ABSTRACT

Title:	TRAFFIC FLOW MODELING WITH REAL-TIME DATA FOR ON-LINE NETWORK TRAFFIC ESTIMATION AND PREDICTION
	Xiao Qin, Ph.D., 2006
Directed By:	Professor Hani S. Mahmassani, Department of Civil and Environmental Engineering

This research addresses the problem of modeling time-dependent traffic flow with real-time traffic sensor data for the purpose of online traffic estimation and prediction to support ATMS/ATIS in an urban transportation network. The fundamental objectives of this study are to formulate and develop a dynamic traffic flow model driven by real-world observations, which is suitable for mesoscopic type dynamic traffic assignment simulation.

A dynamic speed-density relation is identified by incorporating the physical concept in continuum and kinetic models, coupled with the structural formulation of the transfer function model which is used to represent dynamic relationship. The model recognizes the time-lagged response of speed to the influential factors (speed relaxation, speed convection and density anticipation) as well as the potential autocorrelated system noise. The procedures adapted from transfer function theory are presented for the model estimation and speed prediction using the real-time data. Speed prediction is performed by means of minimum mean square error and conditional on the past information.

In the context of real-time dynamic traffic assignment simulation operation, a framework based on the rolling-horizon methodology is proposed for the adaptive calibration of dynamic speed-density relations to reflect more recent traffic trends. To deal with the different time scales in the data observation interval and the traffic simulation interval, an approximation procedure is proposed to derive proper impulse responses for traffic simulation. Short term correction procedures, based on feedback control theory, are formulated to identify discrepancies between simulation and real-world observation in order to adjust speed periodically.

Numerical tests to evaluate the dynamic model are conducted in a standalone manner firstly and then by integrating the model into a real-time DTA system. The overall conclusion from the results is that the proposed dynamic model is preferable in the context of real-time application to the use of conventional static traffic flow models due to its higher responsiveness and accuracy, although many other aspects remain to be investigated in further steps.

TRAFFIC FLOW MODELING WITH REAL-TIME DATA FOR ON-LINE
NETWORK TRAFFIC ESTIMATION AND PREDICTION

By

Xiao Qin

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2006

Advisory Committee:
Professor Hani S. Mahmassani, Chair
Professor Eyad H. Abed
Professor Ali Haghani
Professor Paul M. Schonfeld
Assistant Professor Richard J. La

Dedication

To my son, Eddy Danlin Qiu

Acknowledgements

I would like to express my sincere gratitude to my advisor Dr. Hani S. Mahmassani for his enthusiastic encouragement, continued patience and generous support through my PhD study at University of Maryland and the University of Texas at Austin. It has been an honor to work with Dr. Hani S. Mahmassani and benefit from his passionate attitude and enlightening discussion in scientific research and outlook on life. The challenging academic experience with Dr. Hani S. Mahmassani has provided me an excellent base from which I will start my professional path.

I am also thankful to Dr. Eyad H. Abed, Dr. Ali Haghani, Dr. Paul M. Schonfeld, and Dr. Richard J. La, members of my dissertation committee, for their participation and helps. Their valuable comments and suggestions to my initial work directly contribute to the completion of my dissertation.

I would like to thank my colleagues and friends in College Park, MD and Austin, Texas for the pleasant collaboration and genuine friendship. Special thanks firstly go to Dr. Yi-Chang Chiu, Dr. Nathan N. Huyhn, Dr. Xuesong Zhou, Dr. Khaled Abdelghany, and Dr. Ahmed Abdelghany, for their advices during the beginning of my doctoral studies. I particularly acknowledge Dr. Xuesong Zhou, Robert H. Mahfoud, Hayssam Sbayti, Chung-Cheng Jason Lu, and Xiang Fei for their extensive contribution to and abundant knowledge of the real-time DTA system. I am continuously inspired by them and could not fulfill my dissertation work without their constructive inputs. I am also grateful to the other DTA research team member, Sevgi Erdogan, Jing Dong, Kuilin Zhang, Stacy Eisenman, and Samer Hamdar for their countless helps.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1   _Research motivation and objectives_

Traffic congestion remains a major societal concern across the world, with no visible signs of substantial reduction in the future. The growth rate for the number of U.S. vehicles from 1980 to 2003 is roughly nine times that for the lane-miles of highway [Bureau of Transportation Statistics, 2004]. Statistics for 85 U.S. cities indicate that the annual person-hours of highway traffic delay per person increased by 243 percent from 1982 to 2002. It is widely agreed that physical capacity addition cannot and will not keep up with increasing demand. Modern and efficient management of existing systems is called upon to deliver considerable improvement in transportation service levels and productivity. Intelligent Transportation System (ITS) technologies (including sensing, location, communications and information technologies) have become essential for the modern management of transportation networks [National ITS Architecture, 2003]. Advanced Traveler Information System (ATIS) and Advanced Traffic Management System (ATMS) integrate such technologies (advanced surveillance systems over a road network, digital sensing and communication between a control center and vehicles) to monitor, manage and control vehicular traffic in a road network, and provide travelers information and guidance, in order to mitigate congestion and enhance safety.

The objectives of ATIS/ATMS call for the monitoring and the response actions to occur in _real time,_ which places particular challenges on the models and algorithms used in these applications. The availability of advanced real-time traffic

simulation tools is critical in order to provide a quasi-continuous view of the state of the traffic system over time and space. These tools are intended to perform real-time system-wide traffic estimation and prediction, based on the existing surveillance system, and meet the information requirements for decision making for operators and users of the traffic network. Simulation-based Dynamic Traffic Assignment (DTA)-type models address these needs, in support of complex traffic control and management functions in the dynamic ITS environment. The information provided by DTA systems, generally including descriptive traffic conditions (current and near future) and normative route guidance, provide a basis for reducing delays on major highways, improving the safety, efficiency, and capacity of existing transportation systems, and enhancing area-wide emergency response through information sharing and coordination. All these intelligent functions in the simulation-based DTA tools are predicated on the availability of reliable and robust traffic flow models capable of representing the dynamic evolution of traffic over space and time.

This research addresses the problem of modeling time-dependent traffic flows when real-time traffic sensor data are available on a subset of the network links for the purpose of online traffic estimation and prediction to support ATMS/ATIS in an urban transportation network.

Traffic flow models have developed over nearly seven decades of research and application; these models can be broadly categorized into microscopic, macroscopic, and mesoscopic in terms of level of detail and process representation. Microscopic traffic flow models provide a detailed representation of individual driver behavior processes in situations such as car-following [Herman et al. 1959, Herman

2

and Potts 1959], lane changing and gap acceptance. As such, these models seek to represent the traffic flow process by describing the behavior of the entities that make up the traffic stream as well as their interactions in detail. This makes microscopic traffic flow models suitable for evaluation of complicated traffic operations under fine-grained representation of road geometries, but usually only for a small network due to their intensive computational requirements. On the other hand, macroscopic traffic flow models, such as the LWR model [Lighthill and Whitham 1955, and Richards 1956], describe the traffic state at a certain level of aggregation as volume (density, or speed) without considering its constituent particles. The lower level of detail not only reduces the computational burden but also facilitates model calibration. It is important to note in this regard that the success of macroscopic models of traffic derives from the well-documented existence of so-called "collective effects", resulting from the complex, generally nonlinear, interaction of individual particles in the traffic stream [Herman, 1992]. These collective effects result in relatively simple relations between averages that describe the traffic state, which have proven to be considerably more robust than individual-level microscopic models. Mesoscopic traffic flow models [Prigogine 1961, Prigogine and Herman 1971], at an intermediate level of detail, describe traffic flow dynamics in aggregated forms but distinguish driver behavior individually.

Conforming to the above distinction, traffic simulation modeling has developed either microscopically or macroscopically in the conventional simulation packages like CORSIM (a microscopic simulation model) and FREFLO (a macroscopic simulation model). On the other hand, mesoscopic simulation modeling

is gaining popularity because it is appropriate to larger networks when computation resources must be managed effectively and some level of detail is still needed. Note that it may not be necessary in a so-called mesoscopic simulation model that traffic interactions be represented according to theories that are classified as mesoscopic (e.g. gas-kinetic traffic flow theory). For instance, DYNASMART [Mahmassani et al. 1994 and 2001] traces individual vehicles, but moves them according to local speeds determined consistently with *macroscopic* traffic flow relations. Other representative mesoscopic simulation models are DYNAMIT [Ben-Akiva, 1998] and CONTRAM [Leonard, 1989].

Traffic flow modeling in *real-time* is a new and challenging area of application of traffic theories, motivated by developments in intelligent transportation technologies and their widespread deployment. To the extent that it is used as the fundamental core of traffic estimation and prediction capabilities, traffic flow modeling in real time gives rise to challenging theoretical and methodological questions, which directly affect the quality of the traffic estimation and prediction, and as such would impact decisions made by traffic operators and users on the basis of these predictions. Meanwhile, surveillance installations have dramatically increased the availability of quasi-continuously collected surveillance data, which could enable more accurate and responsive modeling of traffic flows. Generally, the parameters in any traffic flow model should be estimated using actual data to reflect traffic characteristics of the facility under study. Considering the dynamics in traffic time-series data, traffic flow modeling in real-time is meant to interpret the real-time data rapidly so as to calibrate and update the internal traffic flow model in the traffic

state estimation and prediction system. Therefore, it is anticipated that the online traffic simulation results could provide better agreement with actual real-world conditions if based on real-time traffic flow modeling, compared to a traffic flow model pre-calibrated with historical data. As noted, online traffic flow modeling is a relatively new topic, for which the related research and application are, to date, limited, preliminary and far from mature.

Driven by the motivation mentioned above, this study is concerned with how time-dependent traffic flow patterns can be represented sufficiently and efficiently within the mesoscopic type DTA model and how the collected real-time traffic data can be applied to enhance the accuracy and reliability of the traffic estimation and prediction system.

Therefore, the problem addressed in this dissertation research is as follows: Given a stream of real time observations (traffic speed and density) of the time-varying traffic on a subset of links in a network, it is sought to predict in real time the dynamic traffic speeds within the network, which would be used to move vehicles and determine travel time along paths within the simulation-based dynamic traffic assignment system, so as to result in a maximally consistent match between actual data and predicted values of the network traffic states.

To address the above problem, the fundamental objectives of this research include:

1. formulate a dynamic traffic flow relation which is adapted from the higher order macroscopic continuum model, powered by the general concept of dynamic system modeling and driven by quasi-continuous real-world observations.

2. develop and test an efficient algorithmic implementation for the dynamic traffic flow relation to provide speed predictions in real-time.

3. develop an effective framework for integrating the dynamic flow model into the online simulation-based DTA system (e.g. DYNASMART-X).

The first objective is mainly to enhance the traffic flow model in the mesoscopic-simulation-based DTA system from the simple order macroscopic continuum model to the higher order one, recognizing the massive amount of information available in the form of real-time traffic measurements, and the role of adaptive techniques that have been developed in the area of time-series analysis.

The second objective is to configure an algorithmic procedure specific to the proposed dynamic traffic flow model. The major concern for the algorithm is to be able to calibrate the model and make predictions adaptively with sufficient accuracy and computational efficiency.

The last objective recognizes the importance of certain aspects that could impact the model compatibility and efficiency in the entire DTA system. The difference in the time scale of the traffic sampling and the traffic simulation pace is one of the examples. The possibility of having short term correction to further enhance the prediction accuracy is another interesting aspect. The corresponding strategies are to be put forward to address these issues.

In sum, the application of the dynamic traffic flow model is taken as part of an online operational capability for dynamic traffic assignment (DTA) simulation modeling to predict network traffic conditions in real-time, in order to support traffic operations management and information distribution.

1.2   Overview of approach

In the approaches explored in this research, a dynamic speed-density relation is defined by incorporating the physical concept in continuum and kinetic models, with the structural formulation of the transfer function model.

The proposed model explicitly includes phenomena such as speed relaxation, speed convection and density anticipation, which are found to affect the dynamics of traffic speeds. In other words, the average speed in a section is not only dependent on the traffic density in that section but also on the traffic dynamics of the local and neighboring sections.

