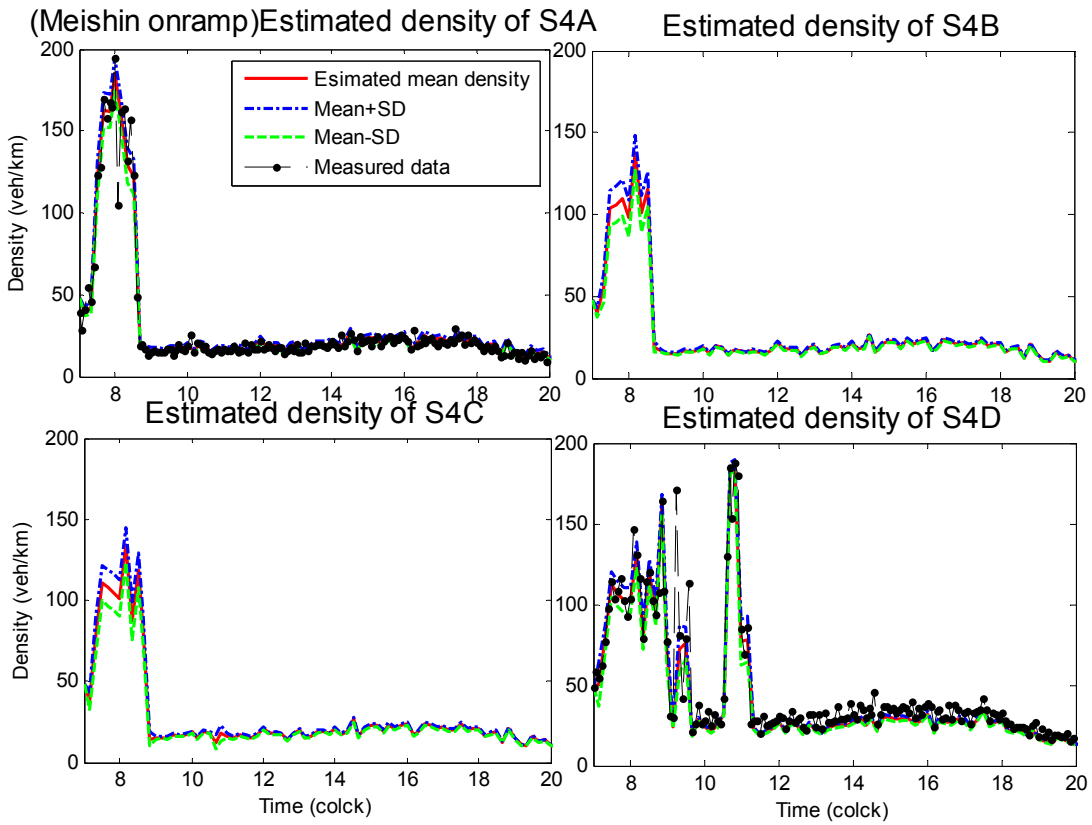
(a). Cell S1A-S1D

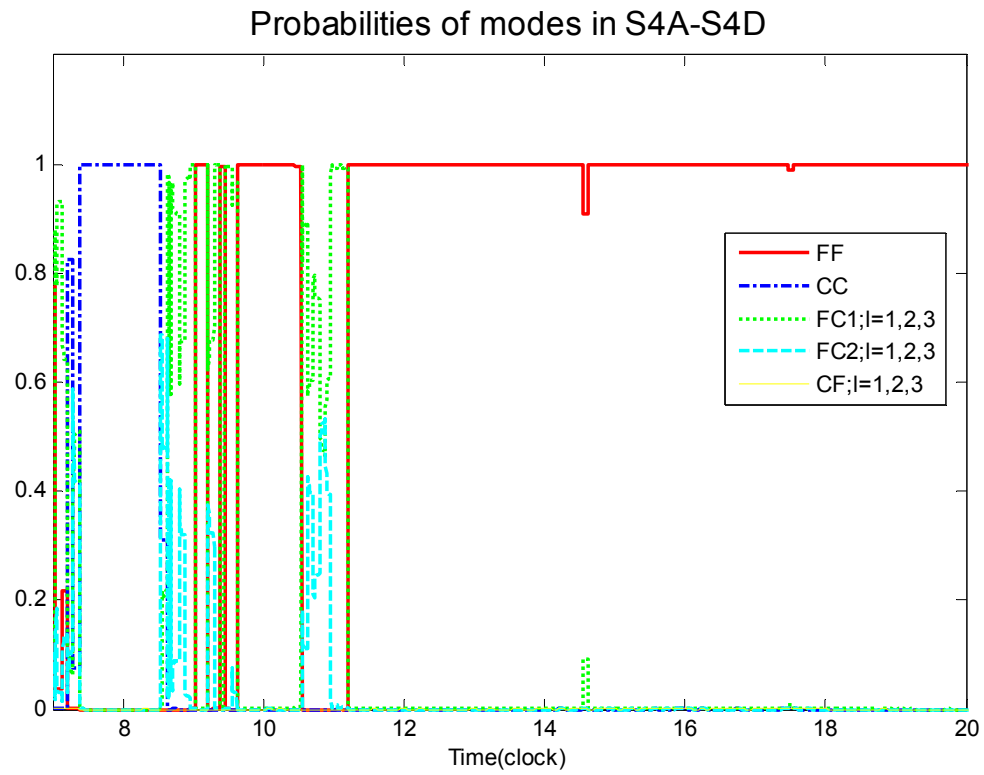


(b). Cell 1E-2A

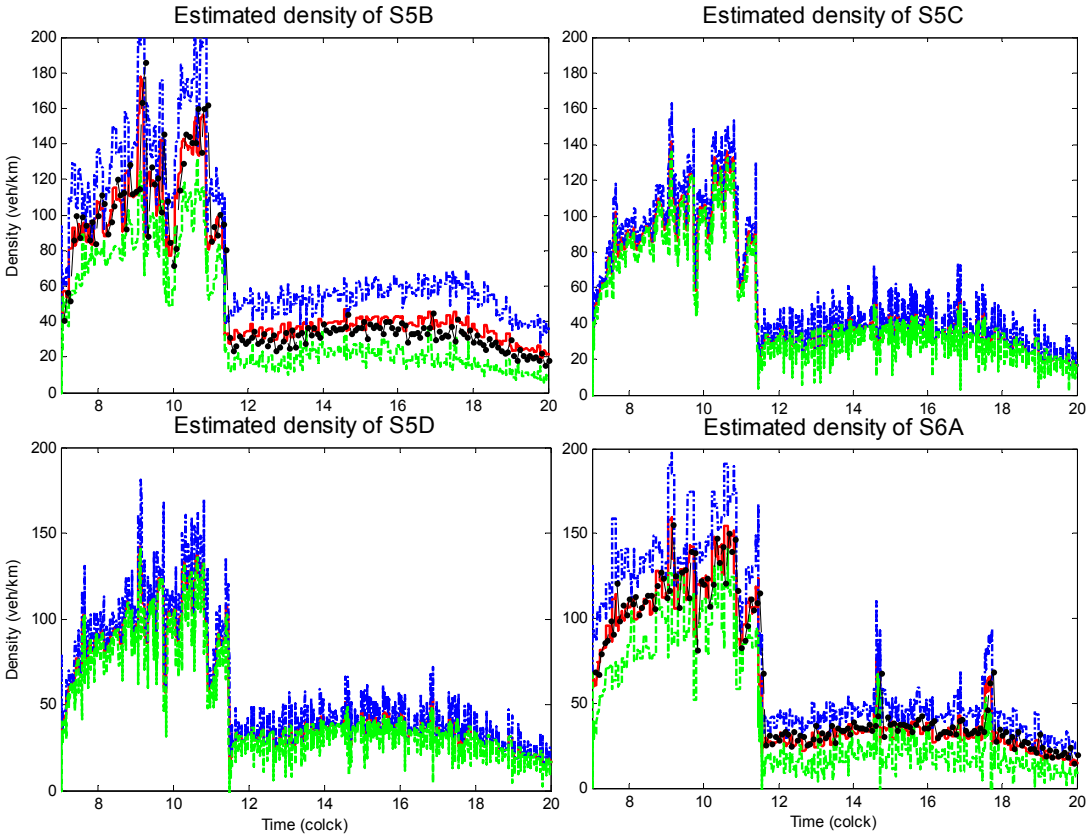

Estimated density of S2B
Estimated density of S2C
Esimated mean density
Mean+SD
Mean-SD
Measured data
Density (veh/km)
Estimated density of S2D(Meishin offramp)
Estimated density of S3A
Time (colck)

(c). Cell 2B-3A

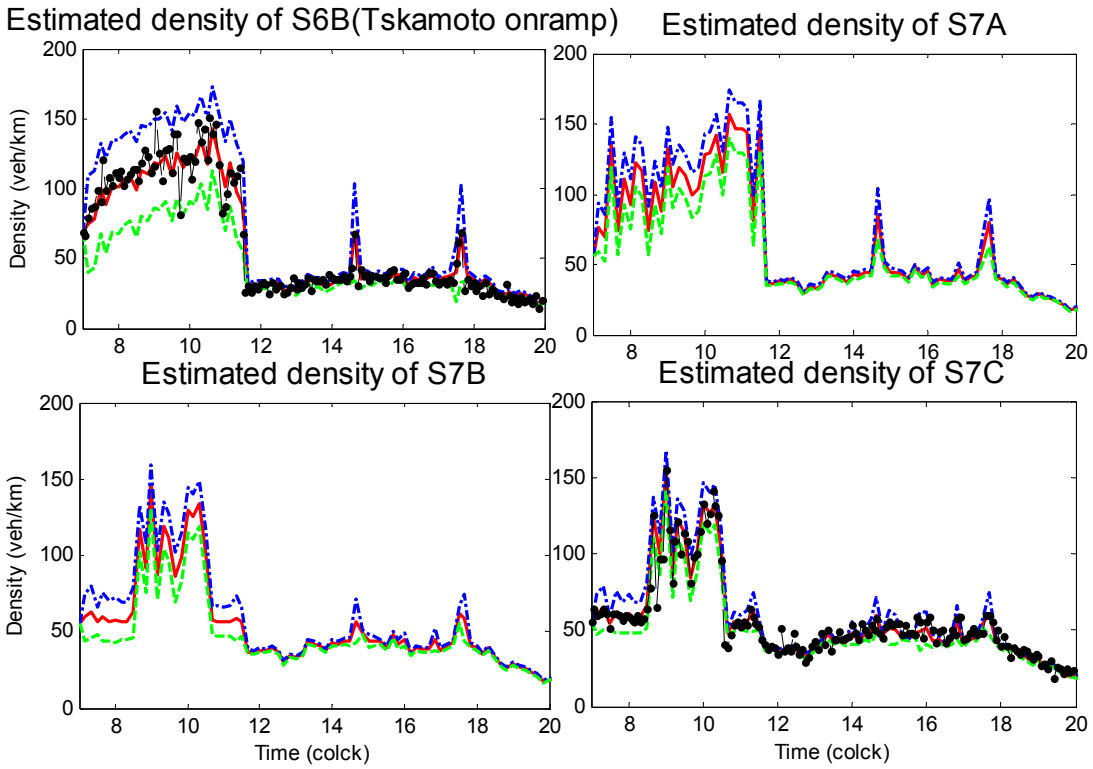

(Meishin onramp)Estimated density of S4A
Estimated density of S4B
Esimated mean density
Mean+SD
Mean-SD
Measured data
Density (veh/km)
Estimated density of S4C
Estimated density of S4D
Time (colck)



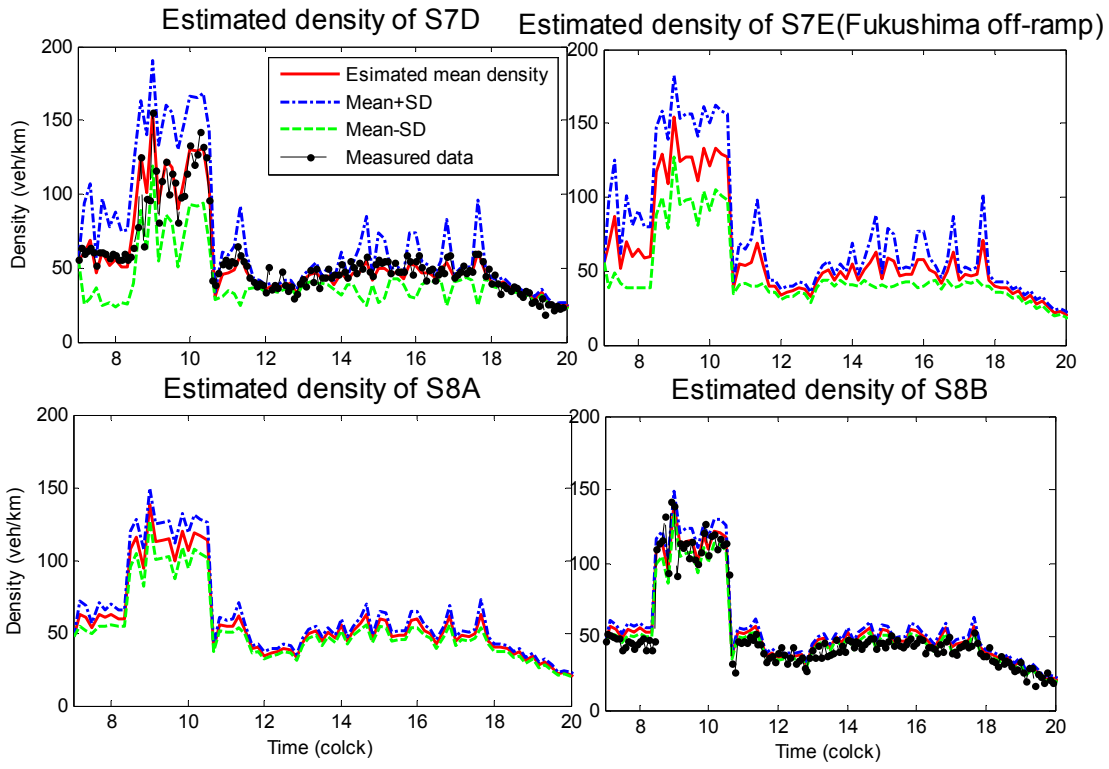
(d). Cell 4A-4D



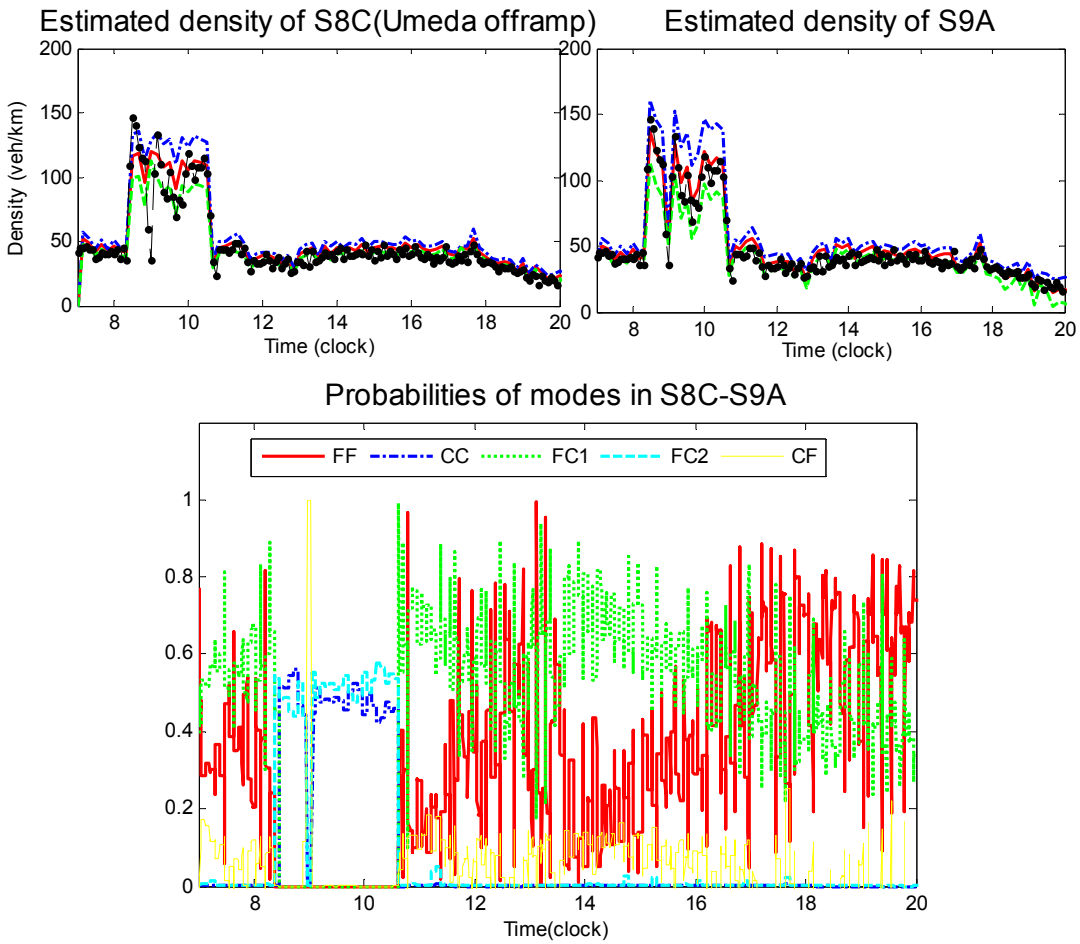
e). Cell 5B-6A



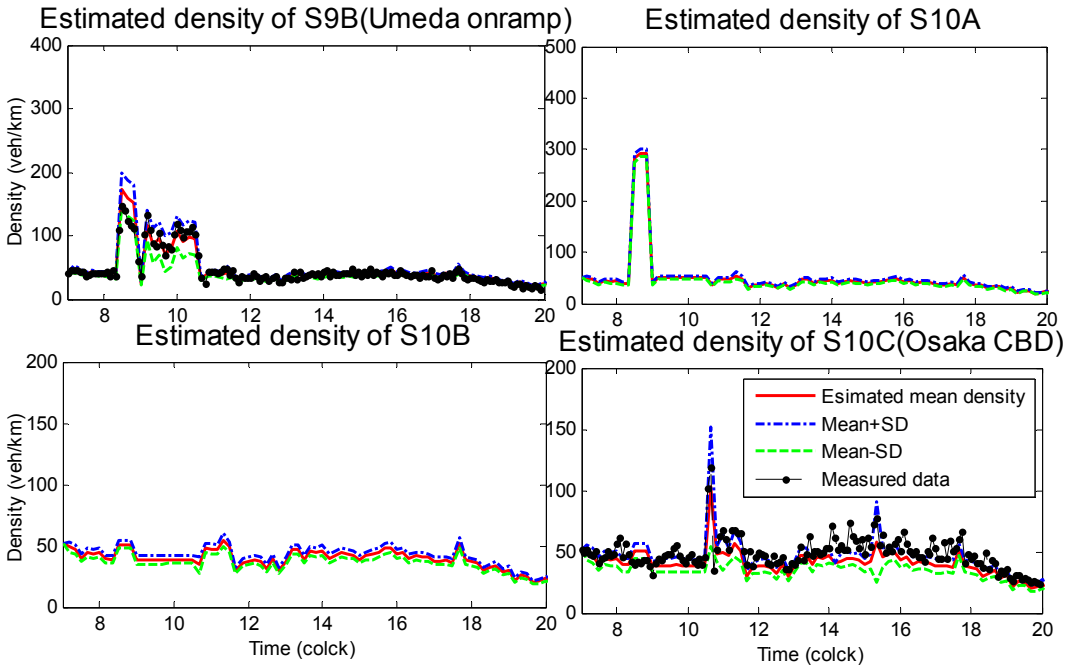
f). Cell 6B-7C



g). Cell 7D-8B



h). Cell 8C-9A



i). Cell 8C-10C

Figure 4.8. Measured and estimated density with 68% confidence interval of ten segments