The proposed model recognizes the time-lagged response of speed to the influential factors as well as the autocorrelated system noise, which forms a general modeling structure for a time-based dynamic system. By applying techniques adapted from time-series theory, procedures based on the least squared method are presented for model estimation using the real-time data. Minimum mean square errors prediction of speeds is given by the expected prediction at present time. This expectation is conditional on the knowledge of the series of the past information.

A rolling horizon framework is proposed for the adaptive calibration of dynamic speed-density relations in the context of online dynamic traffic assignment simulation operation. Such an adaptive mechanism provides a systematic way to maintain an updated traffic flow relation that is consistent with the most recent traffic states.

The integration of the dynamic traffic flow model within network dynamic traffic assignment procedures is then explored. The difference in time resolution of observation intervals and simulation intervals is encountered when applying the calibrated model to the traffic simulation. The problem is resolved by implying a supposed underlying continuous system for a discrete system. The procedure includes the steps of approximation and re-discretisation to generate the parameters with a desired time interval.

Furthermore, to lower the potential inconsistency due to unknown and uncontrolled factors in the whole simulation modeling system, the short term correction procedures are formulated to identify discrepancies between simulation and real-world observations periodically (usually every sampling interval). The discrepancies are used to direct the adjustment of speeds. The adjustments could be triggered by speed-deviations or density-speed-deviations. The associated tuning factors are estimated adaptively using the least-squared-error method, constrained by the DTA model to reach internal and external consistency.

The test results indicate that the proposed model is preferable to the use of conventional static traffic flow models in the context of real-time application.

1.3   *Dissertation organization*

A review of the related literatures is presented in Chapter 2. Classical traffic flow theories, including microscopic, macroscopic and mesoscopic ones, are discussed in general terms. Traffic flow modeling in the mesoscopic simulation-based DTA system is briefly introduced. Particular attention is devoted to the

approaches and techniques pertaining to traffic flow forecasting. In Chapter 3, the formulation of dynamic speed-density relations is presented. The transfer function model from the time-series theory is proposed to provide the formulation structure for the dynamic traffic flow relation. A brief introduction of the transfer function method is followed by description of the model specification based on higher-order continuum traffic flow theories. The approaches to estimate the model and perform forecasting using the real-time traffic data are then presented. To accommodate the requirement of online operations, an adaptive procedure for model calibration and speed prediction is proposed. In Chapter 4, a series of standalone link-level experiments are designed to evaluate the performance of the model presented in Chapter 3 under various scenarios. The model is evaluated via these standalone tests before it can be properly integrated into the dynamic traffic assignment simulation environment. In Chapter 5, the approach for real-time traffic flow modeling is discussed. First, the strategy to deal with the different discrete dynamic systems in the observations and the simulation is provided by recognizing the connection to the underlying continuous dynamic system. In addition, feedback control theory is applied to the online traffic flow modeling to ensure the maximum consistency between simulation and real-world observations. The short term correction procedure which is meant to adjust speed periodically is formulated based on either speed deviations or density-speed deviations. Finally adaptive estimation of control factors for short term correction is proposed. In Chapter 6, the performance of the proposed model and methodology is examined by conducting extensive experiments. In

Chapter 7, overall summary and research contributions, as well as future extensions, are presented.

Chapter 2: Background Review

2.1 *Introduction*

As mentioned in the previous chapter, modeling traffic flow is an essential requirement and capability in a traffic analysis simulation model. The representation of traffic flow has a great impact on the simulation performance. Advanced surveillance technology provides rich data sources to identify the properties of traffic flows. In the following sections, the relevant researches are reviewed. First, traffic flow theories are described briefly since the study presented here is conducted by absorbing the essence of the theories developed and the findings discovered over decades. Then, an introduction of the simulation-based dynamic traffic assignment system DYNASMART-X and description of traffic flow modeling in the traffic simulation part are presented. Finally, the previous studies and various practices regarding traffic flow estimation and prediction with real-time data are summarized.

2.2 *Overview of traffic flow theories*

Modeling link traffic flow behavior has been a subject of considerable interest and brought on much debate in transportation science and its applications for more than a half century. Traffic flow theories are classified in terms of their level of detail, namely, microscopic, mesoscopic and macroscopic traffic flow theory. Traffic flow theory plays a central role in a traffic simulation model which basically imitates traffic dynamics in the real-world to meet the needs of traffic planning and management. In the following parts, physical background, model development and

computation implementation as well as corresponding simulation models are addressed for each type of traffic flow theories. Since the dissertation research presented here is mainly based on macroscopic continuum traffic flow theory, macroscopic models are discussed in more detail than the other two types. Extensive bibliographic information about various traffic flow theories can be found in the recent reviews by Hoogendoorn [1999] and Helbing [2001].

2.2.1   Microscopic traffic flow theories

Microscopic traffic flow theory aims to achieve the highest fidelity about traffic processes, describing the individual behavior of the entities making up the traffic stream as well as their interactions in detail.

Car following models (also called follow-the-leader models), emerging from 1950's as the first and mainstream type of microscopic traffic flow models, describe the dynamics of one vehicle following another and their interactions. The models assume that the acceleration of a vehicle is given by the neighboring vehicles among which the next vehicle ahead (leading vehicle) has the dominant influence [Helbing 2001]. Many variants of microscopic car following models have been developed, such as the general force model as the earliest concept [Pipes 1953], the stimulus response model with sensitivity factors which avoids unrealistic variation of vehicle velocities [Gazis et al. 1961], the optimal velocity model which reflects an adaptation to a headway-dependent velocity [Newell, 1961], and the recent IDM (intelligent driver model) containing a driver response to the relative velocity with respect to the leading vehicle [Treiber et al. 2000]. Car following models are mainly used to

analyze traffic stability [Herman et. al. 1959] and suitable for evaluation of complicated traffic operations, but only for a small network (e.g. corridor, subarea) due to its excessive computation requirement caused by its elaborate nature. Exemplary simulation packages applying car following theory are PARAMICS (by EPCC, Quadstone and SIAS, UK), VISSIM (by PTV, Germany), CORSIM (by FHWA), and MITSIM (by MIT) where lane-changing behavior is incorporated for multilane cases.

Cellular automata (CA) or particle hopping models started to be applied in microscopic traffic flow studies in 80's. CA can model complicated dynamic behavior and is efficient in terms of computation speed. In the model, a street section is split into small cells (typically 7.5 meter) and time is also discretised into small intervals (typically 1 second) which play the role of an updating step of the adaptation time and the safe time clearance. The positions of the vehicles are changed according to the updated speed as a function of the previous speed, maximum speed, headway and probability of spontaneous deceleration [Nagel and Schreckenberg 1992]. Some other sophisticated variants are also developed [Nagel and Paczuski 1995, Brilon and Wu 1999]. Due to their relative compactness, the models are applicable to larger networks compared to car following models. One of the developed simulation packages, OLSIM (by the University of Duisburg), bases the micro-simulation on Cellular Automata.

2.2.2 Macroscopic traffic flow theories

Research on the macroscopic traffic flow modeling starts when the solution of the conservation equation applied to traffic flow is presented [Lighthill and Whitham 1955, Richards 1956]. Typical macroscopic modeling employs the first-order or higher-order continuum representation of traffic flow by analogy with flow of continuous media like fluids, thus also called kinematic traffic theory or hydrodynamic traffic theory. Two basic equations always hold in all the macroscopic traffic flow models. One is the conservation equation which is an accurate physical law. It implies leaving vehicles are equal to entering vehicles plus vehicles stored in the traffic system, expressed as a partial differential equation (see Equation [2-1]). Another equation is the basic traffic flow equation by definition, namely volume equals density times speed (see Equation [2-2]).

$$\frac{\partial q}{\partial x} + \frac{\partial k}{\partial t} = 0 \qquad\qquad [2\text{-}1]$$

$$q = k \cdot u \qquad\qquad [2\text{-}2]$$

where q stands for volume, k for density and u for speed; x is for space; t is for time.

In the first-order continuum model which is also known as LWR model, speed at a temporal and spatial point is merely determined by the equilibrium speed given a concurrent density at the same point, say

$$u = u^e(k) \qquad\qquad [2\text{-}3]$$

where u and k are simplifications of speed $u(x,t)$ and density $k(x,t)$ for convenience; $u^e(k)$ represents the equilibrium speed determined by the equilibrium

speed-density relationship. The relation originates from empirical observations for traffic stream characteristics that reveal the phenomena of decreasing speeds with increasing densities. Several forms for the speed density relations have been proposed over the years through the empirical observations, for instance the Greenshields' model [Greenshields 1935] [Equation 2-4] and the Greenberg's model [Greenberg 1959] [Equation 2-5].

$$u^e(k) = u_f \left[1 - \left(\frac{k}{k_j} \right) \right]$$
[2-4]

$$u^e(k) = u_0 \ln \left(\frac{k}{k_j} \right)$$
[2-5]

where u_f in Equation [2-4] is the free mean speed and u_0 in Equation [2-5] is the optimum speed, respectively; k_j in both equations is the jam concentration.

The LWR model is the most popular model applied in practice due to its simplicity and ability to reproduce the most important features of traffic flow (shock waves and rarefaction waves). However, when apply Equation [2-4] or [2-5] to Equation [2-3], the speed u is adjusted instantaneously to the density k, which is a stationary relation. Mathematical discontinuity (infinite deceleration brought by shock wave) exists in the model. It is a violation of reality in which there exist some time delay and certain traffic propagation along a link, to which speed is adapted. Other shortcomings of the LWR models include inabilities to characterize stop-start waves, hysteresis phenomenon and bifurcation behavior, and non-stationary dynamics [Kuhne and Michalopoulos, 1998].

In many cases, the 'deficiencies' of the LWR models can be tackled by switching to higher-order models. Payne [1971] and Whitham [1974] show that the actual speed of a small ensemble of vehicles is obtained from the equilibrium speed-density relation after a reaction time and from an anticipated location, which is often called acceleration and inertia effects. It results in the higher-order continuum model (or PW model) which introduces higher order terms to form the following momentum equation:

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = \frac{1}{\tau} [u^e(k) - u] - c_0^2 \frac{1}{k} \frac{\partial k}{\partial x} \qquad [2\text{-}6]$$

where τ is a reaction time coefficient; c_0^2 is an anticipation coefficient.

In the partial differential equation [2-6], the speed is no longer modeled as a stationary variable but one deviating from the equilibrium state by an inherent rule associated with flow nature and driving behavior. The left-hand side of Equation [2-6] represents acceleration that is separated into a local acceleration $\frac{\partial u}{\partial t}$ and a convection term $u \frac{\partial u}{\partial x}$. The right-hand side of Equation [2-6] includes a relaxation term which reflects drivers' speed adjustments to the equilibrium one, and an anticipation term which implies drivers' reactions to traffic condition ahead.

In the higher-order model, Payne identified three different aspects for the momentum equation [2-6]: a convection term, $u \frac{\partial u}{\partial x}$, describing how the space-mean speed changes due to the arrival and departure of vehicles at the time-space location (t, x); a relaxation term, $\frac{1}{\tau}[u^e(k) - u]$, describing how vehicles adapt their speed to

the conditions dictated by the fundamental diagram, but with respect to a certain reaction time (as opposed to the instantaneous adaptation in the LWR model); and finally an anticipation term, $c_0^2 \frac{1}{k} \frac{\partial k}{\partial x}$, describing how vehicles react to downstream traffic conditions. As a result, vehicles no longer instantaneously change their speed when crossing a shock wave. The higher-order continuum model can be derived from microscopic car-following theory as done by Payne [1971] and is intended to be a response to the apparent shortcomings of the simple order continuum model.

Since its creation, the higher-order continuum model faces extensive debates about the theoretical soundness and practical usefulness. Some researchers show that the model did not outperform the simple order one, especially in heavier traffic [Hauer and Hurdle 1979, Derzko et.al. 1983, Michalopoulos and Beskos, 1984]. Daganzo [1995] argues forcefully that 1) the inference that wave characteristics resulting from the higher-order continuum model may be faster than the mean speed of traffic is a main reason to disqualify the model; 2) the higher order continuum model given certain circumstances could yield unstable behavior, like negative flows and negative speeds ("wrong way travel") although smoothing out all discontinuities in density. However, Papageorgiou [1998] opposes these arguments. He states that macroscopic mean speed is just the average of the speeds of all individual vehicles and hide the variation of the individual speeds, therefore "it is not a physical contradiction if the speed of wave characteristics in Payne's model is higher than the macroscopic mean speed (but consistent with some maximum value of individual car speeds). The possible negative speeds incurred in the higher-order continuum model are not even thought to be a real obstacle to the model validity after applying a very

minor adjustment to the model. On the other hand, some efforts have been made to extend and improve the model to circumvent its limitations. The anticipation term is adjusted by a factor $k/(k+k_0)$ with a positive constant k_0 by Cremer, M., and M. Papageorgiou [1981] to enhance the model accuracy at small density. Comparison between the two models against real data from different sites under various traffic conditions shows higher accuracy of the higher-order continuum model over the first-order model [Cremer and Papageorgiou 1981, Papageorgiou et al. 1983, Papageorgiou et al. 1989]. A viscosity term, $v_0 \frac{\partial^2 u}{\partial x^2}$, is proposed to add to the right-hand side of Equation [2-6] to smear out sharp shock fronts and present realistic transients [Kühne 1984]. The viscosity term shows contribution to the ability of describing traffic from free flow conditions to highly congested region through the mathematical stability analysis on a basis of a truncated expansion [Kühne and Beckschulte 1993]. Michalopoulos et. al. [1992] proposes a new formulation which mainly replaces the equilibrium speed with free flow speed and introduces a friction term to address the effect of ramp flows. The quantitative testing of the revised higher-order model which is implemented based on a finite difference method (upwind scheme with flux vector splitting) is conducted for freeway segments with entrance and exit ramps. The improved model is declared to have a good capability of describing queue propagation and dissipation. Its promising superiority compared to the simple order model has been concluded based on the test results. More recent progress on improving the higher order continuum models can be found in the publications by Liu et. al. [1998] and Zhang [1998, 1999]. Both models are claimed to be consistent with LWR theory in a limit case with higher superiority and remove

the potential deficiencies (negative speeds and unrealistic disturbance propagation speed) of PW theory described by Daganzo [1995]. Comparisons of numerical results with field data are made in the work of Liu et. al. [1998], which concludes that the improved high-order model yields lower error levels than the simple continuum model.

Typical macroscopic simulation models apply the macroscopic traffic flow theory to determine the condition of traffic by dealing with an aggregation of vehicles rather than treating each individual vehicle. Payne's FREFLO is the most well known macroscopic freeway simulation package. Other simulation models include NETFLO (I and II), TRANSYT (by the University of Florida). The lower level of fidelity makes these models appropriate to static planning applications for typically large areas. To satisfy a higher requirement of traffic flow details, another type of traffic simulation models, mesoscopic ones, are emerging. Individual vehicle is tracked and moved according to *macroscopic* traffic flow relations in these simulation models. A few of the leading mesoscopic simulation models are DYNASMART, DynaMIT and CONTRAM. The attractive aspects of computation efficiency and modeling effectiveness of these simulation packages are gaining increasing attention as well as DTA-related applications. In the section 2.3, the DYNASMART simulation structure is overviewed with zoom-in on the central traffic flow modeling.

2.2.3   Mesoscopic traffic flow theories

Mesoscopic traffic flow models are models that have aspects of both macro and microscopic models and describe traffic flow at an intermediate detail level.

19

Traffic flow dynamics are described in aggregated forms but driver behavior is distinguished individually. Mesoscopic modeling is viewed as a consistent link between microscopic and macroscopic modeling. The most well-know mesoscopic flow models are gas-kinetic traffic flow models. Other developed mesoscopic traffic flow models include cluster models and headway distribution models.

A gas-kinetic traffic flow model was initially proposed by Prigogine and Herman [Prigogine 1961, Prigogine and Herman 1971] who suggest an analogy between traffic and gas behavior. In the model, the concept of the speed distribution function is of importance. In the basic kinetic equation, the temporal and spatial dynamic changes of the speed distribution function are mainly caused by two separate processes, namely relaxation process and interaction process. The relaxation process reflects drivers' intention of driving at desired speeds which are dependent on traffic conditions. The interaction process describes the response of faster vehicles to slower vehicles, say, passing or slowing down with certain probabilities. Munjal and Pahl [1969] questioned the validity of the assumption of vehicle chaos in the model, which in fact neglects correlation inside platoon especially apparent at signalized urban roads. Although the criticism, the earliest gas-kinetic traffic flow model seems interesting to researchers and gets continuous attentions and improvements. Paveri-Fontana [1975] considers the desired speed an independent variable instead of a parameter in the kinetic equation. Recently, Nelson [1995] improved the gas-kinetic model by introducing mechanical and correlation models into the kinetic equation and treating the relaxation and interaction processes in the same principle. Hoogendoorn and Bovy [Hoogendoorn 1997, Hoogendoorn and Bovy 1999] innovated gas-kinetic

multi-class traffic flow models which distinguish the distribution functions for each user class and result in rational asymmetric behaviors for the faster user class and the slower user class. Meanwhile, Helbing [1997] developed a gas-kinetic multilane traffic flow model where lanes are treated specifically and additional terms (i.e., velocity diffusion term, lane-changing term, vehicle entering and leaving rate) are incorporated in the kinetic equation. The latter three contributions led to a generalized platoon-based multilane multi-class traffic flow model by Hoogendoorn [1999]. The models developed so far are dominantly used in theoretical research instead of practical applications due to their implemental complexity and lack of appropriate data available.

2.3 *Traffic flow modeling in mesoscopic simulated-based DTA system*

2.3.1 Introduction of DYNASMART

DYNASMART is a state-of-the-art Traffic Estimation and Prediction System (TrEPS) which supports transportation network planning and operations decisions (in the offline version DYNASMART-P) and ATMS/ATIS capabilities in the ITS environment (in the real-time version DYNASMART-X). The model structure of DYNASMART is illustrated in Figure 2-1. Given the network representation, link characteristics as well as control/operation settings, the simulation component will load a time-dependent OD flow matrix and process the movement of vehicles on links, as well as the transfers between links. These transfers require instructions that direct vehicles approaching the downstream node of a link to the desired outgoing link. The user behavior component is the source of these instructions, as it

21

determines individual path decisions of users in the network. Alternatively, path

decisions may be pre-assigned for some or all users according to a particular

assignment scheme, as is the case when DYNASMART is used as a simulator in the

context of algorithmic procedures (e.g. system optimal dynamic traffic assignment).
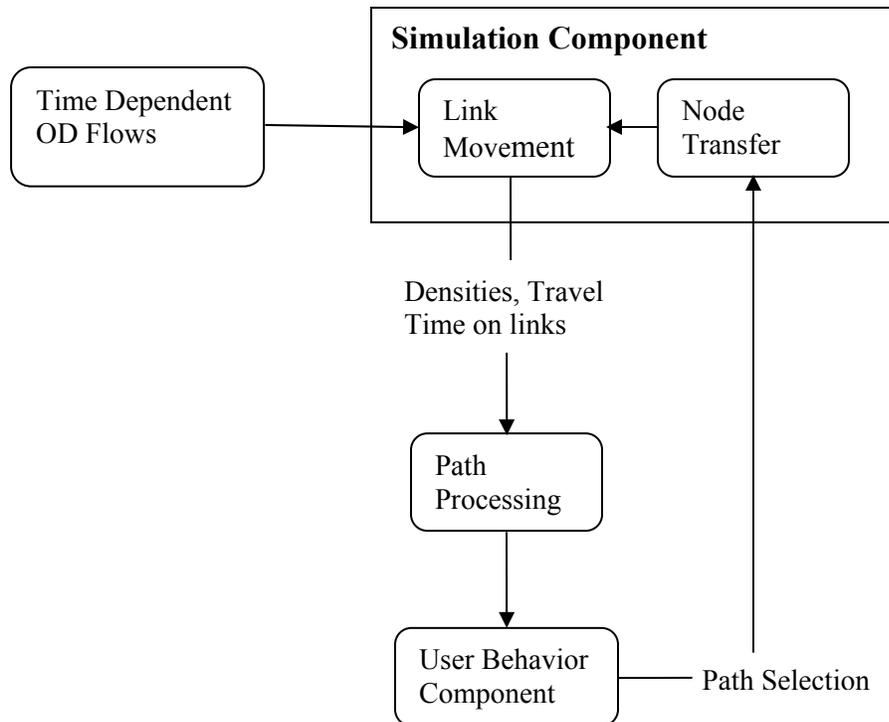
Figure 2-1 Dynasmart Model Structure

The offline version is called DYNASMART-P. It is a dynamic transportation

network design, analysis, planning, evaluation and traffic simulation tool.

DYNASMART-P models the evolution of traffic flows in a traffic network resulting

from the travel decisions of individual drivers. The model is also capable of

representing the travel decisions of drivers seeking to fulfill a chain of activities, at

different locations in a network, over a given planning horizon. The model is an efficient hybrid traffic simulation-assignment approach due to its richer representation of travel behavior decisions, the explicit description of time-varying traffic processes and the explicit representation of traffic network elements (e.g. different link types, signalization and other operational controls). The modeling features chosen for implementation of DYNASMART-P achieve a balance between representation detail, computational efficiency, and input data requirements. DYNASMART-P is carried out in an offline mode whose prime distinction from the online mode (in DYNASMART-X, as described in 2.3.3) is that the OD demand matrix are input externally and are fixed for the analysis period. Also, DYNASMART-P, as an operational planning tool, has no real-time interaction with sensor data collected throughout the network.

In the following sections, the simulation component embedded in DYNASMART is described. Then an introduction is given to the online version DYNASMART-X.

2.3.2 Traffic simulation component in DYNASMART

In DYNASMART, the traffic flow simulation is based on macroparticle simulation concept where the model moves vehicles in discrete macroparticles at the prevailing local speeds determined from the established speed-density relations. The macroparticle concept is adapted from plasma physics (Leboeuf et al. 1979) which exhibits similar properties in this regard. The simulation model is an extension of the macroparticle simulation model (MPSM) (Chang et al., 1985), initially developed as a

special-purpose code for experimental studies of commuter behavior dynamics in congested traffic corridors. In the previous work, 5 to 20 vehicles were used as a macroparticle (Chang et a1. 1985; Mahmassani and Jayakrishnan l988). In its current implementation, DYNASMART uses a macroparticle of *one* vehicle, meaning that it effectively track the movement and location (thereby itinerary) of individual vehicles through a network. However, it does not keep track of the microscopic details of individual traffic maneuvers, such as in car-following models. In this sense, the model is called *mesoscopic* simulation due to the combined aspects of macroscopic relationship and microscopic details.

The traffic simulation in the current DYNASMART uses the equilibrium speed-density relationships in conjunction with the conservation law to describe traffic flow evolution. In general, this approach is practically LWR-type macroscopic traffic flow theory. The continuity equation, expressed in finite difference form, is solved numerically using discrete time steps. Virtually, both average link volume and average link speed are eligible to move the vehicles in the simulation since the identity "volume = density × speed" is hold always. However, for links of finite lengths, moving vehicles according to this identity may lead to physically unrealistic speeds, as discussed in Chang et al. (1985). For this reason, DYNASMART moves vehicles through a network at the prevailing local speeds determined from the equilibrium speed-density relations.

The traffic simulation consists of two primary modules: link movement and node transfer, which are described hereafter in the following sections.

2.3.2.1 Link movement

The link movement module consists of a process for moving vehicles on links during every simulation time step or scanning time interval in the simulation. Note that the network's links are subdivided into smaller sections or segments for traffic simulation purposes. The vehicle concentration prevailing in a section over a simulation time step is determined from the solution of the finite difference form of the above continuity equation, given the concentration as well as inflows and outflows over the previous time-step. Using the current concentration, the corresponding section's speeds are calculated according to a speed-density relation, e.g.,

$$V_i^t = \left(V_f - V_0\right)\left(1 - \frac{K_i^t}{K_j}\right)^{\alpha} + V_0$$

where,

V_i^t , K_i^t = mean speed and concentration in section i during the t-th time step,

V_f, V_0 = mean free speed and minimum speed, respectively,

K_j = jam concentration, and

α = a parameter used to capture the sensitivity of speed to the concentration.