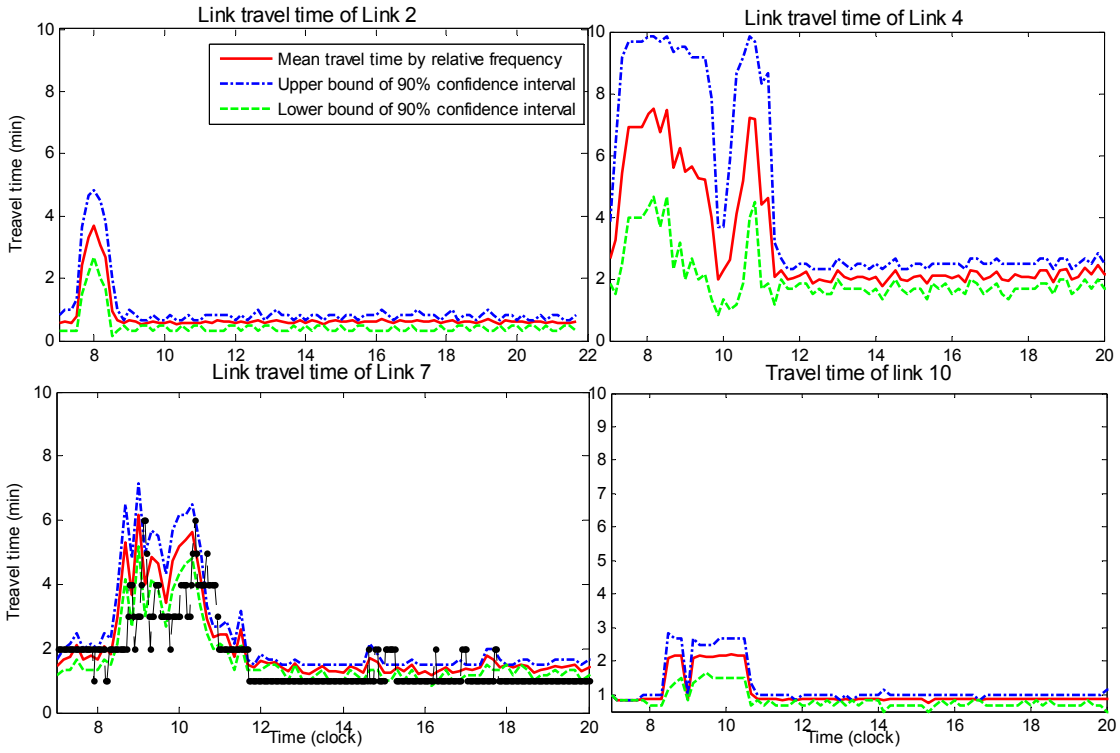


Figure 4.9. Link travel time estimation of four links

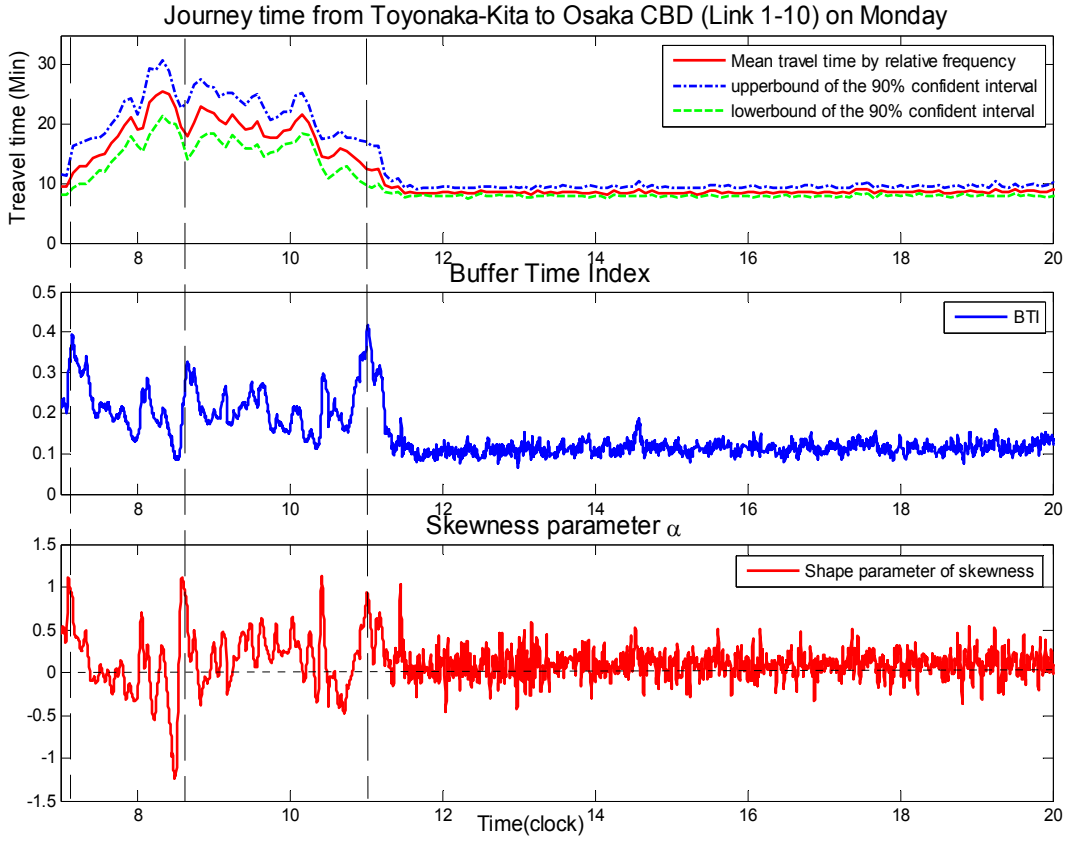


Figure 4.10. Journey time and reliability of Toyonaka-Kita to Osaka CBD

4.5.3 Link travel time and journey time estimations

Given the estimated stochastic dynamic cell densities as shown in Figure 4.8，dynamic link travel time distributions of the 9 links can be evaluated by the PMF based algorithm proposed in this chapter. Figure 4.9 depicts four representative links to illustrate the results, i.e. link 2, 4, 7 and 10, where link 7 is the only link that can provide detected travel time for comparison. The comparison confirms the good performance of the PMF based algorithm except for a little overestimation (indeed, the detected travel time is not accurate).

The figures reflect that the estimated travel times are in accordance with the traffic states on the links. For example, link 2 is congested during 7:30 AM to 8:30 AM as shown in Figure 4.8 (b-c), and the corresponding link travel time (as illustrated by Figure 4.9 (a)) during this period is greater than the travel time in any other time periods. Link 4 suffers from two congestions. As a consequence, its link travel time (as reflected by Figure 4.9 (b)) has two peaks. The estimated link travel time of link 7 (Figure 4.9 (c)) is consistent with the measured one in general, wherein the travel time arises at about 8:30 AM, maintaining a relative large value from 9:00 AM to about 10:15 AM. The congestion starts to dissolve from 10:15, and the dissolving progress lasts more than one hour. Travel time of link 10 reflects the effect of Incident 1599.

Since the journey time over the whole corridor (link 1 to link 10) is not directly provided by the database, Figure 4.10 depicts the journey time estimation results, the reliability index and the skewness parameter. At the beginning of the simulation horizon, the journey time is about 9 minutes. However, the journey time remains at such a small value only for about five minutes and starts to increase sharply since 7:05 AM due to Jam 1898. This congestion period generates a journal time up to 25 minutes, which is about 3 times of the free-flow time. At about 8:00 AM, the journey time reduces a bit due to the dispersion of Jam 1898, while it increases again due to Jam 1909 and Jam 1910 started from about 8:30 AM. Finally, the journey time reduces

to free flow time (9 minutes) after all these congestions are dissolved. Figure 4.10 also presents the Buffer Time Index (BTI) of the journey time in the second figure. The unreliable periods concentrate during 7:05 AM-11:30 AM due to the jams and incidents. The peaks of travel time reliability index appear at three times: (1) the congestion onset progress around 7:12 AM caused by Jam 1898; (2) the intersection of congestion dissolve of Jam 1898 and congestion onset of Jams 1909/1910 around 8:30 AM; (3) the final congestion dissolve around 11:15 AM.

Table 4.4 presents the MAPE (Mean Absolute Percentage Error) of estimated journey time compared with the available measured travel times. The results confirm a satisfactory performance of the proposed method.

Travel time	Path (link 1-2)	Path (link 1-7)	Path (link 1-8)
MAPE	9.93%	9.75%	9.78%

Table 4.4. MAPE of estimated journey time

4.5.4 Sensitivity analysis of travel time estimation

Table 4.5 presents the MAPE of journey time for a path (link1-2) with different pairs of ε for error bound definition of equal and the selection of i and j. The table illustrates that the accuracy of journey time generally remains almost the same given the following settings of and i, j and ε. This implies that the proposed method is not sensitive to the choice of these parameters.

i,j \ ε MAPE	0.5	1	2
$[-\sigma, \sigma]$	10.10%	10.10%	10.11%
$[-3\sigma, 3\sigma]$	9.94%	9.96%	10.00%
$[-4\sigma, 4\sigma]$	9.93%	9.94%	10.00%

Table 4.5. MAPE of journey time for one path (link 1-2)

In this empirical study, $[-3\sigma, 3\sigma]$ is chosen for the definition of sampling region $[k_{lb}, k_{ub}]$ in Equation (4.9) as a greater pair of *i*, *j* cannot improve the journey time estimation significantly but increase the computational effort. $\varepsilon = 1$ is used to evaluate the likelihood of link travel time $P'_{k'|k}$ in Equation (4.8).

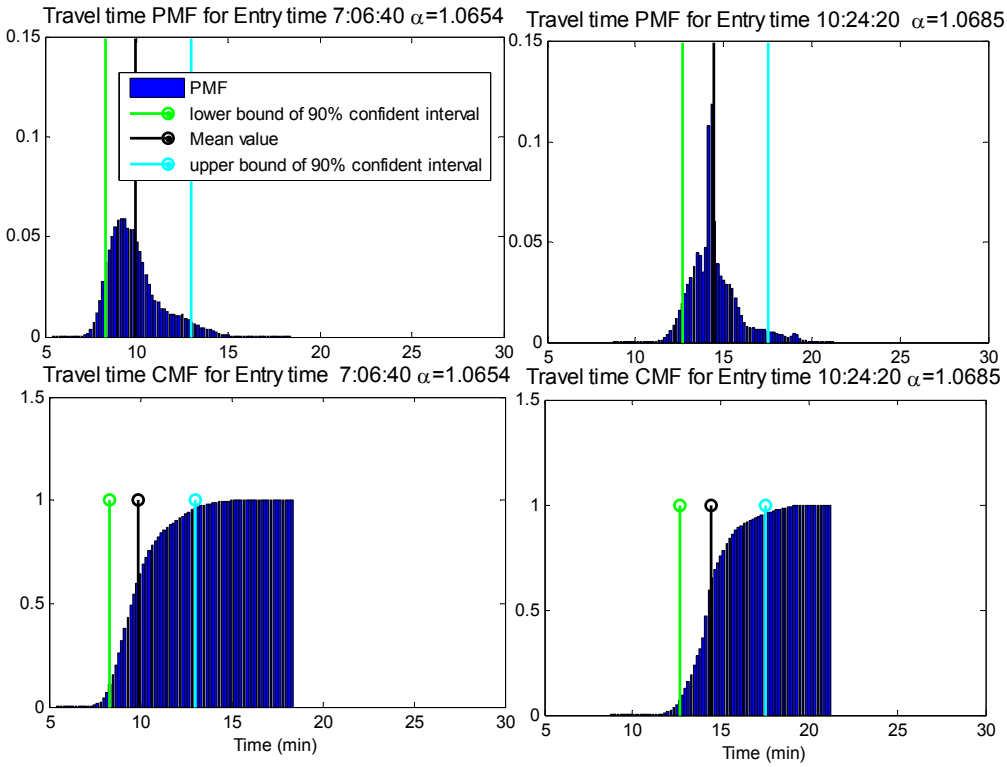
4.5.5 Skewness of journey time

From Figure 4.9 to Figure 4.10, it can be observed that the gaps between the mean travel time with the upper bound of 90% confidence interval and that with the lower bound are not equivalent. This implies that travel time is skewed distributed. Lint *et al.*, (2008) concluded the characteristics of skewness during the different periods of congestion:

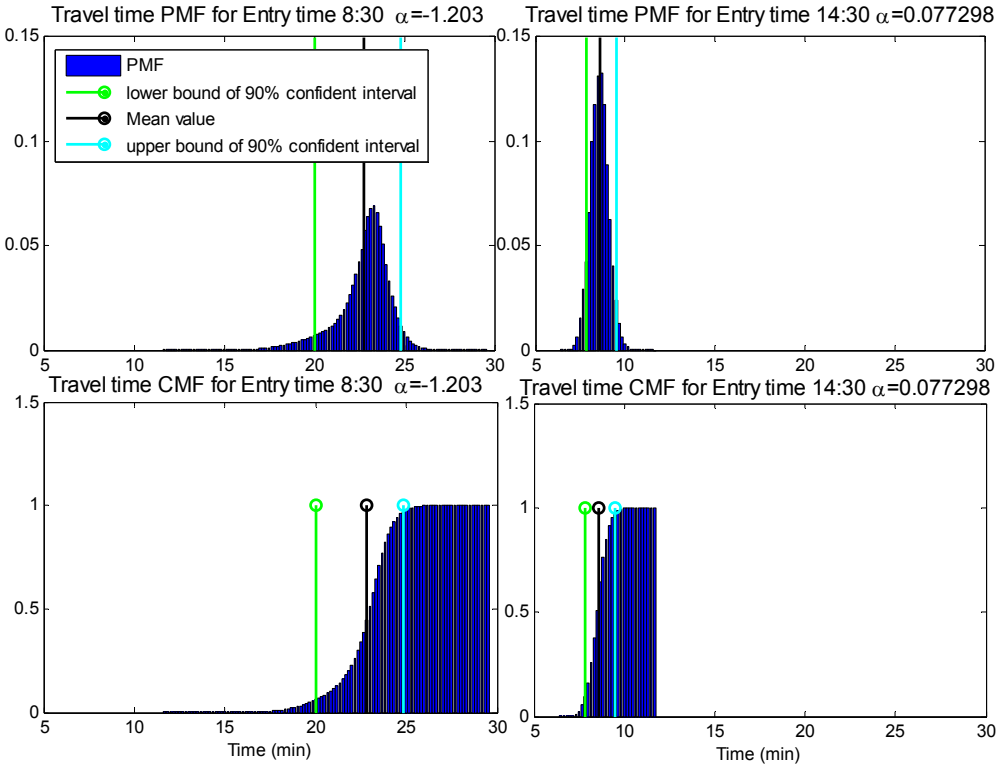
(1). Congestion onset: the mean travel times are increasing, and the distribution is positively skew.
(2). Congestion dissolve: the mean travel times are decreasing, and the distribution is positively skew.
(3). Congestion: the mean travel times are high, while the travel time distribution is wide and either symmetric or slightly negative skew.
(4). Free flow: the mean travel times are low, and the travel time distribution is narrow and either symmetric or slightly negatively skew.

The above conclusions were based on the analysis of statistics. However, in this research, similar results can also be observed from the simulation results for empirical study, such as the shape parameters of the stochastic travel time as depicted in the third figure of Figure 4.10. Generally, the journey times during 7:00 AM-11:30 AM are positively skew, as this period is composed of congestion onset and congestion dissolve except for two short stationary stages of congestion which are negatively skew. After 11:30 am, the journey times are approximately symmetric normally

distributed.



a. Positively skew journey time distribution



b. Negatively and symmetrically ally skew journey time distribution

Figure 4.11. Skewness analysis of journey time

In order to illustrate the characteristics of the skewness of journey time, Figure 4.11 depicts the PMFs and corresponding shape parameter α to assist the comprehension. The positively skew distributions in Figure 4.11 (a) confirm conclusion (1) and (2) by presenting the PMFs of the journey times during congestion onset progress at 7:06:40 AM and congestion dissolve progress at 10:24:20 AM. Figure 4.11 (b) confirms conclusion (3) and (4) by the negatively skew distributions at 8:30 AM under congested condition and the symmetric distribution at 14:30 PM under free flowing condition respectively.

4.5.6 Computational time of simulation

Simulation of this empirical study consists of three main parts: estimation of stochastic dynamic traffic densities for the 38 cells (from 7:00 AM-22:00 PM), evaluation of cumulative flows and PMFs of the link travel times of the 9 links, and calculation of journey time PMFs and skew normal distribution fitting. Table 4.6 lists the computational time for each item based on Matlab 2009b, CPU: Intel Core 2 Duo E8500 @3.16GHz and 4G DDR3 1067MHz.

Computational items	Computational time (second)
Density estimation	404
Link time estimation	140
Journey time estimation	38
Total	582

Table 4.6. Computational time

4.6 Summary of the empirical study and further discussions

Chapter 4 proposed an algorithm to estimate the dynamic stochastic journey time distribution and to access the dynamic journey time reliability based on the stochastic cell transmission model (as described in Chapter 3). The SCTM generates the stochastic cumulative link inflow and outflow profiles. The algorithm for calculating

the probability mass function (PMF) of the stochastic dynamic journey time was proposed by devising the sampling process of the cumulative flows for each entry time interval. The journey time is then estimated by extending the nested delay operator to the stochastic case based on the conditional probability. This chapter also adopted the nonlinear curve fitting method in Matlab to fit the estimated PMF of the journey time to a class of skew normal distribution to determine the skewness of the journey time distribution. The model and the algorithm were validated via an empirical study conducted on a long segment of an expressway between Toyanaka-kita and Osaka CBD. The empirical study illustrated that the model and the algorithm are capable of simulating the complicate traffic conditions caused by the jams due to different incidents and reconstructing the stochastic dynamic travel time distributions.

Appendix

A.1 Variance of the matching error

To calculate the variance of the matching error, it is convenience for us to express it in a vector form so as to make use of the property of multivariate normal distribution (Gut, 2009; Hardle and Simar, 2007; Johnson and Wichern, 2007). Let

$$a = \left(-s_1^{l^m}, -s_2^{l^m}, \cdots, -s_{NC_{l^m}}^{l^m}, \Delta t, \cdots, \Delta t \right)^T \in R^{\ell_1},$$

and

$$\mathbb{Q} = \left(\rho_1^{l^m}(k), \rho_2^{l^m}(k), \cdots, \rho_{NC_{l^m}}^{l^m}(k), q_{out}^{l^m}(k+1), \cdots, q_{out}^{l^m}(k') \right)^T .$$

The matching error is then can be written as $e_k(k') = a^T \mathbb{Q}$. By assumption, $\mathbb{Q}$ has a multivariate normal distribution. Then, $e_k(k')$ is a normally distributed random variable as every linear combination of normally distributed variables is normally distributed (Gut, 2009; H¨ardle and Simar, 2007; Johnson and Wichern, 2007). The variance of the matching error is now

$$Var\left(e_k(k')\right) = Var\left(a^T\mathbb{Q}\right) = a^T Var(\mathbb{Q})a\,,$$

where the covariance matrix $Var(\mathbb{Q})$ can be obtained from the SCTM. For the special case, where the components of $\mathbb{Q}$ are assumed to be independent, $Var(\mathbb{Q})$ is a diagonal matrix which is denoted as Γ_d. By this observation, we may have more interesting results by exploring the relation between the independent case and the dependent case. To begin with, we first introduce the following lemma, which states we can always find a linear transformation, e.g. Mahalanobis transformation, to transfer a vector of multivariate normal distribution with dependent components into a random vector with independent (standard) normal random variables (Gut, 2009; H¨ardle and Simar, 2007; Johnson and Wichern, 2007):

Lemma A.1 For a random vector $\mathbb{Q} \sim N\left(\mu, \Lambda\right)$ has a multivariate normal distribution, we have:

- There exists a random ℓ_2-vector *Z*, whose components are independent standard normal random variables, a ℓ_1-vector μ, and a $\ell_1 \times \ell_2$ matrix *A*, such that $\mathbb{Q} = AZ + \mu$. The covariance matrix of $\mathbb{Q}$ is then given by $\Gamma = AA^T$ with $\text{rank}\,\Gamma = \ell_2$.
- Let $H = C^T\mathbb{Q}$, where the orthogonal matrix *C* is such that $C^T\Lambda C = D$. Then $H \sim N\left(C^T\mu, D\right)$. Moreover, the components of *H* are independent and $VarY_k = \lambda_k, k = 1,2,\cdots,\ell_1$, where $\lambda_1, \lambda_1, \cdots, \lambda_{\ell_1}$ are the eigenvalues of Λ.
- For the Mahalanobis transformation $Y = \Lambda^{-1/2}\left(\mathbb{Q} - \mu\right)$, we have $Y \sim N\left(0, I\right)$, i.e., the components of *Y* are independent standard normal random variables.

By this lemma, we can always find some invertible linear transformation (i.e. isomorphism), such that the original correlated random vector can be transferred to an independent one. Also note that the variance of $e_k(k')$ with correlated components of

$\mathbb{Q}$, say $Var_1\left(e_k(k')\right)$, is a linear functions of that with independent components of $\mathbb{Q}$, say $Var_2\left(e_k(k')\right)$, i.e. $Var_1\left(e_k(k')\right) = b_{k,k'}Var_2\left(e_k(k')\right)$, where $b_{k,k'}$ is certain constant. Thus, the likelihood measure given by (4.8) may be changed. However, the relative frequency given by (4.10) may not change significantly. The idea of coordinate transformation is commonly utilized in statistics and signal processing theory, e.g. the unscented Kalman filter.

A.2 Summations of the nested probabilities

In this appendix, we will show that summations of the nested probabilities over the corresponding sampling time steps equal to one for each departure time *k*. The proof is established by mathematical induction in terms of link sequential order. For the first link l_1 of a route and a given departure time *k*, we have that

$$\sum_{\tau^1=k_{lb}^{l_1}}^{k_{ub}^{l_1}} P_{1,nest}(\tau^1 \mid k) = \sum_{\tau^1=k_{lb}^{l_1}}^{k_{ub}^{l_1}} P_1(\tau^1 \mid k)$$

$$= \sum_{k'=k_{lb}^{l_1}}^{k_{ub}^{l_1}} P_{k'|k} = \sum_{k'=k_{lb}^{l_1}}^{k_{ub}^{l_1}} \frac{P'_{k'|k}}{\sum_{k_{lb}^{l_1}}^{k_{ub}^{l_1}} P'_{k'|k}} = 1.$$

Next, we will show that the nested probability of link l_2 over the corresponding sampling time steps equals to one for the departure time *k*.

$$\sum_{\tau^2=\left\{k_{lb}^{l_1}\right\}_{lb}^{l_2}}^{\left\{k_{ub}^{l_1}\right\}_{ub}^{l_2}} P_{2,nest}(\tau^2 \mid k) = \sum_{\tau^2=\left\{k_{lb}^{l_1}\right\}_{lb}^{l_2}}^{\left\{k_{ub}^{l_1}\right\}_{ub}^{l_2}} \sum_{\tau^1=k_{lb}^{l_1}}^{k_{ub}^{l_1}} P_2(\tau^2 \mid \tau^1) P_{1,nest}(\tau^1 \mid k).$$

Now, we interchange the order of summation so that

$$\sum_{\tau^2=\left(\left(k_{lb}^{1}\right)\right)_{lb}^{l_2}}^{\left(\left(k_{ub}^{1}\right)\right)_{ub}^{l_2}} P_{2,nest}(\tau^2 \mid k) = \sum_{\tau^1=k_{lb}^{1}}^{k_{ub}^{1}} \left(\sum_{\tau^2=\left(\left(k_{lb}^{1}\right)\right)_{lb}^{l_2}}^{\left(\left(k_{ub}^{1}\right)\right)_{ub}^{l_2}} P_2(\tau^2 \mid \tau^1) \right) P_{1,nest}(\tau^1 \mid k)$$

$$= \sum_{\tau^1=k_{lb}^{1}}^{k_{ub}^{1}} P_{1,nest}(\tau^1 \mid k) = 1.$$

Now we assume this is hold for link l_m on the route, that is

$$\sum_{\tau^m=\left(\left(\cdots\left(k_{lb}^{1}\right)\right)\cdots\right)_{lb}^{l_m}}^{\left(\left(\cdots\left(k_{ub}^{1}\right)\right)\cdots\right)_{ub}^{l_m}} P_{m,nest}(\tau^m \mid k) = 1.$$

Then for link l_{m+1}, we have

$$\sum_{\tau^{m+1}=\left(\left(\cdots\left(k_{lb}^{1}\right)\right)\cdots\right)_{lb}^{l_{m+1}}}^{\left(\left(\cdots\left(k_{ub}^{1}\right)\right)\cdots\right)_{ub}^{l_{m+1}}} P_{m+1,nest}(\tau^{m+1} \mid k) = \sum_{\tau^{m+1}=\left(\left(\cdots\left(k_{lb}^{1}\right)\right)\cdots\right)_{lb}^{l_{m+1}}}^{\left(\left(\cdots\left(k_{ub}^{1}\right)\right)\cdots\right)_{ub}^{l_{m+1}}} \sum_{\tau^m=\left(\left(\cdots\left(k_{lb}^{1}\right)\right)\cdots\right)_{lb}^{l_m}}^{\left(\left(\cdots\left(k_{ub}^{1}\right)\right)\cdots\right)_{ub}^{l_m}} P_{m+1}(\tau^{m+1} \mid \tau^m) P_{m,nest}(\tau^m \mid k).$$

By interchanging the order of summation, we have

$$\sum_{\tau^{m+1}=\left(\left(\cdots\left(k_{lb}^{1}\right)\right)\cdots\right)_{lb}^{l_{m+1}}}^{\left(\left(\cdots\left(k_{ub}^{1}\right)\right)\cdots\right)_{ub}^{l_{m+1}}} P_{m+1,nest}(\tau^{m+1} \mid k) = \sum_{\tau^m=\left(\left(\cdots\left(k_{lb}^{1}\right)\right)\cdots\right)_{lb}^{l_m}}^{\left(\left(\cdots\left(k_{ub}^{1}\right)\right)\cdots\right)_{ub}^{l_m}} \left(\sum_{\tau^{m+1}=\left(\left(\cdots\left(k_{lb}^{1}\right)\right)\cdots\right)_{lb}^{l_{m+1}}}^{\left(\left(\cdots\left(k_{ub}^{1}\right)\right)\cdots\right)_{ub}^{l_{m+1}}} P_{m+1}(\tau^{m+1} \mid \tau^m) \right) P_{m,nest}(\tau^m \mid k).$$

By definition of the relative frequency (4.10), we have that

$$\sum_{\tau^{m+1}=\left(\left(\cdots\left(k_{lb}^{1}\right)\right)\cdots\right)_{lb}^{l_{m+1}}}^{\left(\left(\cdots\left(k_{ub}^{1}\right)\right)\cdots\right)_{ub}^{l_{m+1}}} P_{m+1}(\tau^{m+1} \mid \tau^m) = 1.$$

Thus

$$\sum_{\tau^{m+1}=\left(\left(\begin{smallmatrix} l \\ k1 \\ lb \end{smallmatrix}\right)\cdots\right)_{lb}^{l_{m+1}}}^{\left(\left(\begin{smallmatrix} l \\ k1 \\ ub \end{smallmatrix}\right)\cdots\right)_{ub}^{l_{m+1}}} P_{m+1,nest}(\tau^{m+1} \mid k) = \sum_{\tau^{m}=\left(\left(\begin{smallmatrix} l \\ k1 \\ lb \end{smallmatrix}\right)\cdots\right)_{lb}^{l_{m}}}^{\left(\left(\begin{smallmatrix} l \\ k1 \\ ub \end{smallmatrix}\right)\cdots\right)_{ub}^{l_{m}}} P_{m,nest}(\tau^{m} \mid k).$$

But

$$\sum_{\tau^{m}=\left(\left(\begin{smallmatrix} l \\ k1 \\ lb \end{smallmatrix}\right)\cdots\right)_{lb}^{l_{m}}}^{\left(\left(\begin{smallmatrix} l \\ k1 \\ ub \end{smallmatrix}\right)\cdots\right)_{ub}^{l_{m}}} P_{m,nest}(\tau^{m} \mid k) = 1$$

by assumption. Therefore,

$$\sum_{\tau^{m+1}=\left(\left(\begin{smallmatrix} l \\ k1 \\ lb \end{smallmatrix}\right)\cdots\right)_{lb}^{l_{m+1}}}^{\left(\left(\begin{smallmatrix} l \\ k1 \\ ub \end{smallmatrix}\right)\cdots\right)_{ub}^{l_{m+1}}} P_{m+1,nest}(\tau^{m+1} \mid k) = 1,$$

which concludes the proof.

Chapter 5

Traffic state prediction by considering temporal and spatial correlation

As we have mentioned in Chapter 3, the SCTM is originally proposed for traffic state estimation under several assumptions, e.g. the independent/uncorrelated assumption of the underlying stochastic processes governing the demand and supply uncertainties. Note that the traffic flow, by nature, is correlated in both spatial and time domains due to its dynamics and flow propagation. However, the development of the SCTM framework, to be specific the uncorrelated assumption, ignores this issue. The SCTM framework as introduced in Chapter 3 does not address the issue of traffic state prediction which may also prevent the model from a broad range of applications. In this chapter, we aim to extend the SCTM framework to address these problems. First of all, a predictor is adopted as an auxiliary dynamical system to the original SCTM. This predictor will accept the historical spatial-temporal correlated traffic flow data and real-time detected data as inputs to forecast the demand profile and supply functions. The predicted demand profile and supply functions will be taken as inputs to the SCTM to calculate the traffic states. Then we aim to relax the independent/uncorrelated assumption of the model. Moreover, a rolling horizon algorithm is developed to further support traffic state forecasting in real-time manner.

This chapter is organized as: Section 5.1 introduces the phenomena and measurement of temporal and spatial correlations of traffic flow. Section 5.2 introduces a multivariable best linear predictor to extend the SCTM to consider the temporal and spatial correlations of traffic flow. Section 5.3 reviews the three steps in SCTM and adjusts them by considering spatial correlations. Section 5.4 introduces the concept of rolling horizon and proposes a framework for real-time prediction. Finally, Section 5.5 conducts an empirical study for journey time prediction on a segment of I210-W and compares the predicting results with and without considering the correlations.

5.1 Temporal and spatial correlation phenomena and measurement

Spatial-temporal correlations, which may reflect the spatial and temporal similarities of the traffic flow and is recognized as one of the characteristics to describe the relativity of traffic phenomenon, can be utilized to predict short-term traffic state. Rather than introducing the mathematical definitions of these correlations, we would like to give an intuitive example by analyzing the detected traffic flow data from a 2-mile freeway segment on I210-W in California. Figure 5.1-5.4 depicts the temporal and spatial correlated traffic flow phenomena via the historical traffic data of the freeway segment, e.g. the density and inflow profiles detected at locations A, B and C on Tuesday, Wednesday, and Thursday of March and April 2008 and 2009 (see Figure 5.9 and Figure 5.10).

Figure 5.1(a) shows the temporal similarity (or correlation) of the flow profile detected at location A for the three weekdays from 5:00 to 6:00 A.M with a detection frequency (i.e. the sample time interval) of 5 minutes. It is observed that these flow profiles are positively correlated at the adjacent time steps, which implies that the higher value a certain flow profile at a certain time step, the flow profile intends to have a larger value in the near future. For example, the flow profile of March 12 is larger than that of 25 March at 5:05 AM the positive correlation implies this will hold for 5:10 AM. Temporal covariance and correlation coefficient of $q(k)$ between two adjacent time steps k and k-l are defined as:

$$\mathrm{Cov}_{tm}(q(k),q(k-l)) = \frac{\sum_{i=1}^{N}\left(q_i(k)-\overline{q}(k)\right)\left(q_i(k-l)-\overline{q}(k-l)\right)}{N-1} \tag{5.1}$$

$$\gamma_{tm}(q(k),q(k-l)) = \frac{\mathrm{Cov}_{tm}(q(k),q(k-l))}{s_{q(k)}s_{q(k-l)}}, \quad l=1,2\cdots,l_e, l_e \le k-1 \tag{5.2}$$

where $q_i(k)$ is the value of the detected flow during interval k on the i^{th} day, l is a positive integer less than or equal to l_e which is a predefined bound (also an integer)

for the calculation, and $s_{q(k)}$ denotes the standard deviation of the detected flow $q(k)$ for the N sample days. Figure 5.1(b) presents the γ_{tm} of the flow profiles detected at location A with three different values of l (e.g. distances of 5 minutes, 30 minutes and 55 minutes in the figure) from 4:00 AM to 12:00 AM over N=54 days. The result illustrates that the temporal correlations of the flow profiles will generally decrease as l increase, which can be interpreted as the flow pattern of a given time is more similar to a near flow pattern than a distant one, e.g.

$$\gamma_{tm}(q(5:30),q(5:25)) > \gamma_{tm}(q(5:30),q(5:00)) > \gamma_{tm}(q(5:30),q(4:35)).$$

Temporal correlation of the parameters of a fundamental diagram (or supply functions) for a given location also can be analyzed similarly:

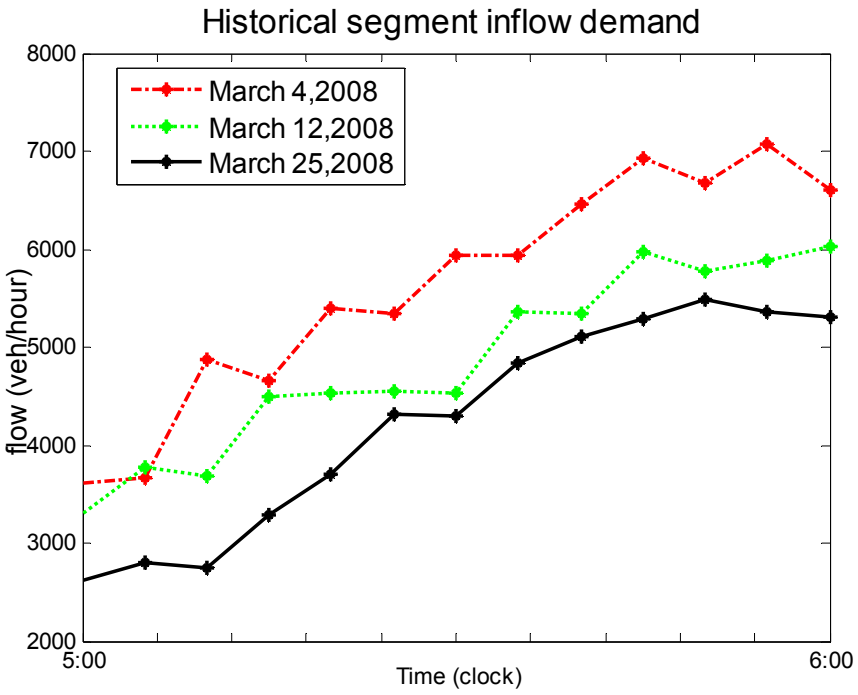
$$\mathrm{Cov}_{tm}(\delta(k_p),\delta(k_p-l_p)) = \frac{\sum_{i=1}^{N}\left(\delta_i(k_p)-\overline{\delta}(k_p)\right)\left(\delta_i(k_p-l_p)-\overline{\delta}(k_p-l_p)\right)}{N-1} \tag{5.3}$$

$$\gamma_{tm}(\delta(k_p),\delta(k_p-l_p)) = \frac{\mathrm{Cov}_{tm}(\delta(k_p),\delta(k_p-l_p))}{s_{\delta(k_p)}s_{\delta(k_p-l_p)}},\quad l_p=1,2\cdots,l_{p,\varepsilon},\, l_{p,\varepsilon}<k_p \tag{5.4}$$

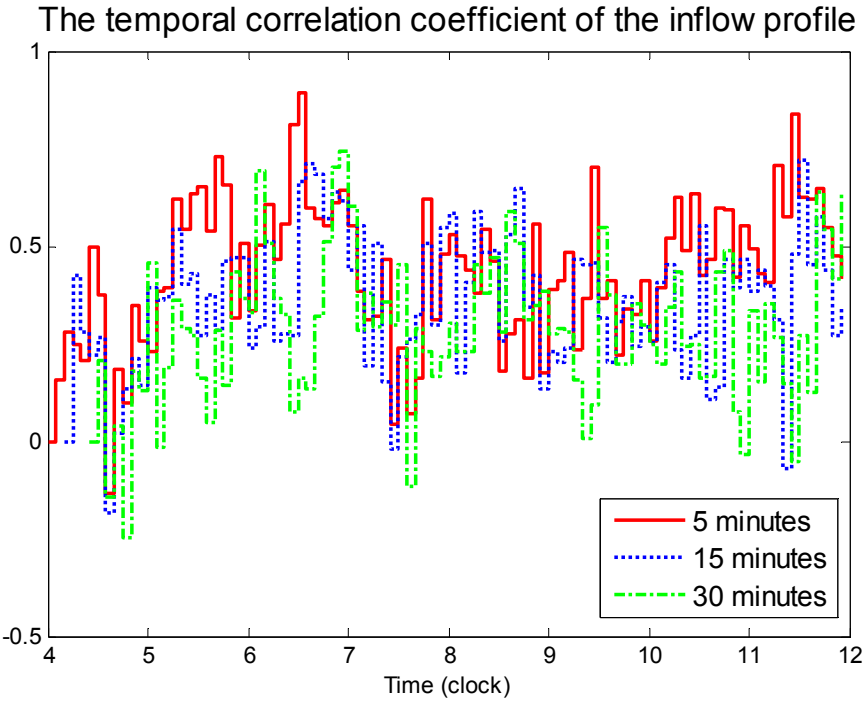
where $\delta_i(k_p)$ is a certain parameter calibrated with data detected during the time period k_p (e.g. half an hour) on the i^{th} day. $\overline{\delta}(k_p)$ is the mean of $\delta\left(k_p\right)$, and $s_{\delta}(k_p)$ is the corresponding standard deviation over the N days.

As shown in Figure 5.2, temporal correlations of free-flow speed v_f during the non-rush hours are much more evident than the correlations of wave-back speed w_c during the rush hours. Compared with the analysis of detected flow patterns, the supply uncertainties are more complicated because the flow-density relationship might not always constitute a complete triangular fundamental diagram, e.g. w_c cannot be observed during 4:00-6:00 AM while v_f cannot be evaluated during rush hours 7:00-9:00 AM. This is reflected in the figure as zero entries. Because the temporal

correlation coefficients of Q_m, ρ_c and ρ_J are very small, the results are omitted for brevity.



a). Flow profiles of the three days between 5:00-6:00



b).Temporal correlation coefficient of the flow

Figure 5.1. Temporal correlation of the inflow profiles

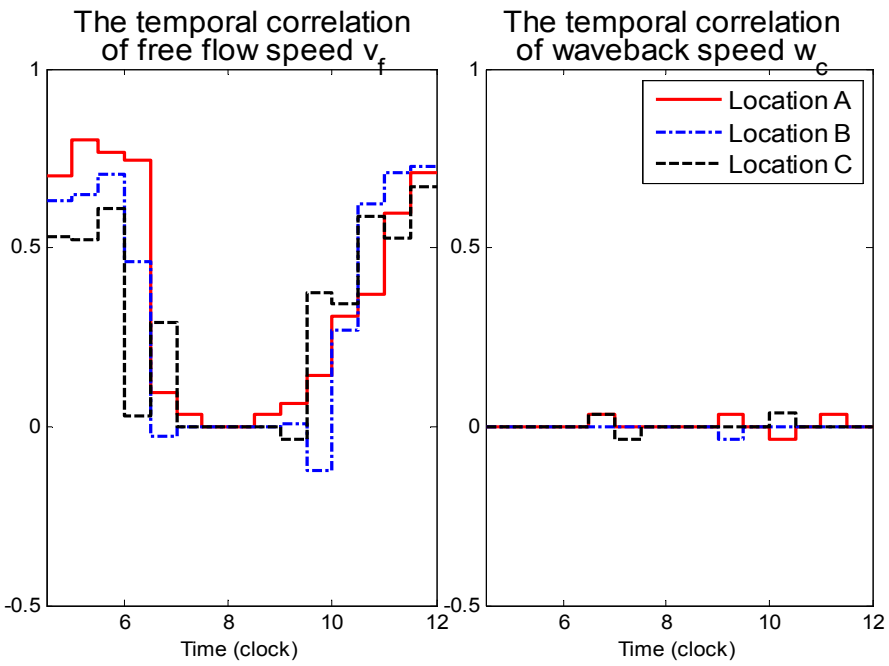


Figure 5.2. Temporal correlation coefficients of the supply functions

Figure 5.3(a) depicts the spatial correlation of the free-flow speeds calibrated at three adjacent sites based on the flow-density data detected during 4:00-4:30AM for two days (April 24, 2008 and April 21, 2009) under different weather conditions. In this example, the free-flow speeds are found to be positively spatially correlated which implies if the free-flow speed at site A (S. Myrtle) increase/decrease, then the free-flow speed at site B (W. Huntington Dr) and site C (N. Santa Anita) will increase/decrease accordingly. The spatial covariance and correlation coefficient of the free-flow speed of three adjacent sites during time interval k_p is defined as:

$$\mathrm{Cov}_{\mathrm{sp}}(\delta_m(k_p),\delta_n(k_p)) = \frac{\sum_{i=1}^{N}\left(\delta_{m,i}(k_p)-\bar{\delta}_m(k_p)\right)\left(\delta_{n,i}(k_p)-\bar{\delta}_n(k_p)\right)}{N-1} \tag{5.5}$$

$$\gamma_{\mathrm{sp}}(\delta_m(k_p),\delta_n(k_p)) = \frac{\mathrm{Cov}_{\mathrm{sp}}(\delta_m(k_p),\delta_n(k_p))}{s_{\delta_m(k_p)}s_{\delta_n(k_p)}} \tag{5.6}$$

where the notations in equation (5.5) and (5.6) are similar with the ones in (5.3) and (5.4) but for the differentiation of $\delta_m(k_p)$ and $\delta_n(k_p)$ denoted by *m* and *n* which represent the adjacent locations. Figure 5.3(b) shows the spatial correlation coefficients of free-flow speed and congestion wave speed among the three locations.

Figure 5.4 presents the spatial correlation of densities by:

$$\mathrm{Cov}_{\mathrm{sp}}(\rho_m(k),\rho_n(k)) = \frac{\sum_{i=1}^{N}\left(\rho_{m,i}(k)-\overline{\rho}_m(k)\right)\left(\rho_{n,i}(k)-\overline{\rho}_n(k)\right)}{N-1} \tag{5.7}$$

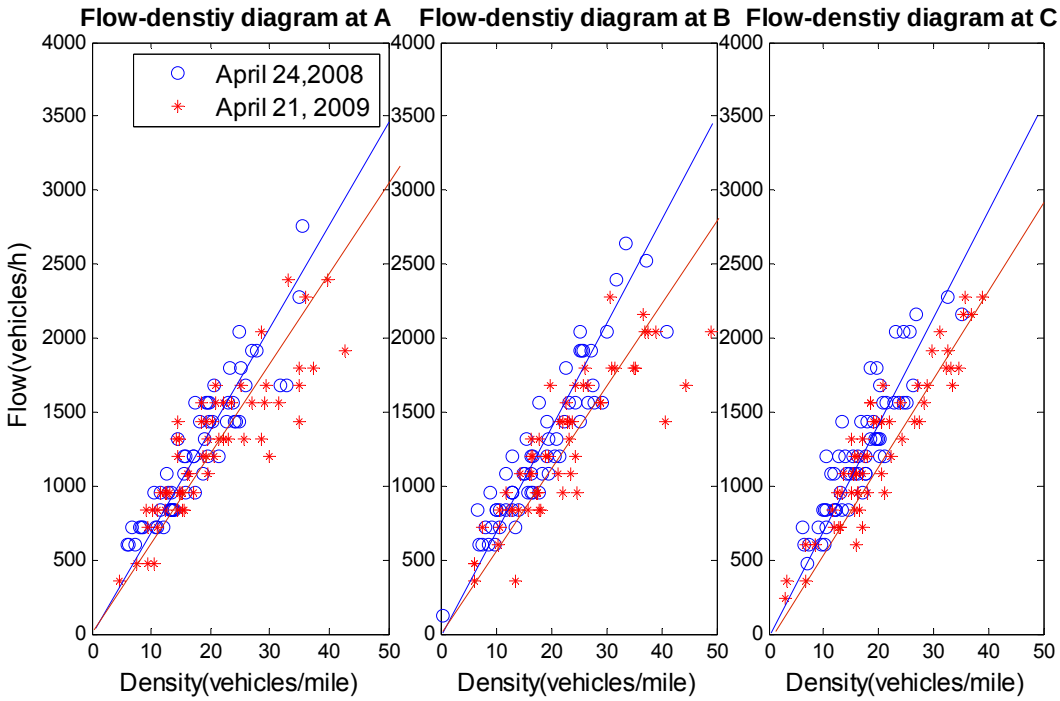
$$\gamma_{\mathrm{sp}}(\rho_m(k),\rho_n(k)) = \frac{\mathrm{Cov}_{\mathrm{sp}}(\rho_m(k),\rho_n(k))}{s_{\rho_m(k)}s_{\rho_n(k)}} \tag{5.8}$$

$\rho_{m,i}(k)$ is the traffic density detected at location *m* during time interval *k* on the i^{th} day, $\overline{\rho}_m(k)$ is the mean density over the *N* days, and $s_{\rho_m(k)}$ is the corresponding standard deviation.