2.3.2.2 Node transfer

The node transfer module performs the link to link or section to section transfer of vehicles at nodes. For interrupted link flow, the node transfer allocates appropriately the right of way according to the control strategy at this intersection. It

25

determines the number of vehicles that are traversing each intersection in the network at each simulation time step as well as the number of vehicles entering and exiting the network. The output of the node transfer includes the number of vehicles that remain in queue and the number added to and subtracted from each link section for each simulation time step. A wide range of traffic control measures for both intersections and freeways are reflected in the outflow and inflow capacity constrains of the node transfer module.

The outflow capacity constraints limit the maximum number of vehicles allowed to leave each approach lane at an intersection. These constraints are described in the following equation which states that the total number of vehicles that enter an intersection (from a given approach) depends on the number of vehicles waiting in the queue at the end of the current simulation interval (time step), AT, and the capacity of this approach. The definition of capacity follows the 1985 Highway Capacity Manual (HCM), and consists of the maximum number of vehicles that can be served under prevailing traffic signal operation.

$$VI_i = \min(VQ_i, VS_i)$$

where,

i: link index;

VI_i: number of vehicles that can enter the intersection from link i during AT;

VQ_i: number of vehicles in queue on link i at the end of AT;

VS_i: maximum number of vehicles can enter the intersection from link i during AT, i.e. $G_i \cdot S_i$;

G_i: remaining effective green time during AT for the movement from link i

(for unsignalized intersection, the calculation is based on the equivalent green time);

S_i: saturation flow rate for the movement from link i; and

AT: the simulation interval.

The inflow capacity constraints determine the maximum number of vehicles allowed to enter a link. These constraints bound the total number of vehicles from all approaches that can be accepted by the receiving link; they include the maximum number of vehicles from all upstream links wishing to enter the receiving link, the available physical space constraint and the section capacity constraint of the receiving link.

$$VO_j = \min\left(\sum_{k \in U} VI_{kj}, VE_j, C_j \Delta T\right)$$

where,

j: link index;

VO_j: number of vehicles that can enter link j

U: set of inbound links into link j (i.e. in the backward star of j)

VI_{kj}: number of vehicles wish that to move from k to j

VE_j: the available space on link j

C_j: the approach capacity of link j

ΔT: duration of a simulation interval

2.3.3 Overview of DYNASMART-X

2.3.3.1 Functionality

With the identical simulation-assignment framework with DYNASMART-P, DYNASMART-X interacts continuously with multiple sources of real-time information, such as loop detectors, roadside sensors, and vehicle probes, which it integrates with its own model-based representation of the network traffic condition. It is designed for Traffic Management Centers (TMCs) to interact with ATMS, ATIS, surveillance systems, incident management systems, and other ITS sub-systems. The system combines advanced network algorithms and models of trip-maker behavior in response to information in an assignment-simulation based framework to provide traffic estimation/prediction and routing guidance.

Besides the most important simulation-based DTA model for traffic state estimation and prediction, consistency checking and updating is a crucial function incorporated in DYNASMART-X to ensure consistency of the simulation-assignment results with actual observations, and to update the estimated state of the system accordingly. Another external supporting function is intended to perform the estimation and prediction of the origin-destination (OD) trip desires that form the load onto the traffic network and are, as such, an essential input to the simulation-assignment core.

2.3.3.2 Description of components and modules

A schematic view of the DYNASMART-X system, as implemented, is shown in Figure 2-2. The arrows represent the data flows between modules and components.
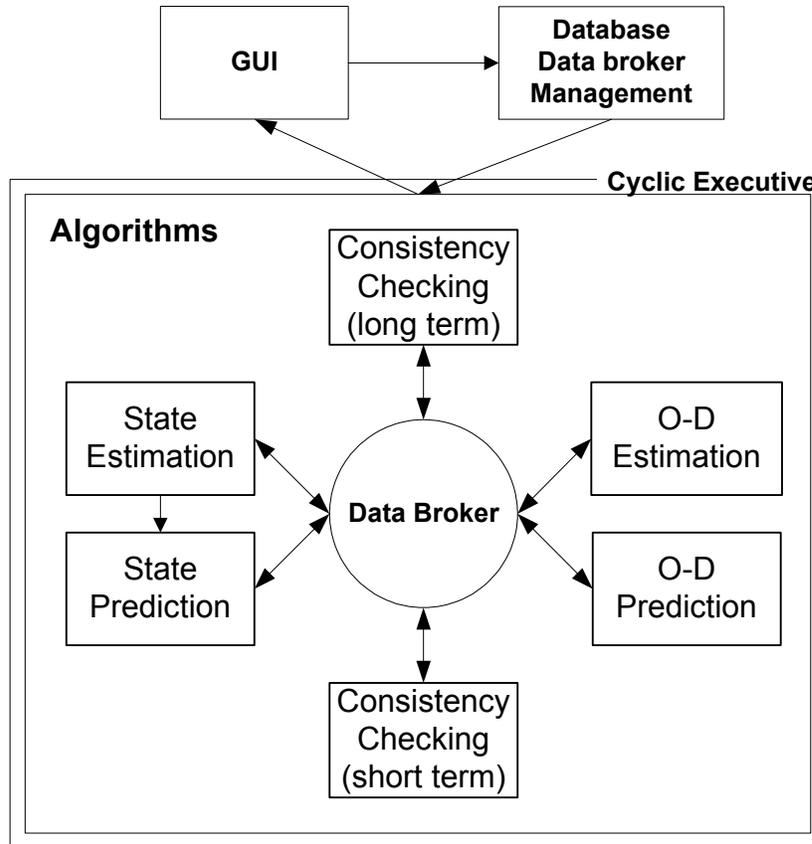
Figure 2-2 DYNASMART-X Schematic View

The algorithmic component is the main entity of the system. It is responsible for implementing various DTA tasks. The purpose of the state estimation module (RTDYNA) is to estimate the prevailing traffic state in the network. The state prediction module (PDYNA), on the other hand, provides future network traffic states

29

for a pre-defined horizon. The OD estimation module (ODE) is responsible for estimating the coefficients of a time-varying polynomial function that describes the OD demand for the current stage. These coefficients are then utilized by the OD prediction module (ODP) to predict the demand - number of vehicles that travel between each OD for every departure time interval in the future stages. Finally, two consistency checking modules (namely LTCC and STCC) are responsible for minimizing the deviation or discrepancy between what is estimated by the system and the real world, in an effort to control error propagation. The module STCC, which stands for short term consistency checking, compares the estimated densities with the observed values, and adjusts the simulated link speeds to minimize discrepancies. Note that this technique of adjusting speeds to minimize density deviations is currently implemented in the prototype system. The technique used currently is advantageous when density or occupancy measurements, instead of speed measurements, tend to be more readily available. Other techniques could be used instead, or in conjunction with the current prototype logic, such as correcting the speed itself, if reliable speed measurements are available. The module LTCC, which stands for long term consistency checking, adjusts the OD demand based on the observed densities. The predicted link proportions are used to match observed densities to their ODs. Based on this matching, correction factors are calculated to update the estimated OD values. These two variants are intended as prototypes that fit within a hierarchical approach to the consistency checking and updating problem. It is the intent to further develop these capabilities in an actual real-time operational setting, because the underlying logic and embedded algorithmic procedures must be

guided by actual traffic system observations. However, the present capabilities have been shown to be effective and provide a good starting point for further elaboration.

The remaining components in the system serve as supporting entities to the algorithmic components. Graphical User Interface (GUI) aims to provide a convenient environment for executing the algorithmic component by allowing users to enter input data, and to view and analyze simulation results "on the fly". Users can see both the current and future network traffic states as generated by the state estimation and state prediction modules, respectively. Traffic statistics are provided at both the link and network levels. Also available are performance plots of the short-term and long-term consistency checking modules. Other features include the ability to view paths, temporal demand patterns, as well as attributes of nodes, links and network.

The intent of the database is to store input and output data. The database primarily stores input data for the state estimation and state prediction modules, as specified by the user through the GUI. The role of the database is two-fold. First, it serves as a gateway through which the DYNASMART-X system communicates with external components (e.g. surveillance, incident detection). Secondly, it allows the system to be responsive to network changes through periodic querying of the database.

In contrast to the database, which is used as a repository for storing long-term data, the data broker provides means for transferring run-time data between modules and components.

Finally, the entire DYNASMART-X is driven by a scheduler, which triggers the execution of each module. Its main duty is to schedule the execution of different processes and perform resource allocation such that all modules satisfy their real-time requirements (soft deadlines are imposed on all modules). Resource allocation refers to the concept of assigning or distributing programs to different computer resources. It is within the scheduler that the data broker is interleaved to manage internal information interchange between modules. That is, the scheduler takes into account data dependency between modules by activating data transfers prior to triggering certain modules. In the current implementation, the scheduling is done in a multi-cycle driven manner. Furthermore, the system is inherently conceived for asynchronous parallel execution. However, all the processes in this version could be executed sequentially besides in real-time. The sequential implementation is intended to accommodate the preferred testing environment requested by the project sponsors.

2.4   *Traffic flow estimation and prediction*

Short term traffic flow forecasting is needed by modern traffic management and control system. Reliable and accurate traffic flow forecasting through continuous contact with advanced surveillance system will support decision-making process on various control strategies and enhance the performance of overall network. Real-time traffic flow estimation and prediction becomes one of the important capabilities for ATMS/ATIS in ITS environment. Many efforts have been focused in this area and various approaches and techniques have been developed to provide traffic forecasts. The works on traffic flow analysis and forecasts take into account the fact that traffic

32

characteristics are stochastic in nature and usually collected as a series of data at regular time intervals from detectors. Time series analysis methods provide a general and systematic approach for capturing the trends in time-ordered data, and representing the variable evolution and even the dynamic interrelationship of system variables. Examples of these models are the smoothing method, the non-parametric regression method, Artificial Neural Network (ANN) model, Kalman filtering model, ARIMA model, and Transfer Function model.

2.4.1 Univariate methods

Smoothing is one of the straightforward and intuitive techniques for combining information contained in a time-series of measurements. The traffic prediction algorithm in the second generation UTCS, UTCS-2 [FHWA, 1973], makes use of smoothed historical traffic data and current traffic measurements from the vehicle detector. The third generation UTCS, UTCS-3 [Lieberman, 1974], relies the prediction on smoothed current traffic measurements only. Stephanedes et al. [1981] provides a more efficient prediction equation compared to the two versions of UTCS. The proposed algorithm predicts the volume during the next time period by linearly combing the current volume, the difference between current volume and previous volume, and the average volume during the previous three, four, or five time periods as the independent variables. The simpler version of the algorithm is obtained by ignoring the first two independent variables, resulting in a simple moving average equation. More recently, Lu [1990] presents an adaptive filtering model to recursively forecast traffic flow on freeway. The state variable in the model is a

smoothing of the past values. The approach uses a simplified least-mean-square algorithm to search for the optimal weights in the smoothing, which results in faster responses. However the solution stability and convergence are impacted by the model structure specified as pointed out by the author.

Recently, a method called non-parametric regression has been considered for traffic flow prediction [Oswald 2000, Smith 2002]. The method searches a collection of historical observations for records similar to the current conditions and uses these to estimate the future state of the system. However, its accuracy is largely dependent on a historical database and does not outperform the stochastic type approach, although it tends to decrease computation time by avoiding data fitting.

Another time-series forecasting technique is Artificial Neural Network (ANN). An ANN is an information processing paradigm that is inspired by the way biological nervous systems (such as the brain) process information. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. The use of ANN computing for transportation applications began recently [Ledoux 1997, Vythoulkas 2000, Yin et al. 2002, Annunziato et al. 2003]. However, the approach is black-box-based and provides little physical insight into the underlying structure of the system [Vythoulkas 2000], which restricts its widespread uses.