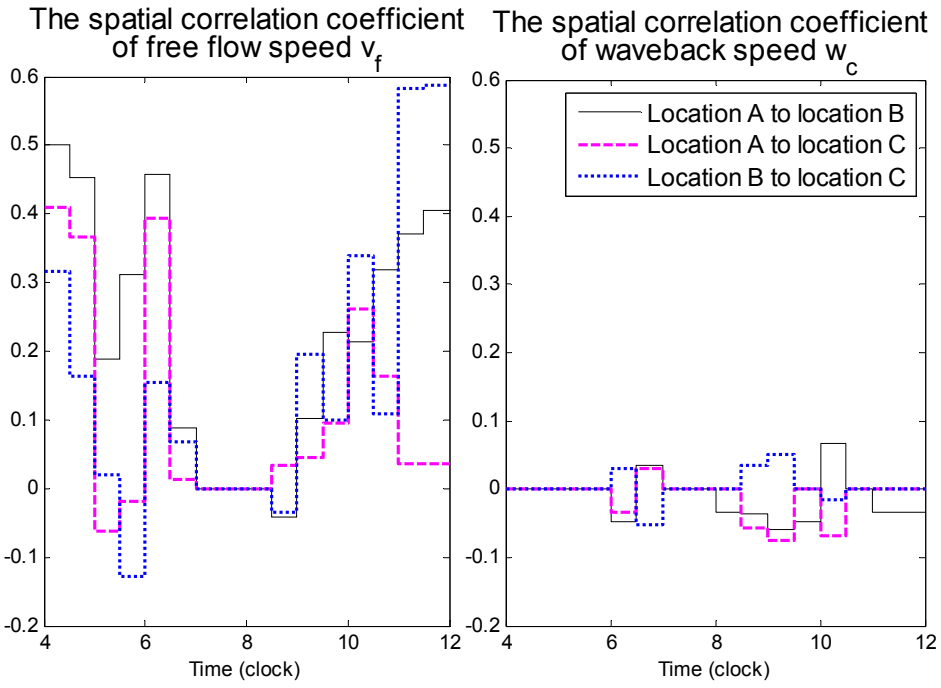
5.2 Temporal correlation and best linear predictor

5.2.1 Best linear predictor

As mentioned in Chapter 3, the accuracy of the prediction heavily depends on the similarity between the historical data and the current trend of the traffic flow when we apply the SCTM to predict traffic state. The prediction may fail when the traffic network is suffering from traffic incident or abnormal weather conditions (see, e.g. Sumalee *et al.* (2010b) and the empirical study in Section 5.5.1). The prediction accuracy would increase by incorporating the spatial and temporal correlations of traffic flow into the SCTM. In order to extend the SCTM for traffic state prediction, we may adopt certain prediction algorithm to forecast the inflow/outflow profiles and supply functions. In this section, the Multivariable Normal Distribution (MND) based best linear predictor (Tong, 1990) is utilized to accomplish this objective.



a). Spatial correlation of free-flow speed



b). Spatial correlation coefficients of free-flow speed and wave-back speed

Figure 5.3. Spatial correlation phenomena of supply side

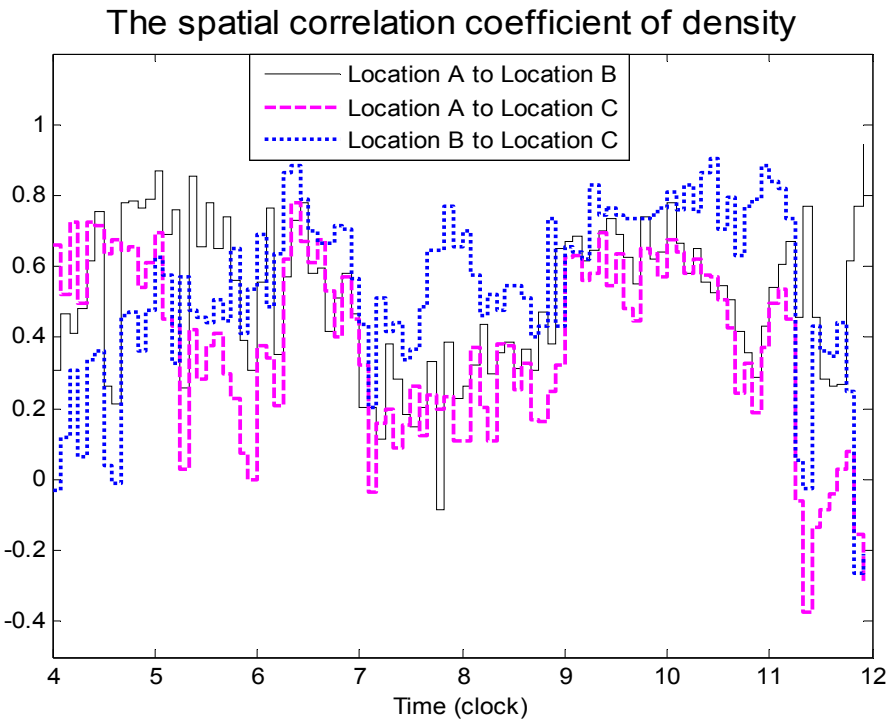


Figure 5.4. Spatial correlation coefficients of densities

Take the prediction of flow profile at a given location as an example; we assume that the dynamic statistical flow profiles are normally distributed. The flow profile vector conditioned on the current measurement can be forecasted by the following best linear predictor:

$$E(\mathbf{q}_f \mid \mathbf{q}_m) = E\left(\widehat{\mathbf{q}}_f\right) + \mathbf{s}_{\widehat{q}_{f,m}} \sum\nolimits_{\widehat{\mathbf{q}}_m}^{-1} \left(\mathbf{q}_m - E\left(\widehat{\mathbf{q}}_m\right)\right) \tag{5.9}$$

where $\mathbf{q}_f = col\left(q\left(k+1\right),\cdots,q\left(k+n\right)\right)$ denotes the extended flow vector to be forecasted for the coming predicting horizon, $\widehat{\mathbf{q}}_f$ is the corresponding historical data during the same period with $E\left(\widehat{\mathbf{q}}_f\right)$ denotes the mean value of $\widehat{\mathbf{q}}_f$, while $\mathbf{q}_m = col\left(q\left((k-l+1)\right),\cdots,q\left(k\right)\right)$ is the measured flow vector for current time period. $\mathbf{s}_{\widehat{q}_{f,m}} = \left(s_{i,j}\right)_{n\times l}$ denotes the covariance matrix between $\mathbf{q}_f$ and $\mathbf{q}_m$, where

$$s_{i,j} = \mathrm{cov}(\widehat{q}_f(i), \widehat{q}_m(j)) = \mathrm{cov}\left(\widehat{q}(k+i), \widehat{q}(k-l+j)\right),\ i = 1,\cdots,n;\ \ j = 1,\cdots,l$$

and $\sum\nolimits_{\widehat{\mathbf{q}}_m}^{-1}$ is the inverse of the covariance matrix of $\widehat{\mathbf{q}}_m$ with $\sum\nolimits_{\widehat{\mathbf{q}}_m} = \left(\sigma_{i,j}\right)_{l\times l}$, and

$$\sigma_{i,j} = \mathrm{cov}(\widehat{q}_m(i), \widehat{q}_m(j)) = \mathrm{cov}\left(\widehat{q}\left((k-l+i)T_p\right), \widehat{q}\left((k-l+j)T_p\right)\right),\quad i,j = 1,\cdots,l\,.$$

The conditional variance matrix of the flow profile can be evaluated as:

$$DX\left(\mathbf{q}_f \mid \mathbf{q}_m\right) = DX(\hat{q}_f) - \mathbf{s}_{\hat{q}_{f,m}} \sum\nolimits_{\hat{\mathbf{q}}_m}^{-1} \mathbf{s}_{\hat{q}_{f,m}}^T \tag{5.10}$$

where $DX\left(\mathbf{q}_f \mid \mathbf{q}_m\right) = \begin{bmatrix} DX_{q(k+1)} & & \cdots & 0 \\ & DX_{q(k+2)} & & \\ \vdots & & \ddots & \\ 0 & \cdots & & DX_{q(k+n)} \end{bmatrix}_{n\times n}$ and

$$DX\left(\hat{\mathbf{q}}_f\right) = \begin{bmatrix} DX_{\hat{q}(k+1)} & & \cdots & 0 \\ & DX_{\hat{q}(k+2)} & & \\ \vdots & & \ddots & \\ 0 & \cdots & & DX_{\hat{q}(k+n)} \end{bmatrix}_{n\times n}$$

where $DX_{q(k+i)} = \mathrm{var}\left(q(k+i)\right)$.

The mechanism of the best linear predictor is depicted in Figure 5.5. We utilize the difference between the historical mean and current measurement of the flow profile (we will refer to this as "error") to correct the prediction. Ideally, if this error is zero, the predicted flow profile "equals to" the historical flow profile, or no adjustment is made. Otherwise, we adjust the prediction by the error weighted by the covariance matrices (or their inverses) if the error is not zero.

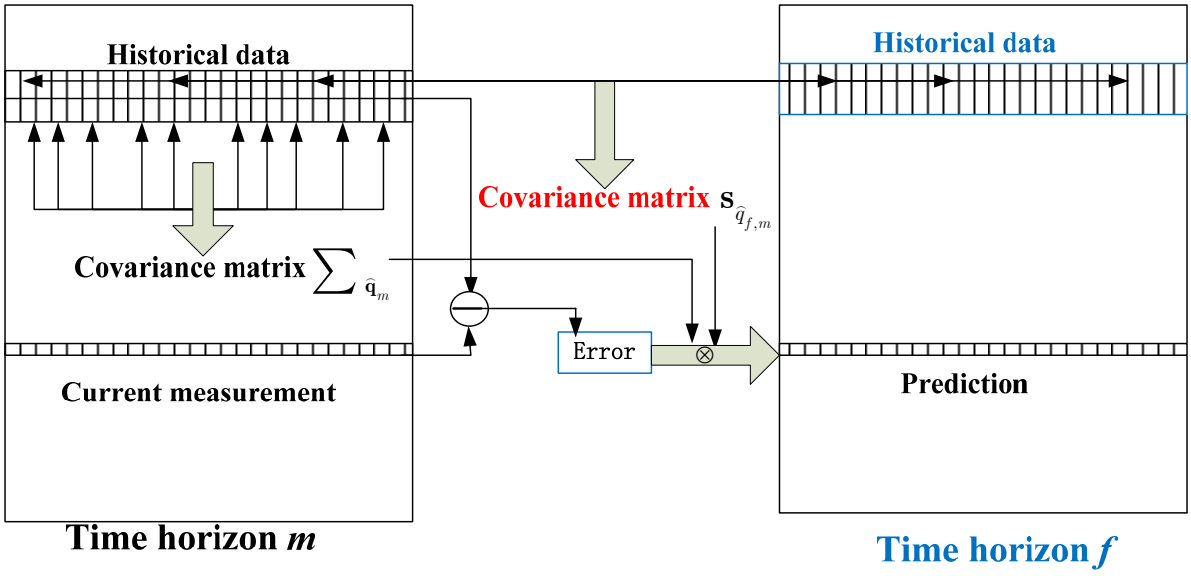


Figure 5.5. Illustration of the mechanism of the best linear predictor

The prediction of supply functions conditioned on the current observation can also be predicted by the same algorithm:

$$E(\delta_f \mid \delta_m) = E\left(\widehat{\delta}_f\right) + \mathbf{s}_{\widehat{\delta}_{f,m}} \sum\nolimits_{\widehat{\mathbf{q}}_m}^{-1} \left(\delta_m - E\left(\widehat{\delta}_m\right)\right) \tag{5.11}$$

where $E(\delta_f \mid \delta_m)$ is the mean value of parameter δ_f given the real-time calibrated δ_m, $\mathbf{s}_{\widehat{\delta}_{f,m}}$ and $\sum_{\widehat{\mathbf{q}}_m}$ are variance vector or matrix respectively.

The best linear predictor can forecast the flow profiles and supply functions by considering their historical statistics and the real-time detection. This method avoids the time consuming approaches such as data mining in database (e.g. *k-NN* algorithm). Thus, it is efficient. Adopting the time-dependent supply functions instead of the fixed parameters renders the prediction of traffic state more reliable. The accuracy of predictor may be influenced by the following factors: the size of $\mathbf{q}_f$ and $\mathbf{q}_m$ (prediction and measurement horizons), the degree of corresponding correlations (i.e. the weighting matrices), and the characteristic of the traffic flow pattern. This issue will be discussed in the empirical study.

5.2.2 Inflow demand and available outflow capacity

Equation (5.9) introduces the prediction algorithm for dynamic short-term forecasting of stochastic inflow, outflow profiles and supply functions for a freeway segment. However, the emphasis of prediction will be different under different traffic conditions. Referring to the short-term prediction of demand profile of a freeway segment, e.g. $u(k) = \left[q_u(k), r_2(k), f_3(k), q_d(k)\right]^T$ introduced in Section 3.4 (i.e. the inflow profile, on-ramp flow, off-ramp flow and outflow profile of the segment). These profiles may not be required simultaneously in prediction. For example, in *Cx* mode (such as *CC* and *CF* mode), $q_u(k)$ is not necessary to be predicted as the inflow to the segment is directly determined by the available capacity of the first cell, i.e.

$q_{u,Cx}(k) = w_{c,1}\left(\rho_{J,1} - \rho_1(k)\right)$. However, for *Fx* mode (such as *FC and FF* mode), $q_u(k)$ must be predicted by utilizing the predictor. Similarly, in *xF* mode (such as *FF* and *CF* mode), $q_d(k)$ is not necessary to be provided by the predictor as the outflow from the segment is determined by the density of the last cell, i.e. $q_{d,xF}(k) = v_{f,z}\rho_z(k)$. However in *xC* mode (*CC and FC* mode), $q_d(k)$ must be forecasted by the predictor by utilizing the temporal correlation of the outflow and the methodology introduced in Section 5.2.1.

5.3 Extension of the SCTM

The consideration of spatial-temporal correlations affects two major components of the SCTM, i.e. the evaluations of probabilities of occurrence of different modes and the traffic flow propagation as mentioned in Section 3.2 and 3.3. In the original SCTM, uncorrelated assumption is enforced to simplify both the probability evaluation and traffic flow propagation, e.g. Equations (3.1) and (3.4) define the probabilities of four basic modes, *FF, CC, CF* and *FC*, while Equation (3.7) simplifies them based on the un-correlation assumption. In fact, as we have illustrated in Section 5.1, both the cell densities and critical densities are spatially correlated. This implies that we should evaluate this joint probability rather than regarding them as independent events, or $\Pr(A \cap B) \neq \Pr(A) \cdot \Pr(B)$.

We here investigate a scenario introduced in Section 3.2.1. To begin with, we define X_{ud} as:

$$\begin{aligned} X_{ud} &= \left[X_u, X_d\right]^T \\ &= \left[\tilde{\rho}_u(k) - \rho_{c,1}(k), \tilde{\rho}_d(k) - \rho_{c,z}(k)\right]^T. \end{aligned}$$

Then the probability density function of X_{ud} is a bivariate normal distribution:

$$pdf(X_{ud}; \mu_{ud}, \Sigma_{ud}) = \frac{1}{\left(2\pi\right)\left|\Sigma_{ud}\right|^{1/2}} e^{-Q(X_{ud};\mu_{ud},\Sigma_{ud})/2} \tag{5.12}$$

where $\mu_{ud} = (\mu_u, \mu_d)^T = \left(\overline{\rho}_u(k) - \overline{\rho}_{c,1}(k), \overline{\rho}_d(k) - \overline{\rho}_{c,z}(k)\right)^T$ is the expectation of X_{ud}, and Σ_{ud} is the covariance matrix, with

$$\sum\nolimits_{ud} = \begin{bmatrix} Var_{Xu} & Cov_{Xu,Xd} \\ Cov_{Xu,Xd} & Var_{Xd} \end{bmatrix}$$

and $Q\left(X_{ud}; \mu_{ud}, \Sigma_{ud}\right) = \left(X_{ud} - \mu_{ud}\right)^T \sum\nolimits_{ud}^{-1} \left(X_{ud} - \mu_{ud}\right)$.

Then Equations (3.1) can be evaluated as:

$$\begin{aligned} P_{FF}(k) &= \Pr\left(\tilde{\rho}_u(k) < \rho_{c,1} \cap \tilde{\rho}_d(k) < \rho_{c,z}\right) \\ &= \Pr\left(X_u < 0 \cap X_d < 0\right) \\ &= \int_{-\infty}^{0}\int_{-\infty}^{0} pdf(X_{ud}; \mu_{ud}, \Sigma_{ud}) d_{Xu} d_{Xd} \end{aligned} \tag{5.13}$$

$$P_{CC}(k) = \int_{0}^{+\infty}\int_{0}^{+\infty} pdf(X_{ud}; \mu_{ud}, \Sigma_{ud}) d_{Xu} d_{Xd}$$

$$P_{CF}(k) = \int_{0}^{+\infty}\int_{-\infty}^{0} pdf(X_{ud}; \mu_{ud}, \Sigma_{ud}) d_{Xu} d_{Xd}$$

$$P_{FC}(k) = 1 - (P_{FF}(k) + P_{CC}(k) + P_{CF}(k)).$$

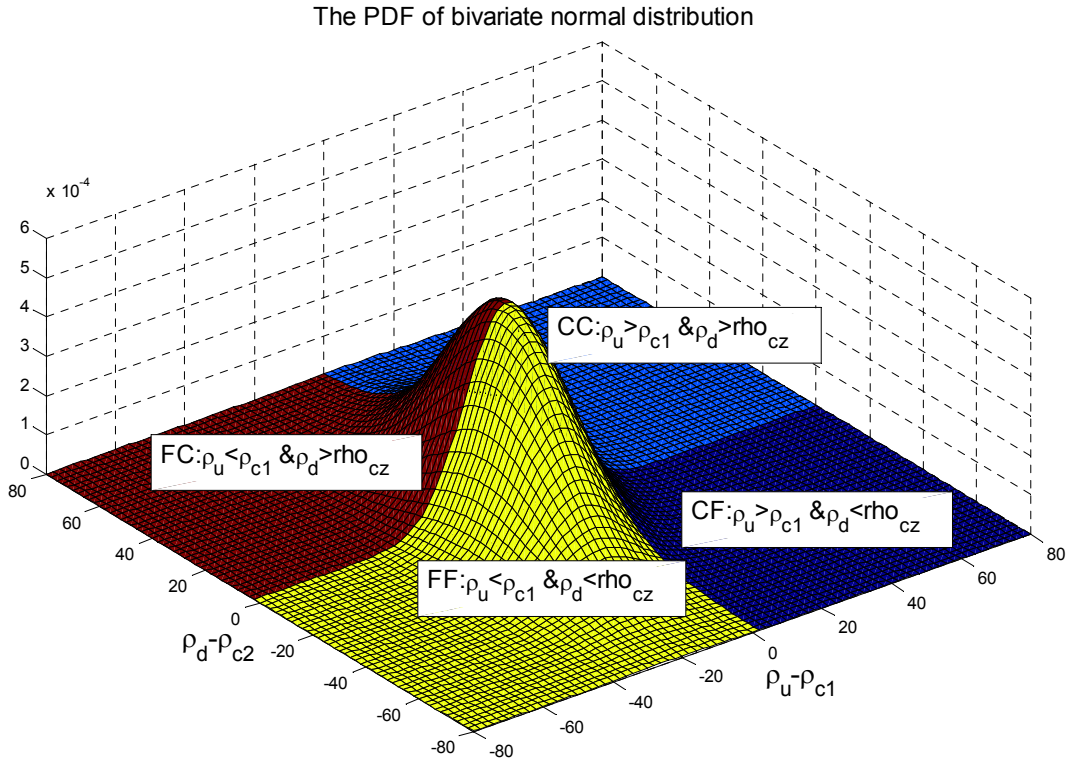


Figure 5.6. The PDF of bivariate normal distribution and probabilities of occurrence of different modes

A demonstration of the PDF of the bivariate normal distribution discussed above is given in Figure 5.6. The probabilities of occurrence of different modes can be evaluated by the area of the corresponding regimes, e.g. the probability of FF mode is given by the yellow cover. Similar method can be applied to evaluate the probabilities of occurrence of transient modes, i.e. $P_{CF,l}(k)$ $P_{FC1,l}(k)$ and $P_{FC2,l}(k)$. To evaluate $P_{CF,l}'(k)$, i.e. the probability of CF mode with the wave-front located at the boundary of cell *l*-1 and cell *l*, we define the following vector:

$$\begin{aligned} X &= \left[x_1, \cdots x_r, \cdots x_z\right]^T \\ &= \left[\tilde{\rho}_1(k) - \rho_{c,1}(k), \cdots \tilde{\rho}_r(k) - \rho_{c,r}(k), \cdots, \tilde{\rho}_z(k) - \rho_{c,z}(k)\right]^T, \quad r = 2, \cdots z-1. \end{aligned}$$

The PDF of *X* is assumed to be multivariate normal distribution:

$$pdf(X;\mu,\Sigma) = \frac{1}{\left(2\pi\right)^{z/2}\left|\Sigma\right|^{1/2}} e^{-Q_z(X;\mu,\Sigma)/2} \tag{5.14}$$

where $\mu = (\mu_1, \mu_2, \cdots \mu_z)^T = (\overline{\rho}_1(k) - \overline{\rho}_{c,1}(k), \cdots \overline{\rho}_r(k) - \overline{\rho}_{c,r}(k) \cdots, \overline{\rho}_z(k) - \overline{\rho}_{c,z}(k))^T$ is the expectation vector of *X*, and Σ is the covariance matrix, with

$$\sum = \begin{bmatrix} \mathrm{var}\left(x_1\right) & \mathrm{cov}\left(x_1, x_2\right) & \cdots & \mathrm{cov}\left(x_1, x_z\right) \\ \mathrm{cov}\left(x_2, x_1\right) & \mathrm{var}\left(x_2\right) & \cdots & \mathrm{cov}\left(x_2, x_z\right) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}\left(x_z, x_1\right) & \mathrm{cov}\left(x_z, x_2\right) & \cdots & \mathrm{var}\left(x_z\right) \end{bmatrix}_{z\times z}$$

where $1 \le i, j \le z$ and $Q_z\left(X;\mu,\Sigma\right) = \left(X-\mu\right)^T \sum^{-1}\left(X-\mu\right)$.

Then

$$\begin{aligned} P_{CF,l}'(k) &= \Pr\left(\tilde{\rho}_1(k) > \rho_{c,1} \cap \cdots \tilde{\rho}_{l-1}(k) > \rho_{c,l-1} \cap \tilde{\rho}_l(k) < \rho_{c,l} \cap \cdots \tilde{\rho}_z(k) < \rho_{c,z}\right) \\ &= \Pr\left(X_1 > 0 \cap \cdots X_{l-1} > 0 \cap X_l < 0 \cap \cdots X_z < 0\right) \\ &= \int_0^{+\infty} \cdots \int_0^{+\infty} \int_{-\infty}^{0} \cdots \int_{-\infty}^{0} pdf(X;\mu,\Sigma) dx_1 \cdots dx_{l-1} dx_l \cdots dx_z \\ &\quad for\ l = 2, \cdots z \end{aligned} \tag{5.15}$$

However the summation of probabilieis $P_{CF,l}'(k)$ for $l = 2, \cdots z$ may not be equal to $P_{CF}(k)$, then the weighting method is applied to overcome this problem:

$$P_{CF,l}(k) = \frac{P_{CF,l}'(k)}{\sum_{lx=2}^{z} P_{CF,lx}'(k)} P_{CF}(k) \tag{5.16}$$

$P_{FC1,l}(k)$ and $P_{FC2,l}(k)$ can be evaluated similar to (3.13) and (5.14)-(5.16).

Evaluation of the autocorrelation matrix of the bilinear system should be adjusted accordingly. Referring to equation (3.9)-(3.12), the following adjustment is made:

$$\chi_{s,i,j} = E\left(\omega_{s,i}(k)\omega_{s,j}(k)\right) - E\left(\omega_{s,i}(k)\right)E\left(\omega_{s,j}(k)\right) = \mathrm{cov}\left(\omega_{s,i}(k), \omega_{s,j}(k)\right).$$

Section 5.3 extends Step 2 and Step 3 in Figure 3.1(b) to consider spatial correlations of stochastic parameters of the fundamental diagram and stochastic densities on adjacent cells.

5.4 The framework of online traffic state and journey time prediction

The flow chart of online journey time prediction is depicted in Figure 5.7(a). The online prediction framework consists of two major parts as described in the previous sections, i.e. the SCTM considering temporal and spatial correlations and journey time evaluation based on the predicted traffic densities via the model. The forecasting framework utilizes the rolling horizon approach which is a concept widely used in online predictions. In this rolling horizon approach, traffic density and journey time are predicted cycle by cycle with pre-defined prediction horizon[1]. Every two neighboring horizons are differentiated by re-predicting horizon which is generally shorter than predicting horizon as illustrated in Figure 5.7(b) .

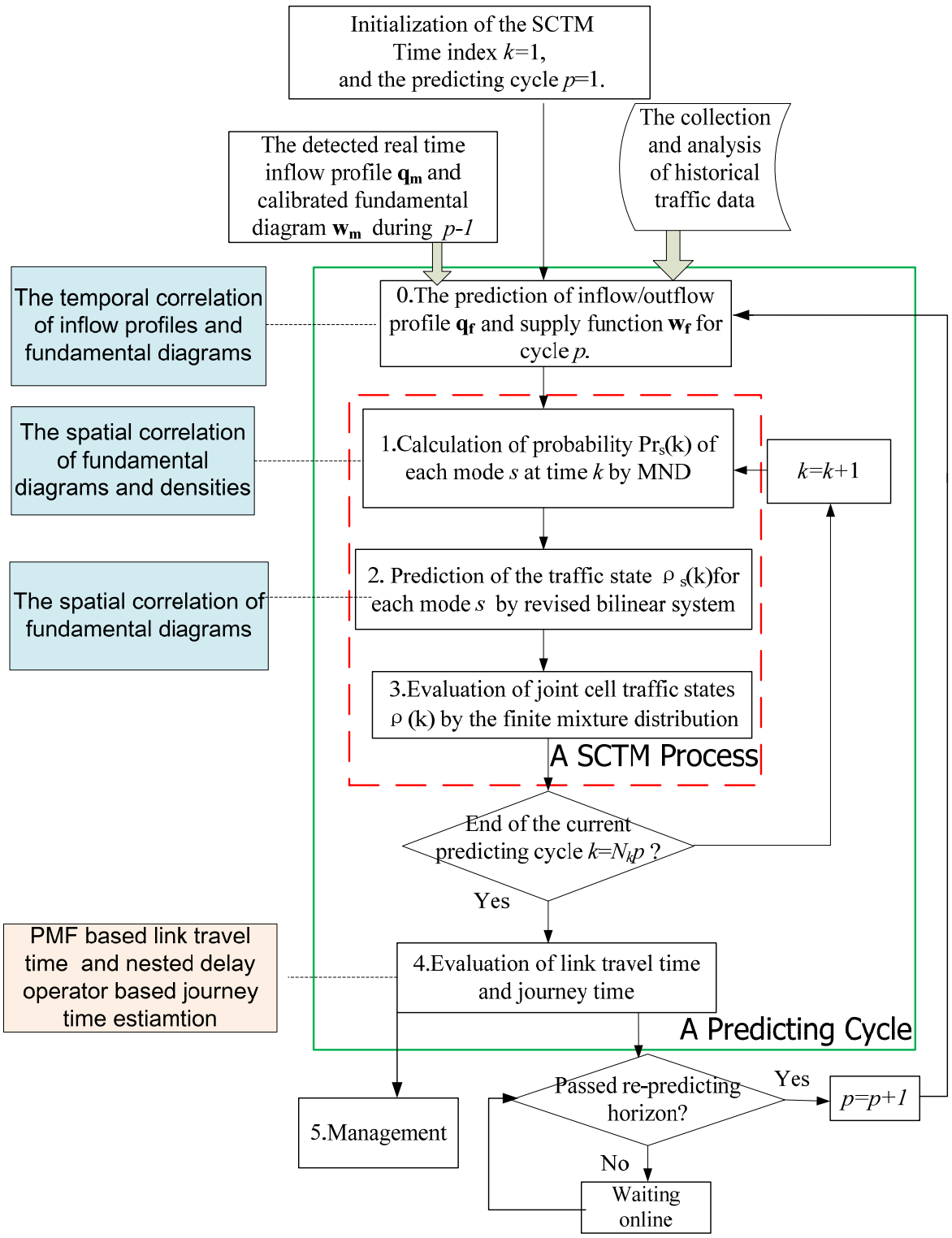
The rolling horizon approach can be interpreted as, for example the route guidance in

[1] This is also known as rolling horizon length which is the time period for which prediction takes place. This length is a function of the maximum trip length. A rolling horizon is usually divided into short time intervals, e.g. rolling horizon step size which specifies how often guidance is renewed.

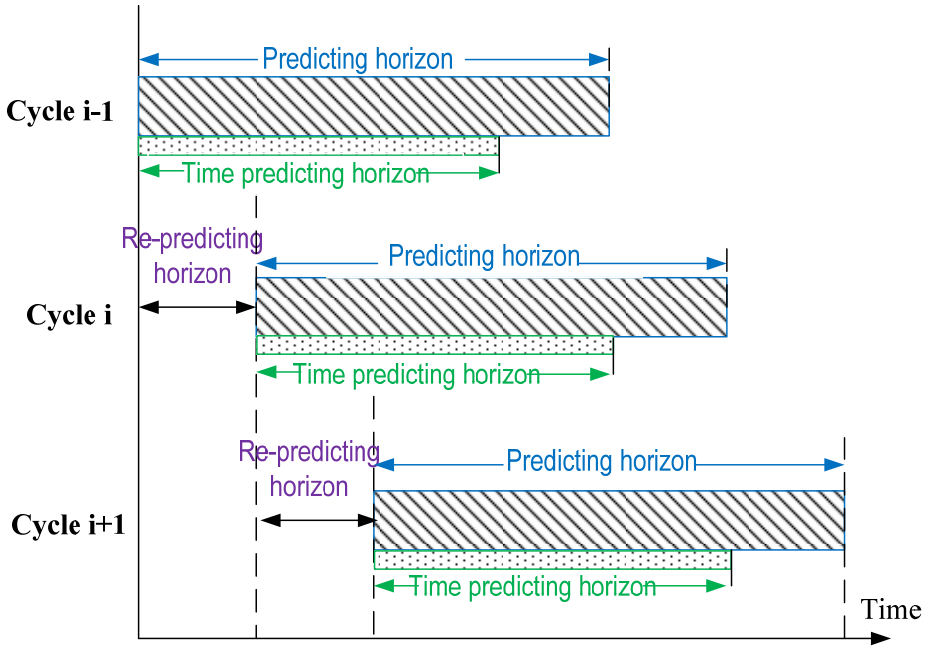
an advanced traveler information system (ATIS), under this model, the route guidance is periodically generated and evaluated for a given time horizon (or the prediction horizon, e.g. 45 minutes) based on the latest information available. However, 45 minutes may be too long for us to trust the prediction (or the prediction may not reflect the current traffic state). On the other hand, the measurement systems keep detecting the real-time traffic state, e.g. the PeMS can provide real-time traffic information every 30 seconds and aggregate traffic state every 5 minutes. We would like to use this information, e.g. the 5 minutes aggregate data, to adjust the prediction which is discussed in the previous sections. Therefore, the current guidance until the next guidance update (e.g. 5 minutes) is actually implemented. The travelers may update their route choices according to the guidance updates. We call this interval of 5 minutes as the rolling horizon step size[2].

Due to the inherent bad detected data, errors of the measurement systems, and communication delays, it may not possible for us to implement a traffic simulator in real-time. This may be overcome by using this rolling horizon approach as depicted in Figure 5.8. The measure data is not directly used by the simulator, e.g. the SCTM is the thesis. On the other hand, it is processed by a predictor/filter, e.g. data filtering (e.g. imputation of bad detections) and the prediction of demand and supply functions as discussed in the previous sections. The filtered data and predicted demand and supply functions are stored in a repository of road networks, which is also a database. The simulator fetches data directly from this repository to conduct real time traffic state estimation prediction.

[2] This step size may be determined based on level of variation in traffic conditions over time (e.g. a longer step size can be used if traffic conditions do not change very much), the resolution of the measurement system, timing of incidents, and available computational resources.



a) Framework of journey time prediction



b) Rolling horizon

Figure 5.7. Framework of the SCTM based traffic prediction with rolling horizon

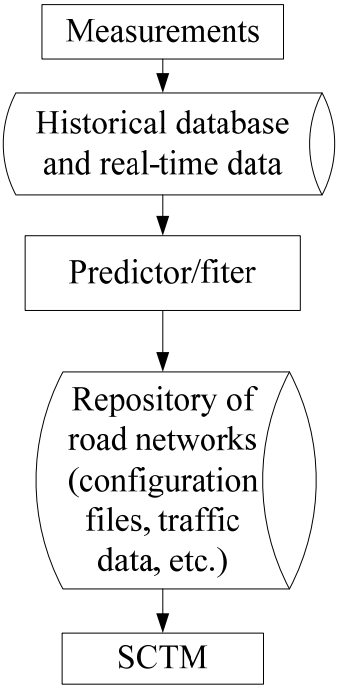


Figure 5.8. Data flow of a real-time traffic state prediction system

5.5 Empirical study 2: journey time prediction on a short segment of I210-W

This empirical study is tested on a 2-miles segment of Interstate 210 Freeway near Los Angeles. The traffic data collected on the morning of March 26, 2008 is selected for

the simulation due to the abnormal traffic conditions observed. The basic information of this segment and the traffic state on that day will be introduced briefly in Section 5.5.1. Section 5.5.2 and Section 5.5.3 compare the journey time predicting results by adopting the original SCTM introduced in Chapter 3 and extended prediction framework proposed in this chapter.

5.5.1 Description of the test site and simulating period

A segment of Interstate 210 West bound, approximately two miles in length, is chosen in this case study as depicted in Figure 5.9. This segment, located in Monrovia, Los Angeles, stretches from S Myrtle Ave (A) through W. Huntington Dr (B) to N Santa Anita Ave (C) with two on-ramps and two off-ramps. The section is instrumented with single-loop inductance detectors, which are embedded in the pavement along the mainline, HOV lane, on-ramps, and off-ramps. The historical data provided by the detectors are used to calibrate the random supply functions.

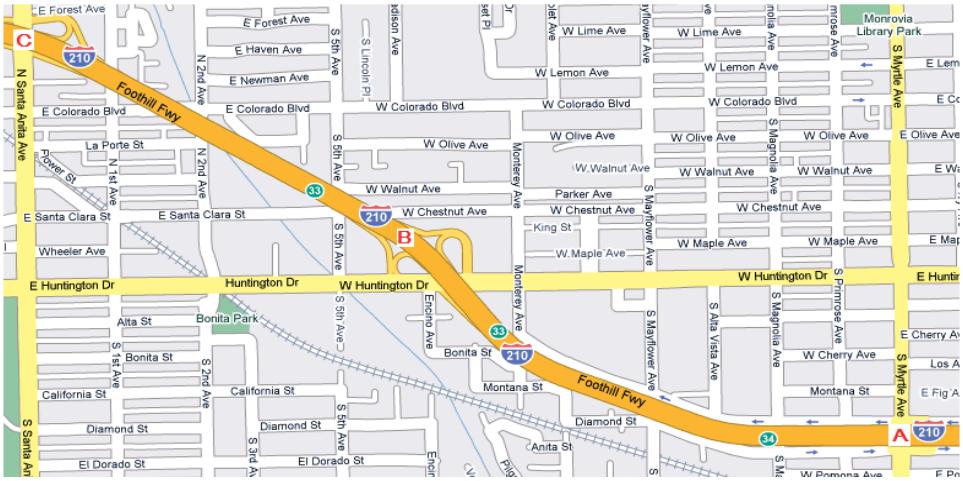


Figure 5.9. Location of the case study

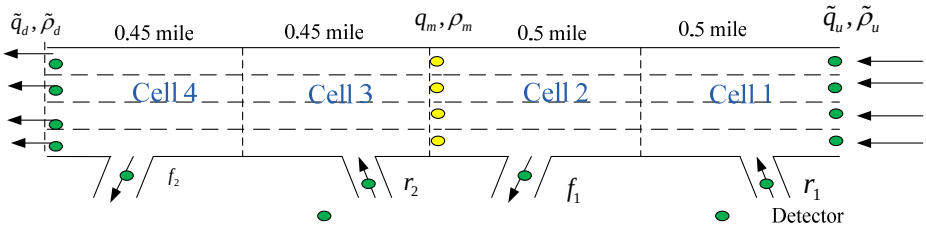


Figure 5.10. Assignment of cells and its detector configuration

Traffic flow data of seven hours (4:00 am-11:00 am) collected on Tuesday, Wednesday and Thursday of March and April 2008 and April 2009(see Table 5.1a) provided by the Performance Measurement System (PeMS) are used in this test. Each loop detector provides traffic volume (veh/time-step) and occupancy (%) of the corresponding lane for every 30 seconds. The densities can be computed for each lane by dividing the occupancy by the *g*-factor which is the effective vehicle length (in miles), provided by the PeMS database. Actually, the simulation prefers using the aggregated 5 minutes data because the 30 seconds raw database suffers from serious data missing problem. Considering the length of links and the accuracy of the travel time estimation, the simulation time step is set to be five seconds. Related time interval assignments are listed in Table 5.1b. Thus a *zero*th-order interpolation is applied to the PeMS data to generate the data with the time step of five seconds. The calibration of the stochastic triangular fundamental diagram is conducted for the four cells by using the historical data over the selected days. The results are shown in Table 5.2 and Figure 5.11. The notations with the hat symbol denote the mean values of the parameters, while σ with the mean notations as subscripts denote the corresponding standard deviations.

Calendar

Sun	Mon	Tue	Wed	Thur	Fri	Sat
						1/03/2008
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31	1/04/2008	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	1/05/2008		

Sun	Mon	Tue	Wed	Thur	Fri	Sat
1/03/2009	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31	1/04/2009	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30		

Selected weekdays

(a).Calendar of the selected weekdays

Items of time interval	Time interval
Raw detection	30 seconds
Data aggregation	5mins
Density estimating/predicting simulation	5 seconds
Travel time estimating/predicting simulation	5 seconds

(b).Time interval assignment of database and simulation

Table 5.1. Simulation settings

	$\hat{v}_f$	$\sigma_{\hat{v}_f}$	$\hat{w}_c$	$\sigma_{\hat{w}_c}$	$\hat{\rho}_c$	$\sigma_{\hat{\rho}_c}$	$\hat{\rho}_J$	$\sigma_{\hat{\rho}_J}$	$\hat{Q}_m$	$\sigma_{\hat{Q}_m}$
Cell 1	65.5	9.0	11.3	0.07	109	1.8	742.2	11.3	7154	623
Cell 2	63.9	6.4	11.9	0.09	113	2.1	725.8	12.5	7264	657
Cell 3	63.9	6.4	11.9	0.09	113	2.1	725.8	12.5	7264	657
Cell 4	65.4	6.4	12.7	0.09	108	2.0	671.0	11.4	7119	615

Table 5.2. Calibration results of the four cells

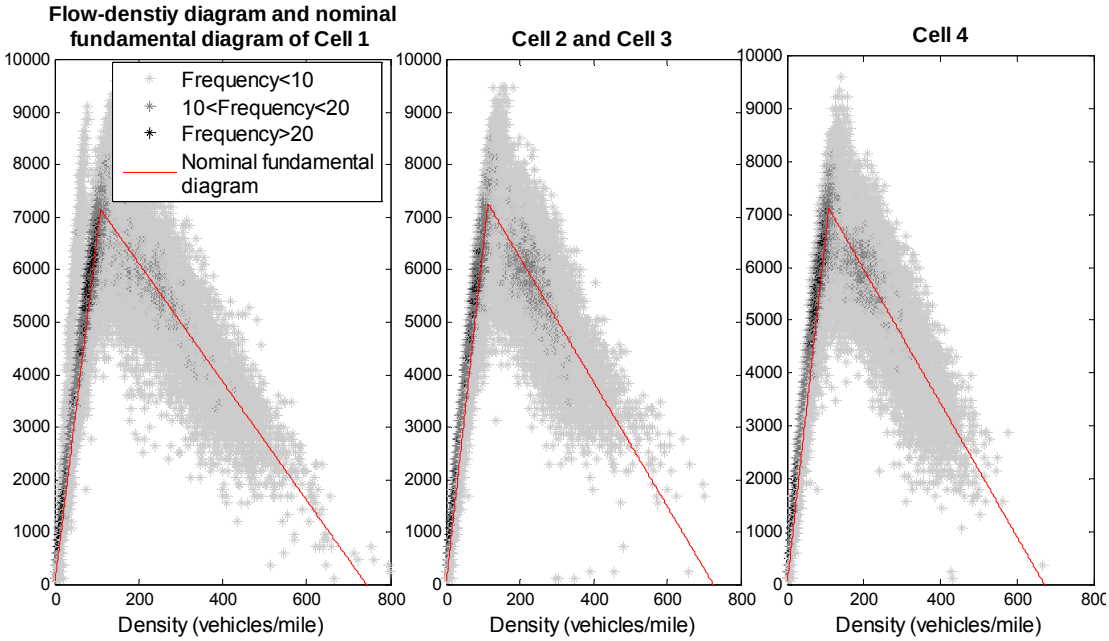


Figure 5.11. Fundamental diagrams of the four cells

On the morning of March 26, 2008, there was an abnormal traffic condition due to certain incident as illustrated in Figure 5.12. As we can observe that the detected inflow and outflow profiles of the day, compared with the distribution of the flows collected over the 54 days shown in Table 5.2a, admitted sudden declines at around 6:30 AM. Note also that the outflow of the segment admitted a decline before the inflow did, which implies this decline was due to the congestion spillback. Clearly, the flows of the tested day are quite different with the mean flows of the 54 sample days especially during the early morning time period 6:15-7:10 AM. If we just directly input the statistical flow profiles of the 54 sample days to a traffic simulator for

short-term prediction, the results, both traffic density and journey time, may be poor in terms of accuracy due to the great difference of the flow profiles.

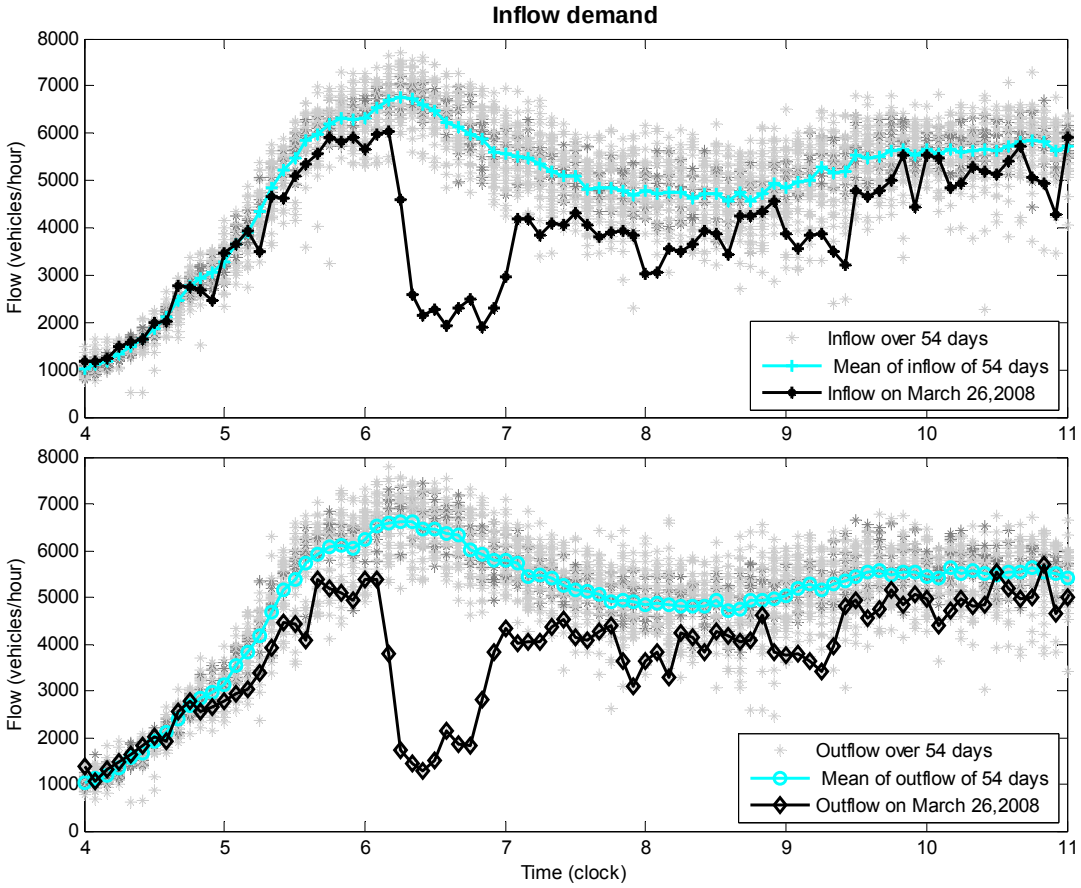


Figure 5.12. Inflow and outflow profile detected on March 26, 2008

5.5.2 Traffic density prediction on March 26, 2008—an inaccurate empirical study without considering correlations

In this part, the traffic state prediction between 5:00-11:00 AM on March 26, 2008 will be conducted every 10 minutes. In order to look into the disadvantages of the framework without considering the correlations, statistics of historical data, i.e. the detected traffic data of 54 sample days and their distributions as shown in Figure 5.12, will be applied for the prediction. The stochastic fundamental diagrams calibrated from historical data as listed in Table 5.2 will be adopted in the simulation.