Kalman filtering is based on theory proposed by Kalman [1960] and applied in modern filter and control theory. It may be applied to short term stationary or non-stationary stochastic phenomena. Gazis, D.C. and C.H. Knapp [1971] proposed a method for estimating the number of vehicles on a section of a roadway from

measurements at the entrance and exit points of the section. In their algorithm, pre-estimated rough densities are adjusted by means of a "sequential estimator" which is derived based on Kalman filter theory. Okutani and Stephanedes [1984] employed the Kalman filtering theory in dynamic prediction of traffic flow. Their models base volume prediction of a link on data from a number of links and appropriately use the most recent prediction error to improve the prediction. The study revealed that the proposed model outperforms UTCS-2 and is even better if the difference between traffic data on the day under study and that on the same day one week before is used as input. Vythoulkas [2000] investigates the adaptive estimation and prediction accomplished by the Kalman filter technology to address the problem of short term forecasting of traffic condition in urban road networks. One of the procedures proposed in that work is to take the ratio between predicted and historical flow as the state variable to be predicted. In such a way, the similarity of traffic flow patterns from day to day is taken into account and hence the procedure conceptually improves the prediction performance. The approaches are tested using real data and show capability of online traffic flow estimation and prediction. The ratio prediction procedure mentioned provides better results.

The well-known ARIMA (autoregressive integrated moving average) model is a univariate time series method which describes how a single time series variable is related to its own past values. ARIMA forecasting is founded on the stochastic system theory. It was first developed in the late 60's but was systemized by Box and Jenkins in 1976 [Box and Jenkins 1976]. ARIMA processes are nondeterministic with linear state transitions. An ARIMA model forecast is typically a weighted

35

average of past values of the series and the number of lags and weights are identified by auto-correlation function (ACF) as well as partial auto-correlation function (PACF). The earliest application of ARIMA in traffic prediction is the work by Ahmed and Cook [Ahmed and Cook 1979]. Traffic time-series volume and occupancy data were analyzed in their work by ARIMA approach and an ARIMA (0, 1, 3) model was found to be most accurate for freeway system in terms of the mean absolute error and mean square error compared to other simple smoothing approaches. Levin and Tsao [1980] compared the performance of two ARIMA model forms using data from Chicago expressway. The study by Williams et al. [1998] provided recommendations for using ARIMA time-series analysis to generate forecasts for traffic management and control systems.

The above methods take into account only the dynamics of single variable, and lack representation of any underlying structural relations that influence or drive these dynamics. For instance, it is well established, both theoretically and empirically, that changes in prevailing speed are strongly related to changes in traffic density. It would be inefficient to ignore the influence of the dynamic pattern of density when forecasting traffic speeds. In this perspective, transfer function model, the transportation applications of which are reviewed in the following session 2.4.2., is more informative to produce more realistic traffic forecasting.

2.4.2   Transfer function model

Unlike the previous univariate predictions, a more sophisticated way to model time-series data is to describe how a time series variable is related to another time

series variable measured with the same resolution. The transfer function model is an example of this strategy.

Transfer function model analysis means that 1) system output is adjusted according to its past values and a series of current and past input values; 2) system disturbance series are correlated [Box and Jenkins 1976, Pankratz 1991].

Tavana and Mahmassani [2000] first investigated the ability of transfer function models to capture traffic flow dynamics for speed estimation given density measurements on a highway section. The model specification was directly based on the classical speed-density forms in first-order continuum models. In the model, deviation of the speed from an equilibrium value given by a static equilibrium speed-density relation, instead of speed itself, is taken as the output time series. In the experimental test using field data from San Antonio, TX, the model exhibited promising results in terms of robustness and satisfactory transferability in estimating speed for ITS application.

Williams [2001] also proposed multivariate vehicular traffic flow prediction through transfer functions with autoregressive integrated moving average errors (called ARIMAX in the paper). Data from upstream sensors are explicitly modeled as impacting factor in the prediction of traffic flow at the location of interest. The real data test revealed that superiority, as well as higher complexity, of the multivariate prediction over the univariate prediction in ARIMA.

The afore-cited woks are both performed offline, which means no adaptations of parameters are made once they are calibrated. However, the particular function parameters are actually varying with time and space. Therefore, in the context of

intelligent transportation systems, the interest is in developing effective methods for parsing large amounts of real-time traffic data adaptively to improve the efficiency of the DTA system for the purpose of satisfactorily estimating and predicting future traffic conditions in real time. However, like on-line estimation with other types of data, adaptive modeling must reflect a compromise between computational effort and accuracy in its choice of methodology. There is limited experience in the use of real-time data for online traffic flow modeling and forecasting.

Huynh et al. [2002] extended Tavana and Mahmassani's work [2000] by applying the transfer function model in a real-time DTA system, and implemented nonlinear optimization algorithm to enable adaptive estimation and prediction. The experimental results suggested that the adaptive model approach outperforms the non-adaptive model, and confirmed the anticipated advantages of transfer function models in on-line application. However, that study was based purely on simulated synthetic pseudo-real time data, and provided no validation for the transfer model form.

Qin and Mahmassani [2004] extend and formalize that approach for real-time implementation in actual systems, and provide validation with actual traffic data from the Irvine network. The model form is slightly different from the former studies [Tavana and Mahmassani 2000 and Huynh et al. 2002] and is derived from the simplified higher-order continuum models with negligence of spatial anticipation term.

2.5   *<u>Summary</u>*

In this chapter, the relevant background concerning traffic flow modeling with real time data is reviewed. Traffic flow theories developed over decades play a fundamental role in the study conducted here. Three types of the theories, microscopic, macroscopic, and mesoscopic traffic flow theories, are overviewed. Since this dissertation research is mainly based on macroscopic continuum traffic flow theory, macroscopic models are discussed in more details. Then the simulation-based dynamic traffic assignment tool DYNASMART-X is briefly introduced and traffic flow modeling in the simulation part is described. Finally, the previous studies and practices regarding traffic flow estimation and prediction through real-time sampling traffic data are reviewed. Among those approaches, the transfer function methods are viewed as the ones with higher potency to model traffic flow dynamic by incorporating other driving factors for the dynamics of the variable of interest.

Chapter 3: Dynamic Speed-density Relation Formulation

3.1   *Introduction*

Performance of a traffic simulator depends to a great extent on the embedded traffic flow model. As a new generation of traffic simulation tools oriented to online traffic management applications, the real-time simulation-based dynamic traffic assignment model needs a reliable traffic flow representation that can rapidly adapt to traffic dynamics observed quasi-continuously from surveillance systems.

Static speed-density relations capture essential traffic stream phenomena. Calibration and parameter estimation of these models for specific facilities are typically performed by applying standard regression techniques to field traffic data. An implicit assumption made in these approaches is that traffic measurements taken over consecutive intervals are independent, and that the resulting static relation holds over a wide range of traffic conditions, without considering potential dynamic effects in the relation. However, traffic flow realizations generally occur in a time sequence over which independence assumptions may not hold. The possibility of a time-lagged relation between speed and density, as well as the possible presence of a temporal pattern of the system noise, are ignored in conventional static models and approaches.

Although the stationarity assumed is a deficiency, the conventional models still form the basis for representing traffic flow in networks in virtually all existing practice and past research. Especially when online modeling is not a requirement and real-time data is not available, the conventional static models perform quite well in terms of reflecting the key traffic stream phenomena. For instance, as mentioned in

section 2.3.2, the traffic flow simulation in DYNASMART is based on mesoscopic simulation concept where the model moves individual vehicles at the prevailing local speeds determined by Greenshields' model from the established speed-density relations under the condition of continuity rule. Traffic flows are represented rationally and efficiently by this simple order continuum based model for the offline DTA operation [University of Maryland DTA Group, 2003] which long term planning could rely on.

However, the models have two limitations when applied to online traffic estimation and prediction. First, the parameters in the static speed-density relation are set as fixed and invariant with respect to time. The adaptive calibration of the static relation using real-time traffic data is not wise because data during any particular time period is not guaranteed to reflect a full data regime to identify an accurate and reliable model form. As such, the model cannot be properly adaptive to the traffic dynamics, which is an essential requirement for real-time traffic management systems. Second, the equilibrium states implied in this model are rarely observed in practice, especially when the time period over which the traffic stream variables are averaged is short (< 1min). In fact, the sampling rate of loop detectors is usually 20-sec or 30-sec.

A traffic flow model which is oriented to online application needs to be amenable to calibration under any traffic condition. In addition, a robust model should be able to reproduce or approach as much as possible the behavior of real traffic. As mentioned earlier, speed is adapted after a certain time delay (time-lagged correlation) and reflects traffic conditions downstream (space-lagged correlation).

41

The higher order continuum model (PW model) uses a momentum equation to describe such speed evolution over time and space.

There are numerous studies on the calibration and application of the higher order continuum model. However, no conclusive result has been reached yet since the coefficients tend to be time-variant and are related to complicated user behaviors. Besides, the temporal and spatial discretisation scheme is also an important tuning factor for the implementation of higher order continuum models. In this dissertation, as a variant and extension of the higher order continuum model, a dynamic speed-density relation is proposed by exploring the capability of the transfer function (TF) method to identify and adaptively calibrate the relation, as well as perform forecasting. The use of transfer function models, based on time series theory, can capture the manner in which system output is related to its past values, and respond to current and past values of system inputs.

In the following sections of this chapter, we will at first introduce the transfer function and discuss why it becomes a powerful method to model and adaptively calibrate dynamic relation with actual data. Then, the derivation of the formulation for the dynamic speed-density relation will be addressed. The model calibration algorithm and the forecasting logic will be described. At last, the rolling horizon scheme is presented as an application of adaptive calibration and forecasting to real time operation.

3.2   Review of transfer function method

Study of dynamic response is a topic of considerable interest in engineering, economics and many other fields. Suppose that in a system, the level of one variable is influenced by the level of another variable, which are called system output and system input, respectively. The response by the system output is usually not immediate but delayed when the system input changes from one level to another. Such a change is referred as dynamic response. It is only when the dynamic characteristics of a system are understood that more reliable forecasting and better system control are possible. A model that describes the dynamic response occurring in a dynamic input-output system is called a transfer function model. Transfer function model should be distinguished with the regular regression model which answer the question of "how is y related to x?". Instead, transfer function deals with 1) lags in the variables to answer the question "how is y(t) related to x(t-k), for various k?"; and 2) autocorrelation in the residuals. In fact, the ordinary regression can be viewed as a particular subset of a transfer function with certain assumptions on the internal structures. As mentioned above, the assumption made in the regression model is not valid in typical time-series traffic data and transfer function shall be a more general methodology to identify the complexity presented in the traffic flow dynamics.

In practice, the system input and output are measured at equispaced intervals of time. Hence, the widely-used transfer function model is in a discrete form. It is often the case that the response of the system output is affected by uncontrollable disturbance (or, noise) other than the system input. To take account not only of the

dynamic relationship associating the input and the output but also of the noise infecting the system, combination of a deterministic transfer function model with a stochastic noise model is necessary.

A general transfer function model of order (r, s) with a stochastic noise model of order (p, q) is given by

$$Y_t = \delta^{-1}(B)\omega(B)X_t + N_t \qquad\qquad [3\text{-}1a]$$

or $\quad Y_t = v(B)X_t + N_t \qquad\qquad [3\text{-}1b]$

where

Y_t is the system output at time interval t.

X_t is the system input at time interval t.

B is the backward shift operator, defined as $B^n Y_t = Y_{t-n}$.

$\delta(B) = 1 - \delta_1 B - \delta_2 B^2 - \cdots - \delta_r B^r$, δ_i $(i = 1,...r)$ are a series of parameters. The estimated value of δ_i could reveal how the current output is related to its past values.

$\omega(B) = \omega_0 - \omega_1 B - \omega_2 B^2 - \cdots - \omega_s B^s$, ω_i $(i = 0,...s)$ are a series of parameters. The estimated value of ω_i could reveal how the current output is related to the current and past values of the input.

$v(B) = v_0 + v_1 B + v_2 B^2 + ...$, or $v(B) = \delta^{-1}(B)\omega(B)$ is called the transfer function (or impulse response function), represented by the ratio of two polynomials $\omega(B)$ and $\delta(B)$. It can be viewed as a linear filter linking the input X_t and the output Y_t.

N_t is the stochastic noise at time interval t which is independent of the input series X_t. N_t is often an autocorrelated sequence and can be specified in an ARIMA model of order (p, q).

$$N_t = \varphi^{-1}(B)\theta(B)a_t \qquad\qquad [3\text{-}2a]$$

$$\text{or} \quad N_t = \psi(B)a_t \qquad\qquad [3\text{-}2b]$$

where

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \ldots - \varphi_p B^p, \ \varphi_i \ (i = 1,\ldots p) \text{ are a series of parameters.}$$

$\varphi(B)$ represents an internal autoregressive process.