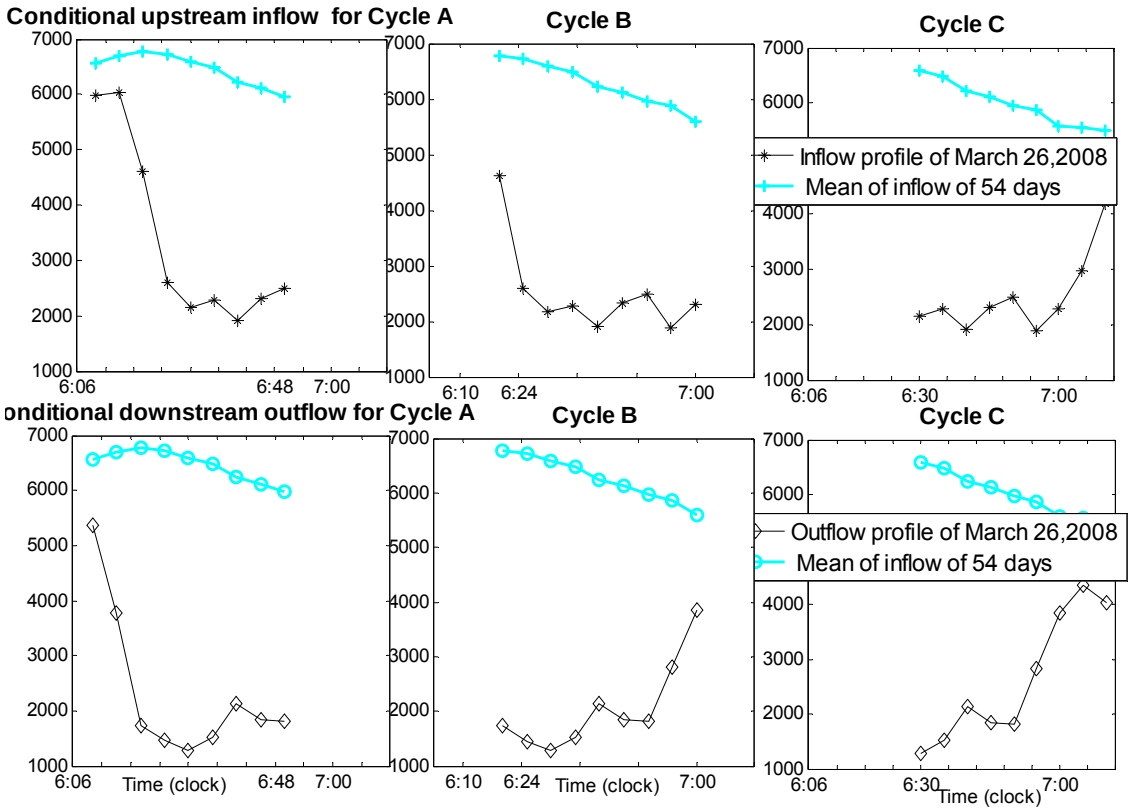
Table 5.3 lists the basic configurations of the simulation. Predicting horizon length is

assigned to be 40 minutes with 10 minutes margin for journey time prediction in order to ensure all the vehicles entering the link within the initial 30 minutes can exit from the link. Rolling horizon step size is 10 minutes which can be adjusted according to the accuracy requirement.

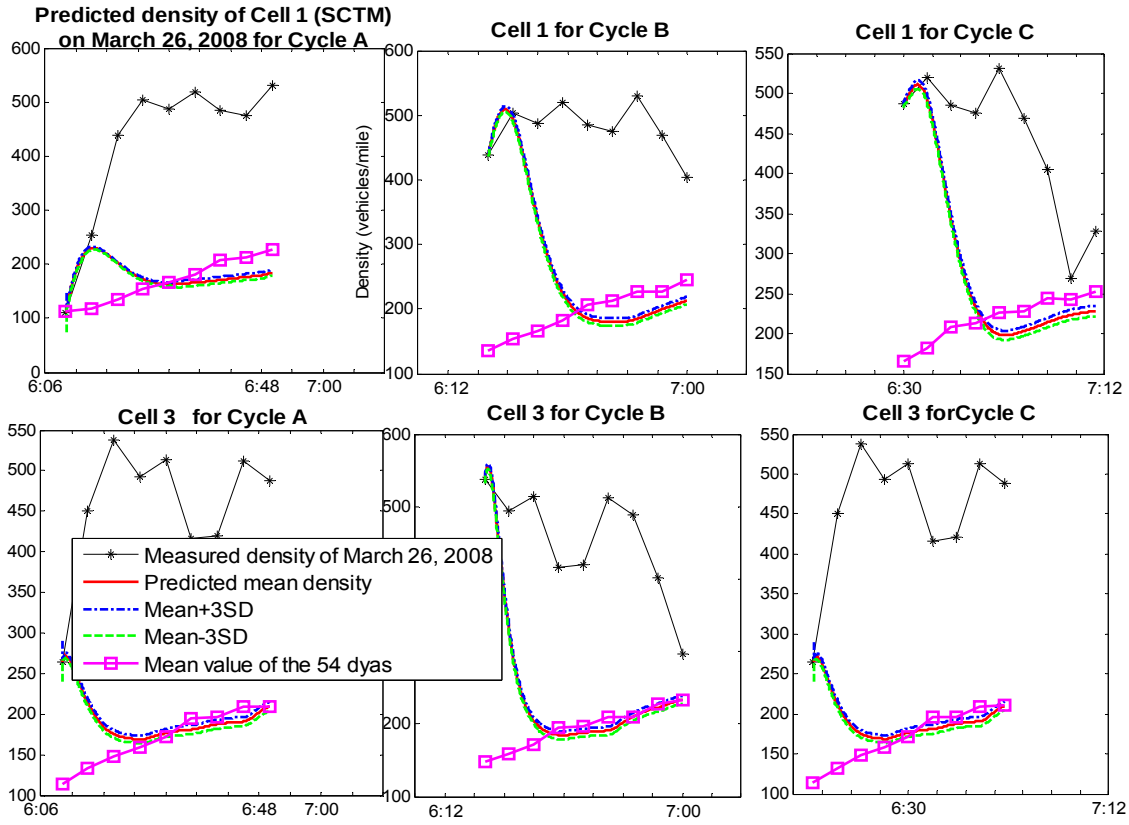
Items of time interval	Time
Raw detection time interval of PeMS	30 seconds
Aggregated time interval of PeMS	5mins
Density estimating/predicting simulation time interval	5 seconds
Predicting horizon	40 minutes
Time predicting horizon	30 minutes
Rolling horizon step size	10 minutes
Test date	March 26, 2008
Test time	5:00-11:00

Table 5.3. Simulation settings

Figure 5.13(a) depicts the statistical value of upstream inflow and downstream outflow (blue "+" lines) of three prediction cycles which begin at 6:10, 6:20 and 6:30 AM on March 26, 2008 compared with the actual detected flow profiles (black star lines), respectively. Figure 5.13(b) depicts the corresponding cell density distributions of cell 1 and cell 3, which are predicted for the three prediction cycles mentioned above, against the actual (black * lines) measurements and mean value of the historical data (pink square lines). Clearly, both of the inflow/outflow profiles and the predicted results are far away from those detected on March 26, 2008 because of the abnormal traffic condition on that day.



a) Conditional inflow and outflow profiles for three cycles



b) Predicted cell densities without considering correlations

Figure 5.13. Prediction results at 6:10, 6:20 and 6:30 on March 26, 2008

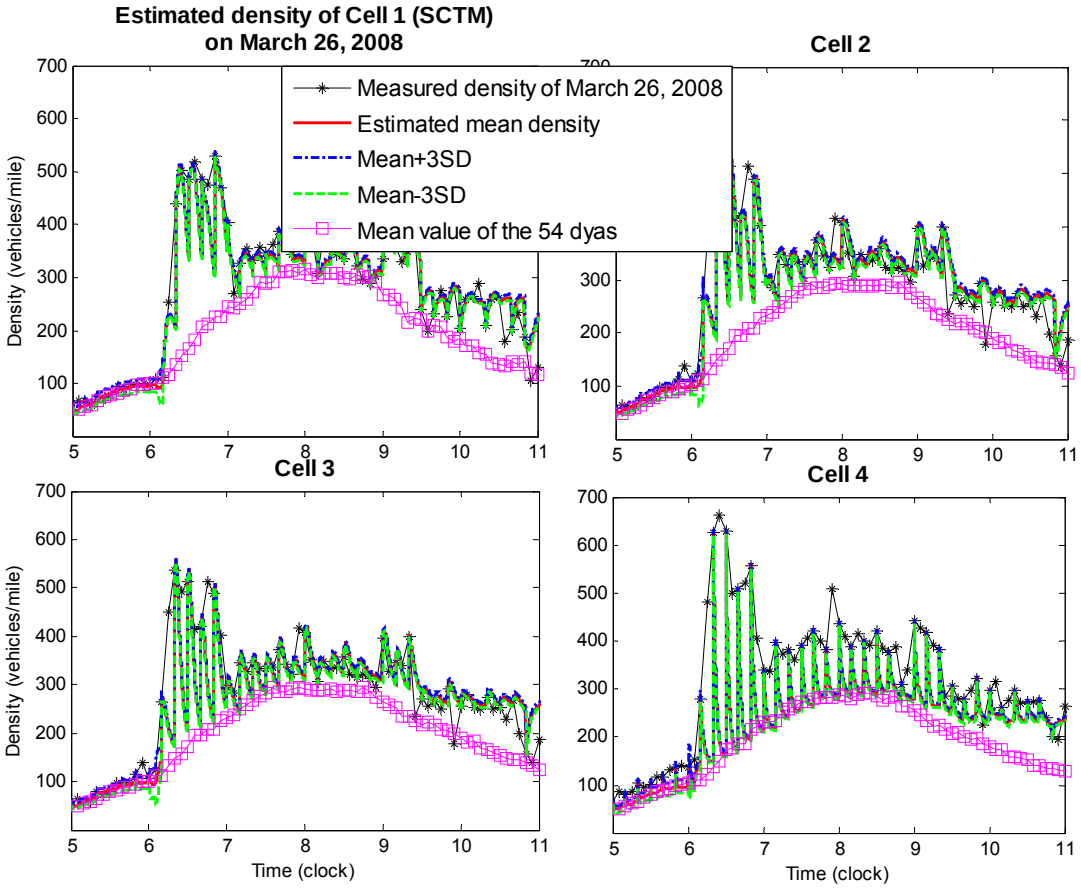


Figure 5.14. Prediction results of cell densities along the whole period

	Inflow	Outflow	Cell 1	Cell 2	Cell 3	Cell 4	Overall
MAPE(Cycle A)	142%	265%	58.9%	53.2%	54.3%	65.2%	57.9%
MAPE(Cycle B)	179%	241%	47.4%	45.8%	49.6%	63.2%	51.5%
MAPE(Cycle C)	148%	154%	37.6%	33.3%	37.2%	53.2%	40.3%
MAPE(Cycle A-C)	156%	220%	47.9%	44.1%	47.2%	60.5%	49.9%
MAPE (6:10-6:20 in Cycle A)	72%	238%	42.9%	27.1%	32.1%	48.3%	37.6%
MAPE (6:20-6:30 in Cycle B)	183%	364%	14.5%	18.1%	32.0%	60.5%	31.3%
MAPE (6:30-6:40 in Cycle C)	191%	256%	11.4%	10.5%	24.6%	54.7%	25.3%
MAPE (6:10-6:40)	149%	286%	22.9%	18.6%	29.6%	54.5%	31.4%

Table 5.4. MAPE of prediction results of three periods

Figure 5.14 provides intuitive results over the whole prediction period by merging the results of all the rolling horizon steps (every 10 minutes). Table 5.4 summarizes the

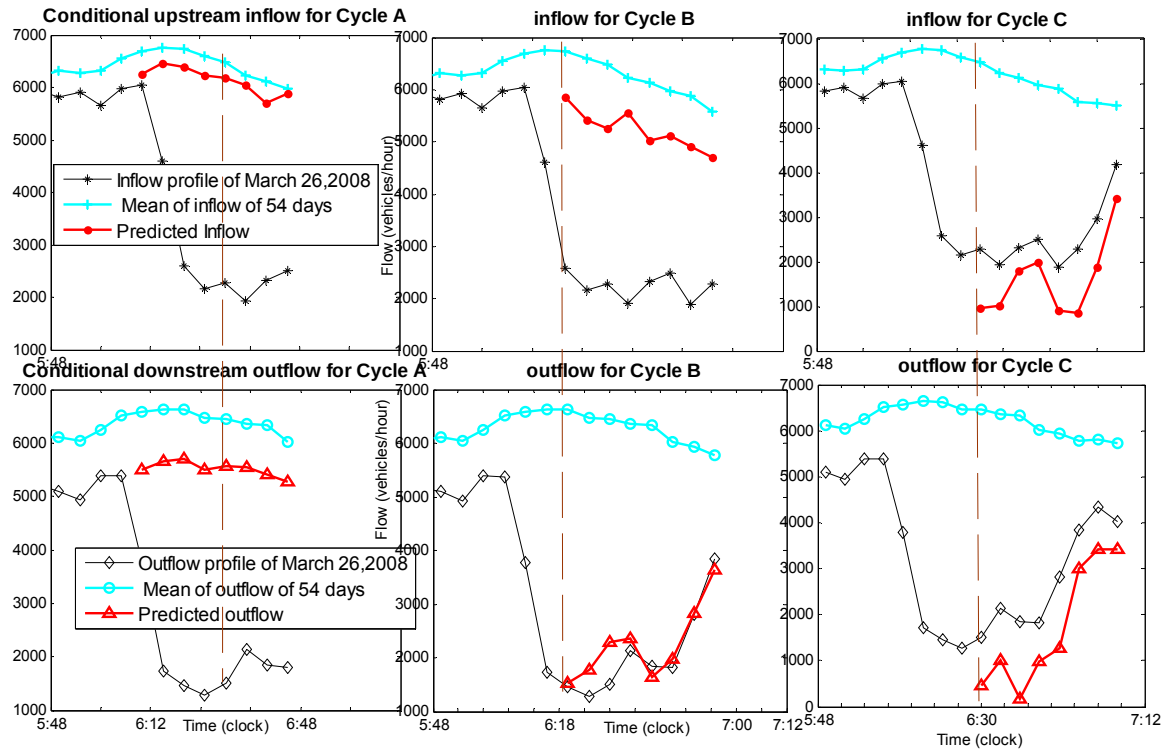
MAPE of the predicted densities during the three periods and the ones of the initial 10 minutes within the three cycles. Note that the MAPEs of predicted densities are smaller than those of flows due to the fact that the factor T_s/l <<1 in the SCTM dynamics, which transfers the flow into density, scales the amplitude. It can be concluded that the prediction is poor in terms of accuracy which reflects the disadvantages of the original SCTM for traffic state prediction under abnormal traffic conditions. It will be shown in the following section that this problem will be resolved by considering the temporal correlations of inflow and outflow profiles for prediction.

5.5.3 Traffic state prediction on March 26, 2008—an accurate empirical study by considering correlations

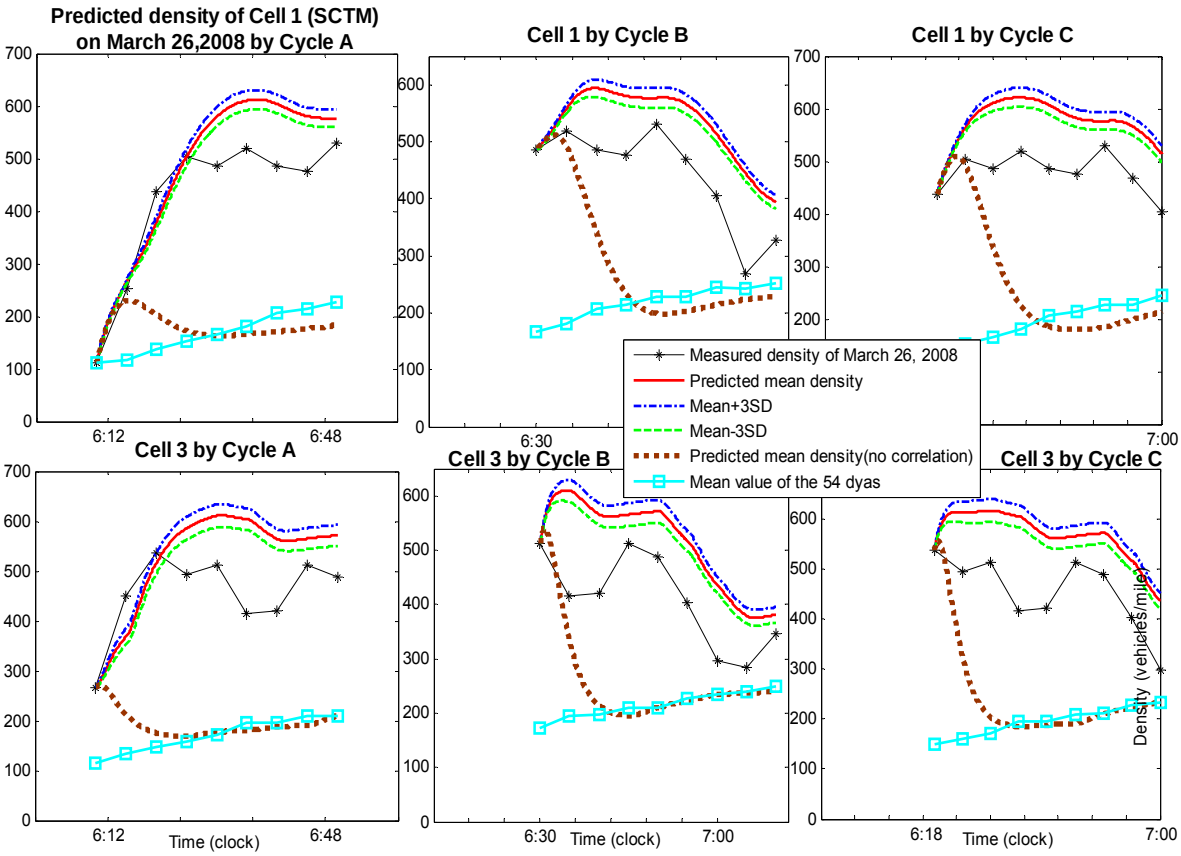
This section conducts an empirical study to validate the significant improvement of prediction results by introducing spatial and temporal correlations to the short-term traffic state forecasting framework as summarized in Figure 5.7(a). Figure 5.15(a) depicts the predicted inflow/outflow profiles for three rolling horizon steps by using multivariate normal distribution based best linear predictor introduced in Section 5.2. As we can see from the figure that in the first rolling horizon step, both inflow and outflow predictions are far away from the actual detections. This is because the abnormal traffic condition occurs after the prediction started at 6:10AM. The prediction cannot make use of the detected flow pattern until the next rolling horizon step, i.e. 10 minutes later. Compared with the prediction based on the original framework (brown dot line), the MAPE of density decreased from 52.4% to 18.0% as listed in Table5.5, and the impulse at the beginning of each rolling horizon step (due to initialization) does not exist anymore. The huge errors between the historical flow pattern and the real-time measurement are indentified by the predictor to adjust the prediction for next rolling horizon step. As reflected by the figure that, the predicted outflow profile meets the detected one in Cycle B. However the decrease of predicted inflow is still unsatisfactory as the decrease at 6:15 is not convincing enough. Figure 5.15(b) plots the predicted densities of cell 1 and cell 3 for the related cycles

accordingly. Compared with Figure 5.13(b), it is clear that the new prediction framework over performs the original SCTM. In Cycle B, the traffic state prediction over-estimates the traffic density but with the MAPE improved from 50.9% (by the original SCTM) to 17.2%. The prediction will keep on adjusting if the error between the historical mean and the current measurement is not zero. Cycle C (6:30-7:10) demonstrates this adjustment by the multivariate normal distribution based best linear operator considering temporal correlations. The level of improvement depends on the relationship between the characteristics of traffic incident and the rolling horizon step size. Of course, we may also use smaller rolling horizon step size to accelerate the rate that the prediction adapts to the abnormal traffic conditions, e.g. from 10 minutes to 2 minutes (with a minimum determined by the resolution of the measurement system, e.g. 30 seconds in PeMS system).

The fundamental diagram has no significant change due to normal weather condition and the fact that the incident does not occur on the underlying segment (as we have mentioned the change of the flow pattern is due to congestion spillback). As illustrated by Figure 5.16, the free flow speed (dash dotted black line) which is calibrated based on the flow-density pairs (blue dots) detected between 5:00-5:30 on March 26 is slightly lower than the historical record (red line). The predicted free flow speed which will be applied for the predicting cycle 5:30-6:00 might also decrease accordingly (green dot line) based on the positive correlation of free flow speed between these two time periods (see the spatial correlation measurement in Figure 5.2). As we have mentioned in Section 5.1, the temporal correlations cannot be calculated for all parameters of the fundamental diagram along the whole period due to the fact that they cannot be calibrated for the whole day, e.g. the congestion wave speed cannot be observed during 4:00-5:00AM. The parameters in Table 5.2 will be applied directly when the parameter is no predictable.