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q, \ \theta_i \ (i = 1,\ldots q) \text{ are a series of parameters.}$$

$\theta(B)$ represents an internal moving average process.

$$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \ldots, \text{ or } \psi(B) = \varphi^{-1}(B)\theta(B) \text{ is represented by the ratio}$$

of two polynomials $\theta(B)$ and $\varphi(B)$. It can be viewed as a linear filter linking the white noise a_t and the colored noise N_t.

a_t is white noise, which is a uncorrected random series with mean zero and stationary variance.

Equivalently, Equation [3-1] or [3-2] can be represented as

$$Y_t = \delta^{-1}(B)\omega(B)X_t + \varphi^{-1}(B)\theta(B)a_t \qquad\qquad [3\text{-}3a]$$

$$\text{or} \quad Y_t = v(B)X_t + \psi(B)a_t \qquad\qquad [3\text{-}3b]$$

The transfer function framework takes into account the possible time-lagged relationship between output and input, as well as the possible time series pattern of the system noise. Figure 3-1 indicates the structure of the single-input transfer

function model with noise model. In fact, it is often the case that the dynamic system to be studied has more than one driving force, so a multiple-input transfer function model is not more complicated but has an additive form as

$$Y_t = \sum_i \delta_i^{-1}(B)\omega_i(B)X_{i,t} + \varphi^{-1}(B)\theta(B)a_t \qquad \text{[3-4a]}$$

$$\text{or} \quad Y_t = \sum_i v_i(B)X_{i,t} + \psi(B)a_t \qquad \text{[3-4b]}$$

White noise a_t → | Linear Filter $\psi(B) = \varphi^{-1}(B)\theta(B)$ | → Noise $N_t = \psi(B)a_t$

Input X_t → | Linear Dynamic System $v(B) = \delta^{-1}(B)\omega(B)$ | → $v(B)X_t$ Output Y_t
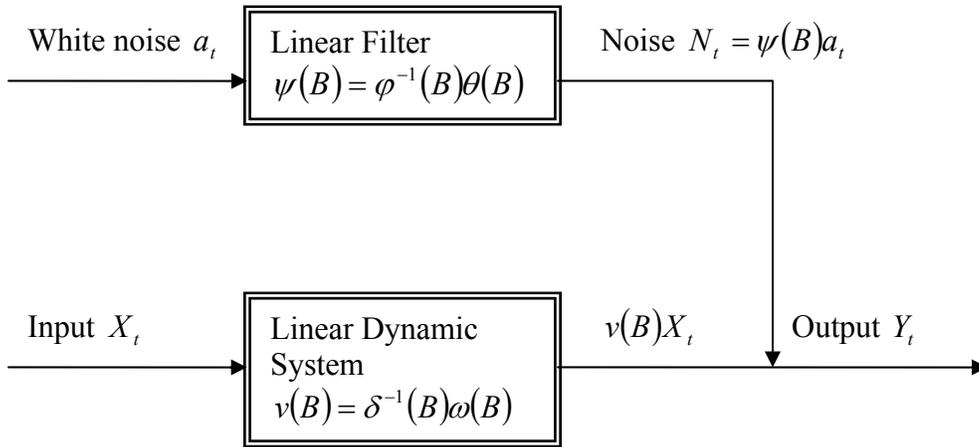
Figure 3-1 Transfer Function Model for Dynamic System with Noise

3.3   *Dynamic speed-density relation*

3.3.1   Determination of system input and output

The application of transfer function approaches in dynamic transportation network modeling brings a new perspective to the specification, calibration and application of speed-density relations. The conventional transfer function model deals with system output that has a linear transferring relation with the system input. Actual speed and density measurements generally exhibit a non-linear relation,

46

reflected in most functional forms considered in the literature [May 1990]. To handle nonlinear transfer function forms, Box et al. [1976] proposed a linearized approximation. But linearization of those traffic stream relations is too unwieldy to use in on-line applications. In this dissertation, determination of system input and output is inspired by the macroscopic higher-order continuum model.

Observations of traffic show that the average speed in a section is not only dependent on the local traffic state in that section but also on the state of the neighboring sections. Three major mechanisms that influence the average speed can be distinguished as: relaxation, convection and anticipation, which are presented in the higher order continuum model (Equation [3-5]):

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \frac{1}{\tau}[u^e(k) - u] - c_0^2 \frac{1}{k}\frac{\partial k}{\partial x} \qquad [3\text{-}5]$$

where

u and k denote the speed $u(x,t)$ and density $k(x,t)$ at location x and at time t, for convenience;

$u^e(k)$ represents the equilibrium speed determined by the equilibrium speed-density relation (e.g. Greenshields model);

τ is the reaction time of a driver with respect to speed change; and

c_0^2 is an anticipation coefficient.

The first term on the right-hand-side of Equation [3-5] is generally interpreted as a relaxation term (of the prevailing speed to the equilibrium value); the second term is an anticipation term (e.g. of density changes ahead); and the second term of the left-hand-side of Equation [3-5] is a convection term. Extensive discussion and

interpretation of the various terms in Equation [3-5] is found in the literature [Kuhne and Michalopoulos, 1998] and a brief review is given in Chapter 2 of the dissertation.

The canonical discretised form of the higher order continuum model is shown in Equation [3-6].

$$u_{t+1}^i - u_t^i = \frac{\Delta t}{\tau}\left[u^e\!\left(k_t^i\right)-u_t^i\right]+\frac{\Delta t}{l_i}\left[u_t^i\!\left(u_t^{i-1}-u_t^i\right)\right]-\frac{\Delta t c_0^2}{l_i}\left(\frac{k_t^{i+1}-k_t^i}{k_t^i}\right) \qquad [3\text{-}6]$$

where i is for geographic section; t is for time interval.

The speed adjustment from one time interval to the next time interval is contributed by the factors of relaxation, convection and anticipation.

Relaxation is proportional to the difference between the actual average speed and the desired average speed (equilibrium speed), $u^e\!\left(k_t^i\right)-u_t^i$. The desired speed is dependent on the prevailing density k_t^i. By observing the density, the drivers tend to accelerate or decelerate towards the desired speed. The larger the difference, the greater speed adaptation will be. The impact of the relaxation is also dependent on the drivers' reaction time which is expressed as τ in the formulation. The shorter the reaction time, the faster the drivers will respond to the speed deviation, thus the greater impact of the relaxation on the speed change, $u_{t+1}^i - u_t^i$.

Convection is proportional to: 1) the speed difference between the two sections i-1 and i, $u_t^{i-1}-u_t^i$. The higher the speed difference, the longer time the vehicles in section i-1 will need to accelerate or decelerate while entering the section i and the greater the impact on the speed adjustment in the section i; 2) the average speed in the section i, u_t^i. The higher the speed, the longer time the vehicles will need to adapt their speed thus the greater the impact on the speed adjustment in the

section i. Convection is also reversely proportional to the length of the section l_i.

The longer the section, the more vehicles drive at their desired speeds thus the lower

impact on the average speed in the section i.

Anticipation is affected by the relative difference of the density in section i+1

and i, $\frac{k_t^{i+1} - k_t^i}{k_t^i}$. If drivers observe more congested traffic ahead, they tent to slow

down; oppositely, if the traffic density ahead is lower, drivers will accelerate.

Anticipation is proportional to the coefficient c_0^2 which has the meaning of the

standard deviation of the vehicular speed distribution. The higher speed variance, the

greater impact of the density difference on the speed adjustment in section i. Besides,

anticipation is reversely proportional to the length of the section l_i. The longer the

section, the more vehicles drive at their desired speeds thus the lower impact on the

average speed in the section i.

Therefore, as the transfer function method is explored to model dynamic

traffic flow, one output and three inputs are identified by the awareness of the

potential dynamic interrelationship mentioned above. The following are the

specifications of the system output Y_t and the system inputs $X_{1,t}$, $X_{2,t}$, and $X_{3,t}$.

· System output: $Y_t = u_{t+1}^i - u_t^i$; [3-7]

· System input 1: $X_{1,t} = u^e\left(k_t^i\right) - u_t^i$; [3-8]

· System input 2: $X_{2,t} = u_t^i\left(u_t^{i-1} - u_t^i\right)$; [3-9]

· System input 3: $X_{3,t} = \frac{k_t^{i+1} - k_t^i}{k_t^i}$. [3-10]

where

$u_{t+1}^i - u_t^i$ is the time-differenced speed between the time t+1 and t at the

section i;

$u^e(k_t^i) - u_t^i$ is the speed relaxation from the equilibrium at the time t and the

section i;

$u_t^i(u_t^{i-1} - u_t^i)$ is the convection term which is the product of the average speed

at the time t and the section i and the differenced speed between the section i-1 and i

at the time t;

$\dfrac{k_t^{i+1} - k_t^i}{k_t^i}$ is the anticipation term which is the ratio of the differenced density

between the section i+1 and i at the time t and the density at the time t and the section

i.

3.3.2   Model formulation

Following the determination of the system inputs and output, the proposed

dynamic speed-density model based on a transfer function model form can be

expressed as follows.

$$Y_t = \delta_1^{-1}(B)\omega_1(B)\cdot X_{1,t} + \delta_2^{-1}(B)\omega_2(B)\cdot X_{2,t}$$

$$+ \delta_3^{-1}(B)\omega_3(B)\cdot X_{3,t} + N_t \qquad\qquad \text{[3-11]}$$

where

$\delta_i(B)$ and $\omega_i(B)$ are polynomials functions for the ith input series. The detailed definitions are as follows, which are similar to the ones described earlier in section 3.2.

$$\delta_i(B) = 1 - \delta_{i1}B - \delta_{i2}B^2 - \cdots - \delta_{ir}B^{r_i}, \quad \delta_{ij} \quad (j = 1,...r_i) \quad \text{are a series of}$$

parameters for the ith input series $X_{i,t}$ (i = 1, 2, 3).

$$\omega_i(B) = \omega_{i0} - \omega_{i1}B - \omega_{i2}B^2 - \cdots - \omega_{is}B^s, \quad \omega_{ij} \quad (j = 0,...s_i) \text{ are another series of}$$

parameters for the ith input series $X_{i,t}$ (i = 1, 2, 3).

N_t is a noise series that is independent on the inputs. Generally, N_t is not a white noise, but an auto-correlated time series which follows a certain ARIMA model of order (p, q), say $N_t = \varphi^{-1}(B)\theta(B)a_t$, where $\varphi(B)$ and $\theta(B)$ are autoregressive and moving average parameters, and a_t is white noise, as specified in Equation [3-2].

Hence, the formulation in [3-11] can also be expressed as

$$Y_t = \delta_1^{-1}(B)\omega_1(B) \cdot X_{1,t} + \delta_2^{-1}(B)\omega_2(B) \cdot X_{2,t}$$

$$+ \delta_3^{-1}(B)\omega_3(B) \cdot X_{3,t} + \varphi^{-1}(B)\theta(B)a_t \qquad\qquad [3\text{-}12]$$

If Equation [3-7] to [3-10] are substituted into Equation [3-12], a dynamic speed-density relation model can be formulated as

$$u_{t+1}^i = u_t^i + \delta_1^{-1}(B)\omega_1(B) \cdot \left[u^e(k_t^i) - u_t^i\right] + \delta_2^{-1}(B)\omega_2(B) \cdot \left[u_t^i(u_t^{i-1} - u_t^i)\right]$$

$$+ \delta_3^{-1}(B)\omega_3(B) \cdot \left(\frac{k_t^{i+1} - k_t^i}{k_t^i}\right) + \varphi^{-1}(B)\theta(B)a_t \qquad\qquad [3\text{-}13]$$

The three key influential elements in the higher order continuum model, namely relaxation, convection and anticipation, are explicitly incorporated into the

dynamic speed-density relation model [3-13] as the system inputs. So, the theoretical basis for the dynamic model is not a new or different one, but is that which has been derived from a physical point of view for decades. However, it should be careful to compare the higher order continuum model and the new dynamic model. In Equation [3-13], more than one time interval is under consideration since the coefficients are basically sequences of weights applied to the past values of the input/output series. How many past values to be considered would be determined by the order r_i and s_i (i = 1, 2, 3). This structure essentially allows flexibility on the time interval of the data sampling because the model captures the various impacts of what happened in the past on the current value of the interest variable. If the data sampling interval is short, the order r_i and s_i might be high; and vice versa.

3.3.3   Model estimation approach

Considering the dynamic speed-density relation in Equation [3-12], we can have the following mathematical transformation:

$$Y_t = \delta_1^{-1}(B)\omega_1(B) \cdot X_{1,t} + \delta_2^{-1}(B)\omega_2(B) \cdot X_{2,t} + \delta_3^{-1}(B)\omega_3(B) \cdot X_{3,t} + \varphi^{-1}(B)\theta(B)a_t$$

$\Rightarrow$

$$\delta_1(B)\delta_2(B)\delta_3(B)\varphi(B)Y_t = \delta_2(B)\delta_3(B)\varphi(B)\omega_1(B) \cdot X_{1,t} + \delta_1(B)\delta_3(B)\varphi(B)\omega_2(B) \cdot X_{2,t}$$

$$+ \delta_1(B)\delta_2(B)\varphi(B)\omega_3(B) \cdot X_{3,t} + \delta_1(B)\delta_2(B)\delta_3(B)\theta(B)a_t \qquad [3\text{-}14]$$

Equation [3-14] can be further simplified as

$$[1 - \Omega_0(B) \cdot B]Y_t = \Omega_1(B) \cdot X_{1,t} + \Omega_2(B) \cdot X_{2,t} + \Omega_3(B) \cdot X_{3,t} + \Theta(B)a_t \quad [3\text{-}15a]$$

or $Y_t = \Omega_0(B) \cdot Y_{t-1} + \Omega_1(B) \cdot X_{1,t} + \Omega_2(B) \cdot X_{2,t} + \Omega_3(B) \cdot X_{3,t} + \Theta(B)a_t \quad [3\text{-}15b]$

if the following definitions are given.

$$\cdot \quad \Omega_0(B) = \left[1 - \delta_1(B)\delta_2(B)\delta_3(B)\varphi(B)\right] \cdot F \qquad\qquad \text{[3-16]}$$

(F is the forward shift operator, defined as $F^n Y_t = Y_{t+n}$. It performs

the inverse operation of the backward shift operator B.)

$$\cdot \quad \Omega_1(B) = \delta_2(B)\delta_3(B)\varphi(B)\omega_1(B) \qquad\qquad \text{[3-17]}$$

$$\cdot \quad \Omega_2(B) = \delta_1(B)\delta_3(B)\varphi(B)\omega_2(B) \qquad\qquad \text{[3-18]}$$

$$\cdot \quad \Omega_3(B) = \delta_1(B)\delta_2(B)\varphi(B)\omega_3(B) \qquad\qquad \text{[3-19]}$$

$$\cdot \quad \Theta(B) = \delta_1(B)\delta_2(B)\delta_3(B)\theta(B) \qquad\qquad \text{[3-20]}$$

Substitute the definitions in [3-7]-[3-10], the following dynamic model is

derived.

$$u_{t+1}^i = u_t^i + \Omega_0(B) \cdot \left[u_t^i - u_{t-1}^i\right] + \Omega_1(B) \cdot \left[u^e\left(k_t^i\right) - u_t^i\right]$$

$$+ \Omega_2(B) \cdot \left[u_t^i\left(u_t^{i-1} - u_t^i\right)\right] + \Omega_3(B) \cdot \left[\frac{k_t^{i+1} - k_t^i}{k_t^i}\right] + \Theta(B)a_t \qquad \text{[3-21]}$$

By performing the simple polynomial function operations, we can see that,

$$\Omega_i(B) = \Omega_{i0} + \Omega_{i1}B + \Omega_{i2}B^2 + \cdots + \Omega_{iM_i}B^{M_i}, \text{ for i} = 0, 1, 2, 3 \quad \text{[3-22]}$$

and

$$\Theta(B) = 1 - \Theta_1 B - \Theta_2 B^2 - \cdots - \Theta_L B^L \qquad\qquad \text{[3-23]}$$

where

Ω_{ij} is the coefficients in the polynomial function $\Omega_i(B)$ (i = 0, 1, 2, 3 and j

= 0, 1, …, M_i);

M_i is the order for the $\Omega_i(B)$ (i = 0, 1, 2, 3);

Θ_j is the coefficients in the polynomial function $\Theta(B)$ (j = 1, 2, ..., L).

L is the order for the $\Theta(B)$;

$\Omega_i(B)$'s in Equation [3-15b] are similar to the transfer function (or impulse response function) $v_i(B)$'s in Equation [3-4b] but they actually have a different definition. $v_i(B)$ is the ratio of two polynomials which are used to describe the deterministic transfer function model, however $\Omega_i(B)$ is the product of several polynomials which are used to describe the transfer function model with the noise model. In addition, the noise term $\Theta(B)a_t$ becomes a pure moving average process since the autoregressive process has been already absorbed into $\Omega_i(B)$'s. The aim of converting Equation [3-12] to Equation [3-15] is to facilitate the model estimation as we can see hereafter.

The best known estimation approach for the transfer function model is due to Box and Jenkins [1976]. Their approach is a procedure for the one-input situation. The sample cross-correlation function between the prewhitened input series and the corresponding filtered output series is studied and used to estimate the transfer function model. Although the method is very capable of solving the one-input situation, it is difficult to generalize it for multiple-input models. Other approaches such as the ones proposed by Priestley [1971] and Haugh and Box [1977] are also appropriate for one-input case. However, the popularity of these methods is limited due to the cumbersome-to-derive structure. As an alternative to time-domain analysis, a frequency-domain approach (spectral methods) has been used to estimate the transfer function models, e.g., Box and Jenkins [1976] and Priestley [1971].

However, it is rather difficult to apply spectral methods in practice. Liu and Hanssens [1982] proposed a procedure which mainly applies linear least-squares estimation on the original or filtered series for transfer function identification. The method performs well according to the test result described in the work. The most important benefit of the method is to handle the multiple-input transfer function model through a simple and straightforward procedure. In this dissertation, the method based on least-squares is explored to perform model estimation for the dynamic speed-density relation. It should be emphasized that although the underlying computing algorithm is same in Liu and Hanssens's paper and this dissertation, the parameters to be estimated are different. In Liu and Hanssens' work, the impulse response function weights are estimated and then continued to be used to identify the lag distribution for the transfer function. Instead of struggling with lag determination for which there is no fast and automatic way, the approach in the dissertation is trying to estimate the values of the parameters in the model directly assuming that a maximum lag order exists. By this way, the subjective and tedious selection of the lag distribution can be avoided. The possible errors resulting from the assumption are expected to be alleviated via the adaptive estimation mechanism (described in the section 3.3.5). The experimental results which are to be presented in the Chapter 4 show the promising performance of the proposed estimation procedure.

The procedure of estimating the parameters in the dynamic model [3-15b] is discussed below.

For each input, select a sufficiently large M_i's. [3-15b] can be expressed as

$$Y_t = \left(\Omega_{00} + \Omega_{01}B + \Omega_{02}B^2 + \cdots + \Omega_{0M_0}B^{M_0}\right) \cdot Y_{t-1}$$

$$+ \sum_{i=1}^{3} \left(\Omega_{i0} + \Omega_{i1} B + \Omega_{i2} B^2 + \cdots + \Omega_{iM_i} B^{M_i} \right) \cdot X_{i,t} + \Theta(B) a_t$$

$$\text{for } t = 1, \ldots, N \tag{3-24}$$

Using

$$M = \mathrm{Max}(M_0, M_1, M_2, M_3)$$

$$L = N - M$$

$$\underline{\beta} = \left[\Omega_{00} \; \Omega_{01} \; \cdots \; \Omega_{0M_0} \; \Omega_{10} \; \Omega_{11} \; \cdots \; \Omega_{1M_1} \; \Omega_{20} \; \Omega_{21} \; \cdots \; \Omega_{2M_2} \; \Omega_{30} \; \Omega_{31} \; \cdots \; \Omega_{3M_3} \right]^T$$

$$\underline{Y} = \left[Y_{M+1}, Y_{M+2}, \cdots, Y_{M+L} \right]$$

$$\underline{X} = \left[\underline{X}_0^0, \underline{X}_0^1, \cdots \underline{X}_0^{M_0}, \underline{X}_1^0, \underline{X}_1^1, \cdots, \underline{X}_1^{M_1}, \underline{X}_2^0, \underline{X}_2^1, \cdots, \underline{X}_2^{M_2}, \underline{X}_3^0, \underline{X}_3^1, \cdots, \underline{X}_3^{M_3} \right]$$

where $\underline{X}_i^j = B^j \underline{X}_i^0$ and $\underline{X}_i^0 = \left[X_{i(M+1)} \; X_{i(M+2)} \; \cdots \; X_{i(M+L)} \right]^T$

$$\underline{\varepsilon} = \left[n_{M+1}, n_{M+2}, \cdots, n_{M+L} \right] \quad \text{where } n_t = \Theta(B) a_t$$

Equation [3-24] can be expressed as an ordinary least-squares problem.

$$\underline{Y} = \underline{\beta} \cdot \underline{X} + \underline{\varepsilon} \tag{3-25}$$

It is known that the solution for the ordinary least-squares estimates of $\underline{\beta}$ is

$$\hat{\underline{\beta}} = \left(\underline{X}^T \underline{X} \right)^{-1} \underline{X}^T \underline{Y} \tag{3-26}$$

It should be pointed out that the noise series $\underline{\varepsilon}$ is not white noise in most cases, therefore the ordinary least-squares estimates of $\underline{\beta}$ is in fact not efficient. Thus, estimation of $\underline{\beta}$ can be improved by using generalized least squares method. If the covariance matrix $\underline{\Gamma}$ of $\underline{\varepsilon}$ is known, the generalized least squares estimator of $\underline{\beta}$ which is consistent and efficient is

$$\hat{\underline{\beta}} = \left(\underline{X}^T \underline{\Gamma}^{-1} \underline{X} \right)^{-1} \underline{X}^T \underline{\Gamma}^{-1} \underline{Y} \qquad\qquad\qquad\qquad \text{[3-27]}$$

It is complicated to compute $\hat{\underline{\beta}}$ directly through [3-27]. Instead, $\underline{X}$ and $\underline{Y}$ are first transformed into $\widetilde{\underline{X}}$ and $\widetilde{\underline{Y}}$ by

$$\widetilde{\underline{X}} = \underline{H}\,\underline{X} \qquad\qquad\qquad\qquad \text{[3-28]}$$

$$\widetilde{\underline{Y}} = \underline{H}\,\underline{Y} \qquad\qquad\qquad\qquad \text{[3-29]}$$

where $\underline{H}$ is the Cholesky decomposition of $\underline{\Gamma}^{-1}$, say $\underline{\Gamma}^{-1} = \underline{H}^T \underline{H}$, thus is $\underline{H}$ an upper triangular matrix.

Therefore, we obtain $\hat{\underline{\beta}}$ by using the ordinary least square estimation on the transformed series $\widetilde{\underline{X}}$ and $\widetilde{\underline{Y}}$.

$$\hat{\underline{\beta}} = \left(\widetilde{\underline{X}}^T \widetilde{\underline{X}} \right)^{-1} \widetilde{\underline{X}}^T \widetilde{\underline{Y}} \qquad\qquad\qquad\qquad \text{[3-30]}$$

The transformations in [3-28] and [3-29] are approximately equivalent to filtering the input and output series by the ARIMA model of the noise series (Ljung and Box 1979). For instance, in [3-25], using the transformations $\widetilde{\underline{Y}} = \Theta^{-1}(B)\underline{Y}$ and $\widetilde{\underline{X}} = \Theta^{-1}(B)\underline{X}$ results in the following equation which can be viewed as a set of ordinary regression equations.

$$\widetilde{\underline{Y}} = \underline{\beta} \cdot \widetilde{\underline{X}} + \underline{a} \qquad\qquad\qquad\qquad \text{[3-31]}$$

where $\underline{a}$ is the vector form of the white noise series a_t.