(a).Predicted conditional inflow and outflow profiles for three cycles



(b).Predicted cell densities for three cycles considering correlations

Figure 5.15. Prediction results of cell densities on 6:10, 6:20 and 6:30 of March 26, 2008

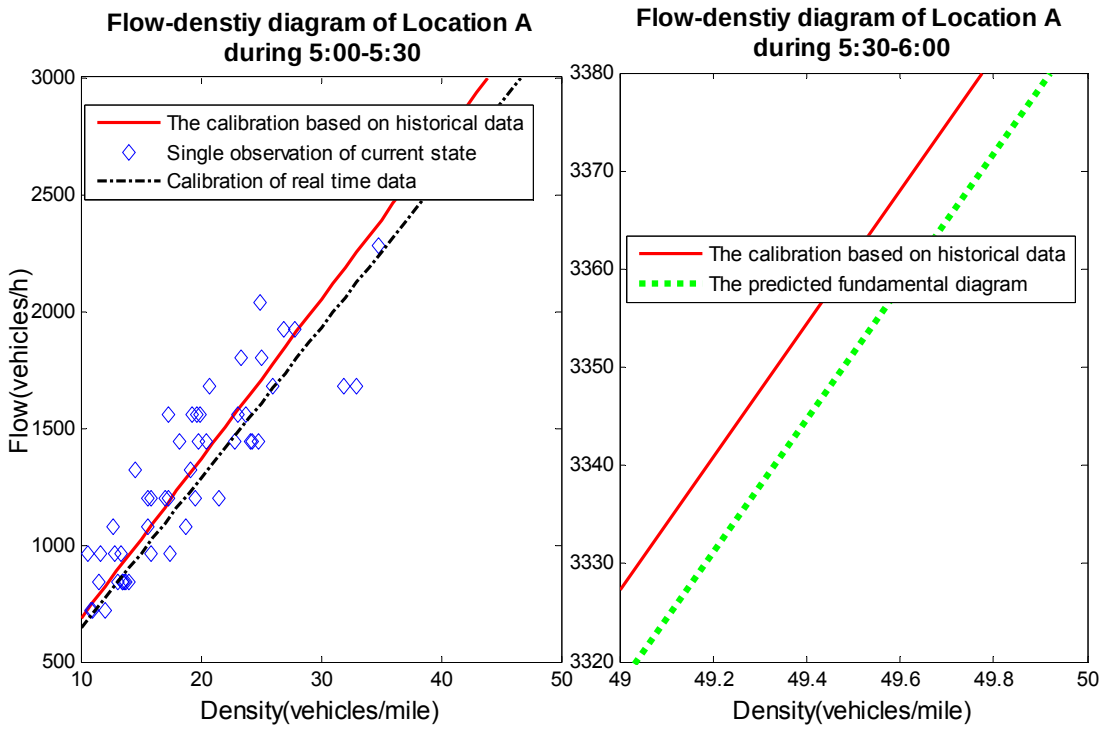


Figure 5.16. The prediction of fundamental diagram for cycle 5:30-6:00

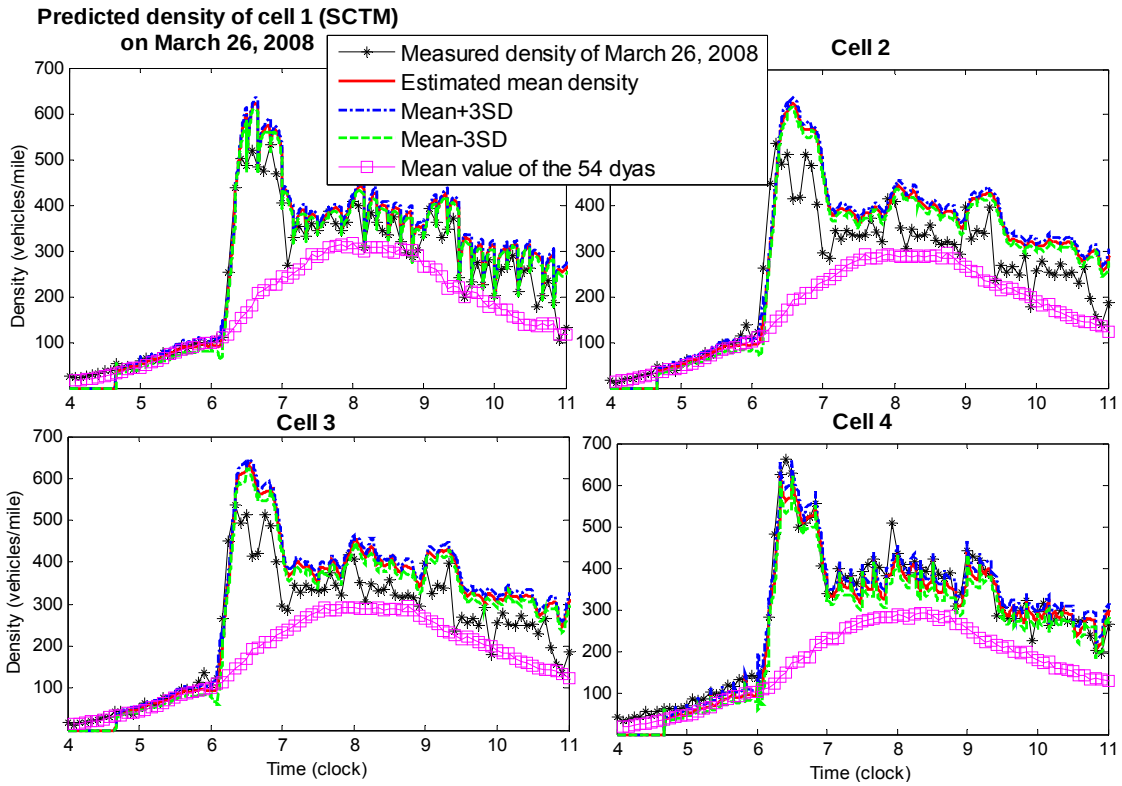


Figure 5.17 Prediction results of cell densities along the whole period

	Inflow	Outflow	Cell 1	Cell 2	Cell 3	Cell 4	Overall
MAPE(Cycle A)	130.1%	213.4%	25.1%	19.3%	19.0%	8.5%	18.0%
MAPE(Cycle B)	135.7%	16.6%	18.4%	22.5%	21.2%	6.8%	17.2%
MAPE(Cycle C)	39.7%	46.5%	19.7%	27.8%	25.0%	4.4%	19.2%
MAPE(Cycle A-C)	102.1%	92.2%	21.1%	23.2%	21.8%	6.6%	18.2%
MAPE (6:10-6:20 in Cycle A)	63%	188%	36.2%	10.1%	8.7%	11.8%	16.7%
MAPE (6:20-6:30 in Cycle B)	136%	32%	20.3%	16.0%	18.3%	10.7%	16.3%
MAPE (6:30-6:40 in Cycle C)	43%	71%	13.3%	28.5%	30.0%	4.7%	19.3%
MAPE (6:10-6:40)	80%	97%	23.3%	18.2%	19.0%	9.1%	17.4%

Table 5.5. MAPE of Predicting results for three periods

Figure 5.17 depicts the traffic density prediction of each rolling horizon step. Compared with the prediction by the original SCTM (the gallery dots), the new framework significantly improves the predictions of cell densities. For this special weekday with long lasting traffic incident during the morning rush hour, the MAPE of predicted densities of thirty six 40-minutes predicting horizons over 6 hours equals to 15.9%.

5.5.4 Journey time prediction on March 26, 2008

Dynamic journey time distributions predicted during the three cycles with corresponding Buffer Time Indexes are presented in Figure 5.18. Each time predicting horizon is 10 minutes shorter than their density predicting horizons as we have mentioned in previous sections. The predicted journey sharply rises from about 100 seconds to 600 seconds from about 6:00, and keeps on oscillating at high value until 9:20 AM around.

The dynamic stochastic journey time is reliable most of the time except for the period suffering from the spillback effect of traffic jam; the index reaches its highest value around 6:30 which means the driver should assign 60% of mean travel time as extra time in order to assure she/he can arrival on-time (or have less than 5% chance to arrive late).

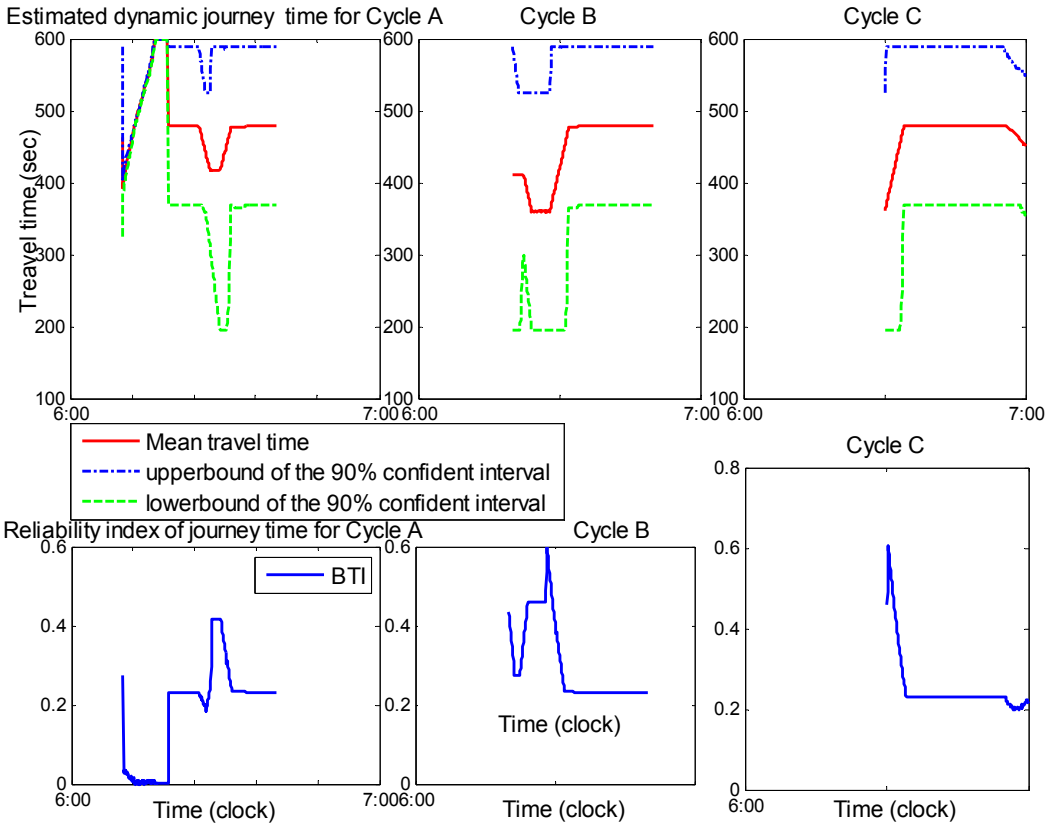


Figure 5.18. Predicted journey times on three cycles

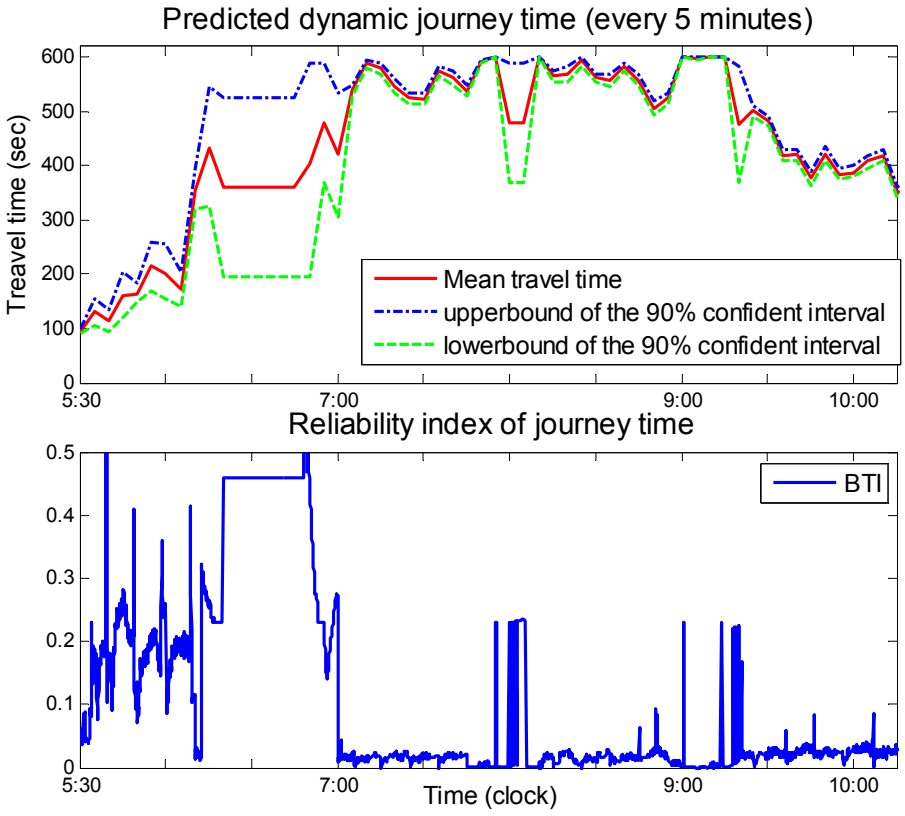


Figure 5.19. Summary of journey time

5.5.5 Discussion of computational time and parameter sensitivity

Table 5.6 gives the computational times of related items in one predicting cycle of Empirical 2-2. The study is also simulated on Matlab 2009(b), CPU: Intel Core 2 Duo E8500 @3.16GHz and 4G DDR3 1067MHz, which is the same configuration with Empirical 1(Journey time estimation on Hanshin Expressway). The table indicates that the prediction based on this framework is acceptable, as the system just needs 14.5 seconds to predict 40 minutes traffic densities for 4 cells plus 30 minutes journey time for 2 links.

Computational items	Computational time (sec) for one cycle
Inflow and outflow prediction	0.0013
SCTM propagation	5.2
Relative frequency generation and distribution fitting of link 1	2.0
Relative frequency generation and distribution fitting of link 2	1.8
Nested operation of the PMF and the distribution fitting of journey time	5.9
The total time cost in traffic state and travel time prediction	14.5

Table 5.6. The computation time

The selection of n in equation (5.9) is determined by the balance of the accuracy of prediction and computational time. Table 5.7 presents the MAPE of inflow and outflow profile prediction on March 26, 2008 over three prediction periods (7:10, 7:20, 7:30) and the whole time scale (5:00-11:00) with n=8, 4, 2, 1 on the condition of l=*8*. That is to test the 40-minutes short term prediction of inflow and outflow by referring to different length of measurement such as the past 40 minutes, 20 minutes, 10 minutes or 5 minutes. Compared with the MAPE of historical record with the test day, the result shows that the best linear prediction has obvious effect, and generally it becomes more effective as n get larger for $n \le 4$, that is to say a longer observation of real-time trend will benefit the prediction. However, n does not need to be infinitely long as too much consideration on the period far away from current time may decrease the sensitivity of prediction such as n=8. Additionally, the computational time also

becomes greater. Finally, n=4 is considered to be the best selection for equation (5.9) with l=8.

	n=8		*n* =4		*n* =2		*n* =1		Historical record	
	Inflow	*Outflow*	*Inflow*	*Outflow*	*Inflow*	*Outflow*	*Inflow*	*Outflow*	*Inflow*	*Outflow*
MAPE(Cycle A)	126%	226.0%	130.9%	213.4%	131.7%	227.2%	132.3%	232.1%	141.8%	256.2%
MAPE (Cycle B)	131.3%	119.4%	135.7%	16.6%	136.4%	31.2%	136.7%	54%	179.1%	240.5%
MAPE (Cycle C)	66.1%	33.4%	39.7%	46.5%	18.12%	11.4%	18.0%	14.6%	148.5%	154.5%
MAPE (5:00-11:00)	28.7%	31.4%	21.9%	23.9%	21.7%	25.7%	24.5%	29.0%	37.4%	49.1%
Average Computational time (Second/horizon)		0.0017		0.0013		0.0004		0.0004		--

Table 5.7 Influence of n to the precision of prediction

5.6 Summary

Chapter 5 developed an online framework by extending the stochastic cell transmission model (SCTM) to consider the spatial-temporal correlations of the uncertain flow profiles and supply functions for the short-term traffic state prediction. The traffic inflow/outflow profiles and the fundamental diagrams, conditioned on the real-time observation, are predicted via the Multivariable Normal Distribution based best linear predictor given the statistics of the historical data. The predictor is able to adjust the value of flow profiles by capturing the error between the real-time measurements and the historical mean. Meanwhile, the key processes of the SCTM, that is the evaluation of probabilities and propagation of traffic states, are improved by considering the spatial correlations of supply and traffic cell densities. The empirical study on a short freeway segment confirmed the benefit of the extended framework by demonstrating the significant improvement via comparisons of the simulation results.

Chapter 6

Conclusions and future works

This chapter presents concluding remarks on this research, highlights its significance, and suggests directions for future research. This chapter is organized in two sections. Section 6.1 summarizes the research, discusses associated conclusions and highlights the significance of the research. Some recommendations and potential values of the research to the practitioners and decision makers are given. Section 6.2 discusses possible extensions and directions for future research.

6.1 Summary and conclusions

Significant interests of researchers and other traffic engineering practitioners have been spent on the problems of estimating and predicting travel time in the past decades. However, as reviewed in Chapter 2, the existing methods either rely on the availability of the data from automatic vehicle identification (AVI) system (which may not be available in some cases) or are only able to produce deterministic estimation of travel (or journey) time. Also, these methods may not be applicable to the data provided by traditional traffic measurement systems (e.g. loop detector or auto-scope system) which can provide flow counts, density, and average speed during fixed detecting time intervals. Nevertheless, few of these existing methods can generate statistical distributions of travel time over different time periods. This thesis contributes to the literature on travel time estimation and prediction for freeway networks in two aspects: stochastic dynamic traffic flow modeling and stochastic dynamic travel time estimation and prediction.

This research extended the framework of Stochastic Cell Transmission Model (SCTM) proposed by Sumalee et al. (2011) in two major aspects. First, a practical method to approximate dynamic travel time distribution from the outputs of the SCTM was

proposed in Chapter 4. The original framework of SCTM provides the stochastic traffic density on each cell along a freeway corridor. The method proposed in Chapter 4 uses the stochastic traffic density to estimate the statistical distributions of dynamic link travel time and journey time of a given route. To be specific, Chapter 4 first introduced the definition of (deterministic) link travel time in terms of dynamic flow matching under the first-in-first-out (FIFO) principle. This definition was then extended to the stochastic version by defining a likelihood concept that measures the probability of the difference between the cumulative stochastic link inflow and outflow profiles be less than or equal to a prescribed bound. Based on this likelihood measure, the probability mass function (PMF) of the link travel time was evaluated over an appropriate sampling interval. The dynamic link travel time distribution was constructed by fitting the PMF with a skew normal distribution. After obtaining the link travel time distributions along a route, the deterministic nested delay operator was extended to evaluate stochastic journey time distribution of a given route. The PMF of journey time was defined as a series of "nested" conditional probabilities along the links on the route. By fitting the PMF of the journey time by skew normal distribution, the stochastic journey time distribution was deduced.

The empirical study was conducted on a long segment of an expressway between Toyanaka-kita and Osaka CBD to validate the proposed method for travel time estimation under incident scenarios with limited traffic measurements and incident records. The results showed a satisfactory performance of the proposed algorithm for travel time distribution estimation/prediction if the SCTM can provide an accurate estimation of the cell densities. The results on the shewness analysis of the travel time distribution are also consistent with the empirical observations in Lint et al (2008). The statistical distribution of dynamic travel time can also be used to calculate travel time reliability index. In Chapter 4, the buffer time index (BTI) was used as the travel time reliability index. The BTI was calculated for the case study reported in Chapter 4. This allows the analysis of travel time reliability, which has been discussed previously only in the context of static analysis in the literature, with the dynamic consideration.

Due to the lack of a proper prediction algorithm for demand and supply profiles and the inherent independent/uncorrelated assumptions in the original SCTM framework, the SCTM cannot predict traffic states accurately for on-line applications. This prediction error in traffic states caused the inaccurate prediction of the stochastic travel time. Chapter 5 was devoted to tackle this issue wherein the SCTM framework is extended for on-line traffic state prediction.

As discussed earlier, the difficulty to obtain accurate short-term traffic state (including travel time) estimation/prediction is due to the high traffic density and congestion in the network as well as the interaction of the demand and supply uncertainties along with the dynamic nature of traffic flow. The demand and supply uncertainties are correlated in both space and time domains. The traffic flow models and traffic forecasting methods can be further refined if the spatial correlation, i.e. systematic dependence between the observations at each location and the observations at neighboring locations, is taken into account in the modeling and prediction frameworks. The second contribution of this research is the extension of the original SCTM framework to consider spatial-temporal correlations of the model inputs (on both demand and supply sides) and traffic states. The travel time evaluation method proposed in Chapter 4 was also extended to the SCTM with spatial-temporal correlations.

A multivariate normal distribution based best linear predictor was utilized to predict the inflow demand and supply functions of the SCTM. The statistical correlations of demand and supply as analyzed from the historical traffic data are used in conjunction with the recent updated traffic data (from online measurement) to predict the demand and supply in the short-term future. The predicted demand and supply are taken as inputs to the SCTM to perform short-term traffic state prediction. For real-time or online application, the prediction is conducted in a rolling horizon manner. The extended SCTM model and prediction algorithm were tested with an empirical data from the PeMS database. As shown by the test, incorporating the prediction algorithm

for demand and supply profiles improved the accuracy of traffic state prediction of the SCTM. The empirical results also revealed the potential application of the proposed method for traffic state prediction under abnormal traffic conditions, e.g. incidents and adverse weather conditions.

6.2 Future research

6.2.1 Reliability-based stochastic dynamic traffic assignment

The proposed methods have potential applications in stochastic dynamic traffic assignment studies. In the deterministic dynamic traffic assignment framework, four key components are required:

(i) models of link and path delays;

(ii) flow dynamics;

(iii) flow propagation constraints; and

(iv) a route/departure-time choice model.

All these components are either missing from the literature on stochastic dynamic traffic assignment or not very well defined. The proposed SCTM and its extension in this thesis offer an approach to describe components (ii) and (iii). The link and route travel time calculation methods developed in this thesis offer the first component. The fourth component is an interesting direction for future studies of this research.

Indeed, the reliability analysis in this thesis provides half of the fourth component for a reliability-based stochastic dynamic traffic assignment. The remaining challenges are to model traveler's departure time/route choice behavior under uncertainties and to define dynamic user equilibrium under stochastic environments. Some preliminary results on this stream were achieved by Li (2009), Zhong (2011) and Szeto et al. (2011). In Li (2009), the author investigated travelers' choice behavior modeling under travel time uncertainty and applied the deterministic DUE to study traveler's

departure time/route choice behavior. However, the uncertainty considered in Li (2009) is mainly on the random fluctuations of the capacity from day-to-day, while other supply and demand functions are assumed to be deterministic, e.g. a fixed number of travelers, constant from day-to-day, is presumed. The DUE was solved by a simulation based solution approach. Szeto et al. (2011) proposed a Monte–Carlo cell-based multi-class dynamic traffic assignment formulation that considers random evolution of traffic states. Travelers are assumed to select routes based on perceived effective travel time, where effective travel time is the sum of mean travel time and safety margin (as in the deterministic case, this may not capture the route travel time accurately). All drivers are assumed to select routes based on the dynamic extension of the reliability-based stochastic user equilibrium principle (Shao et al., 2006) called the reliability-based stochastic-dynamic–user-equilibrium (RSDUE) principle. This RSDUE principle states that for each class of drivers departing at any time, they select routes with the minimum perceived effective travel time at the time of departure. The RSDUE problem was then investigated by certain heuristic simulation based approach.

On the other hand, Zhong (2011) investigated the stochastic system optimal traffic scheduling problem by applying stochastic dynamic programming theory to a simplified version of the SCTM. A set of optimal control laws were obtained analytically in terms of recursive coupled Riccati difference equations. As the optimal control may be fragile with respect to the model miss-specifications, e.g. errors in calibrations and sharply variations under abnormal traffic conditions, he further pursued the robust (optimal) decision policy, which would act robust with respect to the parameter miss-specifications in the traffic flow model, and to attenuate the effect of disturbances in the freeway network (wherein demand uncertainty is usually taken as a kind of disturbance). To overcome the curse of dimensionality of dynamic programming, it was suggested by Zhong (2011) that the approximate dynamic programming (ADP) approach would provide a better numerical solution algorithm. In the future works, we may adopt the stochastic dynamic programming approach as adopted in Zhong (2011) in conjunction with the RSDUE travelers' behavior

assumption as adopted in Szeto et al. (2011) to address the stochastic DUE traffic assignment problem and apply the ADP approach to obtain an efficient solution algorithm for the SDTA problem.

6.2.2 Incident detection

This work aims to propose a new methodology aims at detecting freeway incidents in real-time and characterizing incidents in terms of time-varying lane-changing fractions (or probabilities), queue lengths, sharply changes in demand (or boundary inflow profiles) or supply functions, and delays caused by incidents, etc.

Recent studies, e.g. Laval and Daganzo (2006), have revealed that a lane changing vehicle acts as a moving bottleneck on its destination lane while accelerating to the speed prevailing on the lane. As a consequence, such disruption triggers other lane changes. To more realistically capture traffic dynamics along a freeway corridor, the lane-changing behavior has to be taken into account when developing traffic flow models. On the other hand, time-varying lane-changing fractions and queue lengths play an important role in detection and subsequent traffic control of lane blocking incidents (Sheu and Ritchie, 2001; Sheu, 2004).

In the future works, we will further extend the SCTM to consider the lane-changing behavior. In the deterministic counterpart, the multi-lane hybrid cell transmission model (CTM) was proposed by Laval and Daganzo (2006). The multi-lane hybrid CTM was shown to be able to capture both the mandatory and the discretionary lane-changing behaviors, and to treat lane changing vehicles as moving bottlenecks. However, just like the CTM, this hybrid CTM is a non-linear non-differentiable transformation from one traffic state to another, particularly with the lane-choice probability involved. Similar to Sheu and Ritchie (2001) and Sheu (2004), we will define the lane changing fractions (or probabilities), e.g. lane changing fraction from the original lane to adjacent lanes and return-lane-changing fraction from adjacent

lanes to the original lane in the downstream. These lane changing fractions are then extended as state dynamics parallel to the stochastic traffic density in the existing SCTM framework. In this manner, we can represent the extended SCTM as a discrete-time nonlinear stochastic system, which characterize cell traffic densities, levels of congestion, queue lengths, inter-lane and intra-lane traffic states in terms of mean and variance or probability of occurrence. Contrary to the original SCTM wherein aggregated traffic measurements are used to calibrate and simulate the model, the extended SCTM requires lane traffic count and occupancy as two major types of input data.

The above extended stochastic cell transmission model (ESCTM) improves the original SCTM by involving a best linear predictor and an additional dynamics governing the lane-changing behavior. The best linear predictor as proposed in Chapter 5 utilizes the spatial-temporal correlations of traffic flow to forecast the demand and supply functions. Then the predicted demand and supply functions are loaded into the SCTM as exogenous signals to predict short-term traffic states in terms of traffic density and queue length. During and after a lane blocking incident, the traffic flow characteristics along the related links may change substantially, e.g. the demand and supply functions and traffic state. By comparing the real-time measurements and estimations of these characteristics with the predicted ones, such abrupt changes may be identified, and hence the incident occurrence may be recognized, leading to corresponding incident alarms.

However, this judgment may be not sufficient for real-time incident detection, especially under free-flowing traffic conditions during non-rush hour (no queue and no significant changes in the demand and supply functions would be observed). Therefore, we further utilize the lane-changing fractions given by the ESCTM to develop a more comprehensive incident detection algorithm and subsequent incident characterization. Considering the stochastic input/outputs of the SCTM framework for traffic state estimation/prediction and lane-changing fractions, probabilistic functions can also be

defined to characterize the likelihood of an incident. Furthermore, we will define several time-varying thresholds, which may be a combination of the lane-changing fractions, traffic density, queue lengths as well as the demand and supply functions, for incident detection. Similar to Sheu (2004), finally, incident detection will be formulated as a pattern recognition problem by comparing the likelihood functions with the time-varying thresholds. Incident induced delay is important to characterize the incident and for the purpose of incident management. We will also extend the methodology proposed in Chapter 4 to calculate the distribution of stochastic delay induced by an incident.

References

Asakura, Y., and Kashiwadani, M., 1991. Road network reliability caused by daily fluctuation of traffic flow. *Proceedings of the 19th PTRC Summer Annual Meeting*, Seminar G, pp. 73-84.

Bates, J. , Polak, J., Jones, P., and Cook, A., 2001. The valuation of reliability for personal travel. *Transportation Research Part E*, 37, pp. 191-229.

Bertsekas, D., 2005. Dynamic Programming and Optimal Control. *Athena Scientific*.

Bell, M., 1999. Measuring network reliability: a game theoretic approach. *Jounral of Advanced Transportation*, 33, pp. 135-146.

Bell, M., and Cassir, C., 2000. Reliability of Transport Networks. London, U.K.: *Research Studies Press*.

Bell, M., 2000. A game theory approach to measuring the performance reliability of transport networks. *Transportation Research Part B*, 34, pp. 533-546.

Bell, M., and Cassir, C. 2002. Risk-averse user equilibrium traffic assignment: An application of game theory. Transportation Research Part B, 36, pp. 671-682.

Bogers, E. and Van Zuylen H., 2004. The importance of reliability in route choice in freight transport for various actors on various levels. *In Proceedings: European Transport Conference*. Strasbourg, France.

Carey, M., Ge, Y., in press. Comparison of methods for path flow reassignment for dynamic user equilibrium. *Networks and Spatial Economics*.

Cassir, C., Yang, H., Kakan, L., Tang, W., and Bell, M., 2001. Travel time versus capacity reliability of a road network: Reliability of transport networks, *United Kingdom: Research Studies Press*.

Chandra, S., Al-Deek, H., 2009. Predictions of freeway traffic speeds and volumes using vector autoregressive models. *Journal of Intelligent Transportation Systems,* 13, pp. 53–72.

Chang, T., Nozick, L., and Turnquist, M., 2005. Multiobjective path finding in stochastic dynamic networks, with application to routing hazardous materials shipments. *Transportation Science*, 39, pp. 383-399.

Chen, C., 2003. Freeway performance measurement system (PeMS). *PhD dissertation*, University of California, Berkeley.

Chen, C., Jia, Z., and Varaiya, P., 2001. Causes and cures of highway congestion. *IEEE Control Systems Magazine*, 21, pp. 26-33.

Chen, C., Skabardonis, A., and Varaiya, P., 2003. Travel-time reliability as a measure of service. *Transportation Research Record*, 1855, pp. 74-79.

Chow, A., 2007. System optimal traffic assignment with departure time choice. *PhD dissertation*, University of London.

Chui, C., Chen, G., 2009. Kalman filtering with real-time applications, 4th edition. Springer-Verlag Berlin Heidelberg.

Clark, S., Dougherty, M., and Kirby, H., 1993. The use of neural networks and time series models for short-term traffic forecasting: a comparative study. In: Transportation planning methods, *Proceedings of the PTRC 21st annual summer meeting*, Manchester.

Coifman, B., and Cassidy, M. 2002. Vehicle re-identification and travel time measurement on congested freeways. *Transportation Research Part A*, 36, pp. 899-917.

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

Daganzo, C., 1995. The cell transmission model: Network traffic. *Transportation Research Part B*, 29, pp. 79-93.

Dion, F., and Rakha, H., 2006. Estimating dynamic roadway travel times using automatic vehicle identification data for low sampling rates. *Transportation Research Part B*, 40 (9), 745-766.

Federal Highway Administration, 2005. Traffic Congestion and Reliability: Trends and Strategies for Advanced Mitigation. U.S. Department of Transportation.

Federal Highway Administration. 2007. Travel Time Reliability: Making It There on Time, All the Time. U.S. Department of Transportation.

Fei, X., Lu, C., Liu, K., 2011. A Bayesian dynamic linear model approach for real-time short-term freeway travel time prediction. *Transportation Research Part C*, 19 (6), pp. 1306-1318.

Fu, L., and Rilett, L., 1998. Expected shortest paths in dynamic and stochastic traffic networks. *Transportation Research Part B*, 32, pp. 499-516.

Gao, S., and Chabini, I., 2006. Optimal routing policy problems in stochastic time-dependent networks. *Transportation Research Part B*, 40, pp. 93-122.

Gao, S., Frejinger, E., Ben-Akiva, M., 2008. Adaptive route choice models in stochastic time-dependent networks. *Transportation Research Record*, 2085, pp. 136-143

Gao, S., Frejinger, E., Ben-Akiva, M., 2010. Adaptive route choices in risky traffic networks: A prospect theory approach. Transportation Research Part C, 18(5), 727-740.

Gao, S., Huang, H., 2012, Real-time traveler information for optimal adaptive routing in stochastic time-dependent networks. *Transportation Research Part C*, 21 (1), pp. 196-213.

Haberman, S., 1977, Maximum likelihood estimates in exponential response models. **Annals of Statistics**, 5, pp. 815-841.

Haghani, A., Hamedi, M., Sadabadi, K. F., Young, S., Tarnoff, P., 2010. Data collection of freeway travel time ground truth with Bluetooth sensors. *Transportation Research Record*, 2160, pp. 60-68.

Herrera, J.,Work, D., Ban, X., Herring, R., Jacobson, Q., and Bayen, A., 2010. Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment. *Transportation Research Part C*, 18(4), 568-583.

Highway capacity manual, 2000. *National Research Council*, Washington, D.C.

Hinsbergen, C., Van Lint, J., and Van Zuylen, H. J., 2009. Bayesian committee of

neural networks to predict travel times with confidence intervals. *Transportation Research Part C*, 17, pp. 498-509.

Hollander, Y., and Liu, R., 2008. Estimation of the distribution of travel times by repeated simulation. *Transportation Research Part C*, 16, pp. 212-231.

Jula, H., Dessouky, M., and Ioannou, P., 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, pp. 97-110.

Kharoufeh, J., and Gautam, N., 2004. Deriving link travel-time distributions via stochastic speed processes, *Transportation Science*, 38, pp.97-106.

Karlaftis, M., and Vlahogianni, E., 2011. Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. *Transportation Research Part C*, 19, pp. 387-399.

Kwon, J., Varaiya, P., and Skabardonis, A., 2003. Estimation of truck traffic volume from single loop detectors with lane-to-lane speed correlation. *Transportation Research Record*, 1856, pp. 106-117.

Lam, W., Chan, K., Tam, M., and Shi, W., 2005. Short-term travel time forecasts for transport information system in Hong Kong. *Journal of Advanced Transportation*, 39, pp. 289-305.

Laval, J.A., and Daganzo, C.F., 2006. Lane-changing in traffic streams. *Transportation Research Part B*, 40, pp. 251–264.

Li, H., 2009. Reliability-based dynamic network design with stochastic network. *PhD thesis*, University of Delft.

Liu, H., Ma, W., Wu, X. and Hu, H., 2010. Real-time estimation of arterial travel time under congested conditions. *Transportmetrica*, doi: 10.1080/18128600903502298

Lo, H., 2002. Trip travel time reliability in degradable transport networks. In *Procdeeing of the 15th International Symposium On Transportation and Traffic Theory (ISTTT)*. Adelaide, South Australi.

Lo, H., 2001. A cell-based traffic control formulation: Strategies and benefits of

dynamic timing plans. *Transportation Science*, 35, pp. 148-164.

Lo, H., and Szeto, W.Y., 2002. A cell-based variational inequality formulation of the dynamic user optimal assignment problem. *Transportation Research Part B*, 36, pp. 421-443.

Lomax, T., Schrank, D., Turner, S., and Margiotta, R., 2003. Selecting travel reliability measures. *Texas Transportation*.

Lyman, K., and Bertini, R., 2008. Using travel time reliability measures to improve regional transportation planning and operations. *In Proceedings of the Transportation Research Board's 87th Annual Meeting*, Washington, DC., USA.

Mahmassani, H., Haas, C., Logman, H., and Shin, H., 2002. Summary and analysis: Integration of point-based and link-based incident detection and traffic estimation. *Research project report of centre for transportation research*, the University of Texes.

Miller-Hooks, E., and Mahmassani, H., 1998. Optimal routing of hazardous materials in stochastic time-varying transportation networks. *Transportation Research Recor*d, 1645, pp.143-151.

Miller-Hooks, E., and H. Mahmassani, 2000. Least expected time paths in stochastic, time varying transportation networks, *Transportation Science*, 34, pp. 198-215.

Min, W., and Wynter, L., 2011. Real-time road traffic prediction with spatio-temporal correlations. *Transportation Research Part C*: *Emerging Technologies*, 19, pp. 606-616.

Mun, J., 2007. Traffic performance models for dynamic traffic assignment: An assessment of existing models. *Transport Reviews*, 27, pp. 231-249.

Munoz, L., Sun, X., Horowitz, R., and Alvarez, L., 2003. Traffic density estimation with the cell transmission model. *Proceedings of the American Control Conference*, Denver, Colorado, pp. 3750-3755.

Ng, M., and Waller, S., 2009. Dynamic route choice model in face of uncertain capacities. *Proceedings of the Transportation Research Board's 88th Annual Meeting*, Washington, D.C., USA.

Ngoduy, D., 2011. Multiclass first order model using stochastic fundamental diagrams. *Transportmetrica*, 7, pp. 111-125.

Nie, Y., and X. Wu, 2009. Shortest path problem considering on-time arrival probability. *Transportation Research Part B,* 43, pp. 597-613.

Nie, X., and Zhang, H., 2005. A comparative study of some macroscopic link models used in dynamic traffic assignment. *Networks and Spatial Economics*, 5, pp. 89-115.

Pal, R. and Bose, I., 2009. An optimization based approach for deployment of roadway incident response vehicles with reliability constraints. *European Journal of Operational Research,* 198, pp. 452-46.

Papageorgiou, M., and Kotsialos, A., 2002. Freeway ramp metering: An overview. *IEEE Transactions on Intelligent Transportation Systems*, 3, pp. 271-281.

Papageorgiou, M., Diakaki, C., Dinopoulou, V., Kotsialos, A., and Wang, Y., 2003. Review of road traffic control strategies. *Proceedings of the IEEE*, 91, pp. 2043-2067.

Pattanamekar, P., Park, D., Rilett, L., Lee, J., and Lee, C., 2003. Dynamic and stochastic shortest path in transportation networks with two components of travel time uncertainty. *Transportation Research Part C,* 11 , pp. 331-354.

Pearce, V., 2001. Can I make it to work on time? T*raffic Technology International*, 10, pp. 16-18.

Peeta, S., and Ziliaskopoulos, A., 2001. Foundations of dynamic traffic assignment: The past, the present and the future. *Networks and Spatial Economics*, 1 (3/4), pp. 233-266.

Petty, K., Bickel, P., Ostland, M., Rice, J., Schoenberg, F., Jiang, J., and Ritov, Y., 1998. Accurate estimation of travel times from single-loop detectors. *Transportation Research Part A*, 32 (1), pp. 1-17.

Pulugurtha, S. and Pasupuleti, N., 2010. Assessment of link reliability as a function of congestion components. *Journal of Transportation Engineering*, 136, pp. 903-913.

Samaranayake, S., Blandin, S., Bayen, A., in press. A tractable class of algorithms for reliable routing in stochastic networks. *Transportation Research Part C.*

Schönhof, M., and Helbing, D., 2007. Empirical features of congested traffic states and their implications for traffic modeling. *Transportation Science*, 41, 135-166.

Shao, H., Lam, W., and Tam, M., 2006. A reliability-based stochastic traffic assignment model for network with multiple user classes under uncertainty in demand. *Networks and Spatial Economics*, 6, pp.173-204.

Sheu, J. B., and Ritchie, S. G., 2001. Stochastic modeling and real-time prediction of vehicular lane-changing behavior. *Transportation Research Part B*, 35, pp. 695-716.

Sheu, J. B., 2004. A sequential detection approach to real-time freeway incident detection and characterization. *European Journal of Operational Research*, 157 (2), 471–485.

Sheu, J., Lan, L., and Huang, Y., 2009. Short-term prediction of traffic dynamics with real-time recurrent learning algorithms. *Transportmetrica*, 5, pp. 59-83.

Smith, B., Williams, B., and Oswalsd, R., 2002. Comparison of parametric and nonparametric models for traffic flow forecasting. *Transportation Research Part C*, 10, pp. 303–321.

Soriguera, F., Rosas, D., and Robusté, F. 2010. Travel time measurement in closed toll highways. *Transportation Research Part B*, 44, pp.1242–1267.

Stathopoulos, A., Karlaftis, M., 2003. A multivariate state-space approach to urban traffic flow prediction. *Transportation Research Part C*, 11, pp. 121–135.

Sumalee, A., and Wang, J., 2012. Probabilistic fusion of vehicle features for re-identification and travel time estimation using video image data. Preprint submitted to the *91st Annual Meeting of Transportation Research Board*.

Sumalee, A., Zhong, R., Pan, T., and Szeto, W., 2011. Stochastic cell transmission model (SCTM): A stochastic dynamic traffic model for traffic state surveillance and assignment. *Transportation Research Part B*, 45, pp.507-533.

Sun, X., Munoz, L., Horowitz, R., 2003. Highway traffic state estimation using improved mixture Kalman filters for effective ramp metering control. *Proceedings of the 42nd IEEE Conference on Decision and Control*, Hawaii, pp.6333-6338.

Szeto, W.Y., Jiang, Y., and Sumalee, A., in press. A cell-based model for multi-class doubly stochastic dynamic traffic assignment. *Computer-Aided Civil and Infrastructure Engineering*.

Tam, M.L. and Lam, W., 2008. Using automatic vehicle identification data for travel

time estimation in Hong Kong. *Transportmetrica*, 4, pp. 179-194.

Tu, H., 2008. Monitoring Travel Time Reliability on Freeways. TRAIL Thesis Series, the Netherlands.

Van Lint, J., 2004. Freeway travel time reliability maps, using the shape of the day-to-day travel time distribution. TRAIL Thesis Series, the Netherlands.

van Lint, J., 2008. Online learning solutions for freeway travel time prediction. *IEEE Transactions on Intelligent Transportation Systems*, 9, pp. 38-47.

van Lint, J., Van Zuylen, H., and Tu, H., 2008. Travel time unreliability on freeways: Why measures based on variance tell only half the story. *Transportation Research Part A*, 42, pp. 258-277.

van Lint, J., and van Zuylen, H., 2005. Monitoring and predicting freeway travel time reliability. *Transportation Research Record*, 1917, pp. 54-62.

Vlahogianni, E., Golias, J., and Karlaftis, M., 2004. Short-term traffic forecasting: Overview of objectives and methods. *Transportation Reviews*, 24, pp. 533-557.

Vlahogianni, E., Karlaftis, M., Golias, J., 2005. Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach. *Transportation Research Part C*, 13, pp. 211-234.

Vlahogianni, E., Karlaftis, M., Golias, J., 2007. Spatio-temporal short-term urban traffic volume forecasting using genetically optimized modular networks. *Computer-Aided Civil and Infrastructure Engineering*, 22, pp. 317–325.

Waller, S., Ziliaskopoulos, A., 2002. On the online shortest path problem with limited arc cost dependencies. *Networks*, 40(4), pp. 216-227.

Wang, H., Li, J., Chen, Q., and Ni, D., 2009. Speed-density relationship: From deterministic to stochastic. *In Proceedings of the Transportation Research Board's 88th Annual Meeting*, Washington, DC., USA.

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

Wen, Y., 2008. Scalability of dynamic traffic assignment. *PhD dissertation*, Massachusetts Institute of Technology.

Wu, C., Ho, J., and Lee, D., 2004. Travel-time prediction with support vector regression. *IEEE Transactions on Intelligent Transportation Systems*, 5, pp. 276-281.

Yang, H., Bell, M., and Meng, Q., 2000. Modeling the capacity and level of service of urban transportation networks. *Transportation Research B*, 34, pp. 255-275.

Yeon, J., Elefteriadou, L., and Lawphongpanich, S., 2008. Travel time estimation on a freeway using discrete time Markov chains. *Transportation Research Part B*, 42, pp. 325-338.

Yuan, J., Mills, K., 2005. A cross-correlation-based method for spatial-temporal traffic analysis. *Performance evaluation*, 61, pp. 163-180.

Zhong, R., 2011. Dynamic assignment, surveillance and control for traffic network with uncertainties. *Ph.D Thesis*, The Hong Kong Polytechnic University.

Zhong, R., Sumalee, A., Pan, T., and Lam, W.H.K, in press. Stochastic cell transmission model for traffic network with demand and supply uncertainties. Submitted to Transportmetrica, Transportmetrica..