However, the noise model is in fact unknown but can be identified from ARIMA analysis on the OLS residuals (Liu and Hassens 1982). An iterative

procedure of filtering and the ordinary least square estimation could then be set up

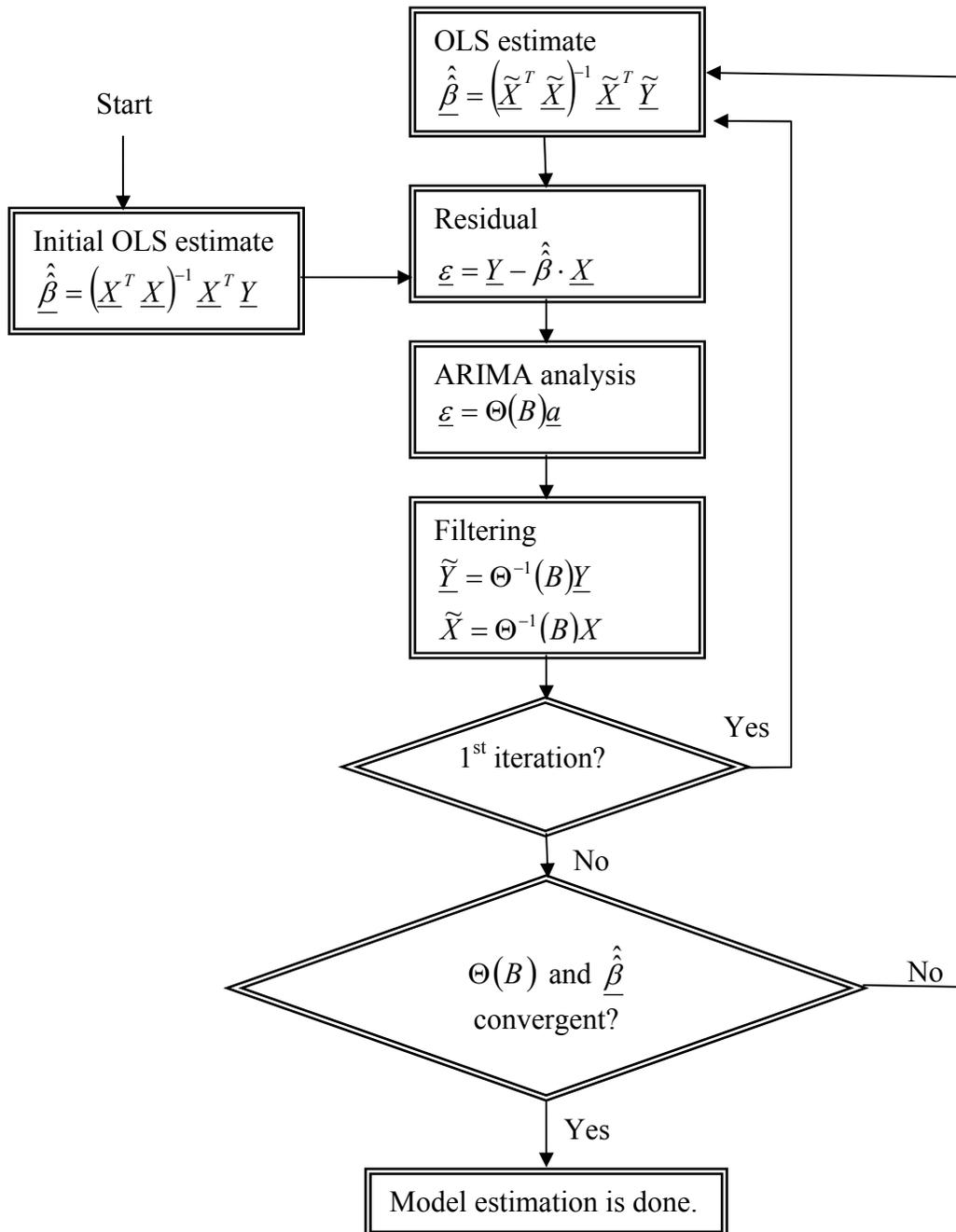until the estimates of $\hat{\underline{\beta}}$ and the ARIMA model converge (Figure 3-2).

Figure 3-2 Iterative Procedure of Filtering and OLS Estimation

Two issues related to the model estimation should be considered. First, how to determine the value of M_i for each input i? Secondly, how to deal with the non-stationary data?

The first problem is actually to choose most significant past time intervals at which the exogenous variables occur. In other words, one group of the explanatory variables of Y_t is X_t, X_{t-1}, X_{t-2},...... X_{t-g},...... It is reasonable to assume that the larger the lag g, the smaller the influence of X_{t-g} on Y_t. There exists a maximum lag, M_i, beyond that the influence can be neglected. The magnitude of M_i is quite dependent on the length of time interval on which the time series is sampled. The longer the time interval, the smaller M_i could be; vice versa. In practice, we can set a sufficiently large value to M_i and continue with the ordinary least square estimation. The t-statistics for each estimates of $\hat{\underline{\beta}}$ could help to rule out the insignificant variables, which basically is to reduce the size of M_i. As we can see from the notation for Equation [3-25], smaller M_i means less coefficients to be estimated and the model becomes more parsimonious which is a desirable property. With newly updated equations, another ordinary least square estimation can be performed. If it is necessary, an iterative procedure could be used to obtain the convergence of the values of M_i's.

The second problem is usually resolved by filtering the time series data using difference method. A stochastic stationary process, which has a constant mean level

and finite variance over time, can make the transfer function model well-conditioned and with controllable forecasting. A non-stationary process can transform to a stationary process by the dth difference [Box and Jenkins 1976], as follows:

$$W_t = (1 - B)^d Z_t \qquad \qquad \text{[3-32]}$$

W_t is the stationary series. Z_t is the non-stationary series. In practice, d is usually 0, 1, or at most 2 to obtain a stationary series. Time series might transit between stationary and non-stationary state back and forth gradually. Therefore, employing the differenced input and output series in [3-15b] rather than the original series is more advantageous especially in infinite online estimation framework. For instance, one degree time differencing is applied to the inputs and output and we obtain

$$Y_t \Leftarrow (1 - B)Y_t = Y_t - Y_{t-1} \qquad \qquad \text{[3-33]}$$

$$X_{i,t} \Leftarrow (1 - B)X_{i,t} = X_{i,t} - X_{i,t-1} \qquad \text{for } i = 1, 2, 3 \qquad \text{[3-33]}$$

Once the differenced series are prepared, the subsequent estimation procedure remains unchanged.

3.3.4   Minimum mean square error forecast

To obtain an optimal speed forecast using information from both density and speed occurring in the near past, we first estimate the dynamic model [3-15b] in the manner already outlined in the section 3.3.3. Suppose, using the previous notations, that a dynamic model is

$$Y_t = \Omega_0(B) \cdot Y_{t-1} + \sum_{i=1}^{3} \left[\Omega_i(B) \cdot X_{i,t} \right] + \Theta(B) a_t \qquad [3\text{-}35a]$$

$$\text{or} \quad Y_t = \left(\sum_{j=0}^{M_0} \Omega_{0j} B^j \right) \cdot Y_{t-1} + \sum_{i=1}^{3} \left[\left(\sum_{j=0}^{M_i} \Omega_{ij} B^j \right) \cdot X_{i,t} \right] + \left(\sum_{j=0}^{L} \Theta_j B^j \right) \cdot a_t \qquad [3\text{-}35b]$$

where $\Theta_0 = 1$.

Assume that stochastic ARIMA models [Box and Jenkins 1976] for $X_{i,t}$ and Y_t are

$$Y_t = \psi_Y(B) \beta_t = \sum_{j=0}^{\infty} \psi_{Y,j} \beta_{t-j} \qquad [3\text{-}36]$$

where $\psi_Y(B) = \varphi_Y^{-1}(B) \theta_Y(B)$ and $\psi_{Y,0} = 1$

$$X_{i,t} = \psi_{X_i}(B) \alpha_{i,t} = \sum_{j=0}^{\infty} \psi_{X_i,j} \alpha_{i,t-j} \qquad [3\text{-}37]$$

where $\psi_{X_i}(B) = \varphi_{X_i}^{-1}(B) \theta_{X_i}(B)$ and $\psi_{X_i,0} = 1$.

Now [3-35b] may be written as

$$Y_t = \left(\sum_{j=0}^{M_0} \Omega_{0j} B^j \right) \cdot \sum_{j=0}^{\infty} \psi_{Y,j} \beta_{t-1-j} + \sum_{i=1}^{3} \left[\left(\sum_{j=0}^{M_i} \Omega_{ij} B^j \right) \cdot \sum_{j=0}^{\infty} \psi_{X_i,j} \alpha_{i,t-j} \right] + \left(\sum_{j=0}^{L} \Theta_j B^j \right) a_t$$

$$= \sum_{j=0}^{\infty} \widehat{\psi}_{Y,j} \beta_{t-1-j} + \sum_{i=1}^{3} \left[\sum_{j=0}^{\infty} \widehat{\psi}_{X_i,j} \alpha_{i,t-j} \right] + \sum_{j=0}^{L} \Theta_j a_{t-j} \qquad [3\text{-}38]$$

So the theoretical value for the output at time $t+l$ is:

$$Y_{t+l} = \sum_{j=0}^{\infty} \widehat{\psi}_{Y,j} \beta_{t+l-1-j} + \sum_{i=1}^{3} \left[\sum_{j=0}^{\infty} \widehat{\psi}_{X_i,j} \alpha_{i,t+l-j} \right] + \sum_{j=0}^{L} \Theta_j a_{t+l-j} \qquad [3\text{-}39]$$

Suppose that the best forecast $\hat{Y}_t(l)$ of Y_{t+l} made at origin t, which is a linear

function of current and previous β_t's, α_t's and a_t's, is of the form

$$\hat{Y}_t(l) = \sum_{j=0}^{\infty} \psi^*_{Y,(l+j)}\beta_{t-1-j} + \sum_{i=1}^{3}\left[\sum_{j=0}^{\infty}\psi^*_{X_i,(l+j)}\alpha_{i,t-j}\right] + \sum_{j=0}^{\infty}\Theta^*_{l+j}a_{t-j} \qquad [3\text{-}40]$$

Then

$$\text{if } l < L: \ Y_{t+l} - \hat{Y}_t(l) = \sum_{m=0}^{l-1}\widehat{\psi}_{Y,m}\beta_{t+l-1-m} + \sum_{i=1}^{3}\left[\sum_{m=0}^{l-1}\widehat{\psi}_{X_i,m}\alpha_{i,t+l-m}\right] + \sum_{m=0}^{l-1}\Theta_m a_{t+l-m}$$

$$+ \sum_{j=0}^{\infty}\left(\widehat{\psi}_{Y,l+j} - \psi^*_{Y,(l+j)}\right)\beta_{t-1-j} + \sum_{i=1}^{3}\left[\sum_{j=0}^{\infty}\left(\widehat{\psi}_{X_i,l+j} - \psi^*_{X_i,(l+j)}\right)\alpha_{i,t-j}\right]$$

$$+ \sum_{j=0}^{L-l}\left(\Theta_{l+j} - \Theta^*_{l+j}\right)a_{t-j} + \sum_{j=L-l+1}^{\infty}\left(-\Theta^*_{l+j}\right)a_{t-j} \qquad [3\text{-}41]$$

$$\text{if } l \geq L: \ Y_{t+l} - \hat{Y}_t(l) = \sum_{m=0}^{l-1}\widehat{\psi}_{Y,m}\beta_{t+l-1-m} + \sum_{i=1}^{3}\left[\sum_{m=0}^{l-1}\widehat{\psi}_{X_i,m}\alpha_{i,t+l-m}\right] + \sum_{m=0}^{L}\Theta_m a_{t+l-m}$$

$$+ \sum_{j=0}^{\infty}\left(\widehat{\psi}_{Y,l+j} - \psi^*_{Y,(l+j)}\right)\beta_{t-1-j} + \sum_{i=1}^{3}\left[\sum_{j=0}^{\infty}\left(\widehat{\psi}_{X_i,l+j} - \psi^*_{X_i,(l+j)}\right)\alpha_{i,t-j}\right]$$

$$+ \sum_{j=L-l+1}^{\infty}\left(-\Theta^*_{l+j}\right)a_{t-j} \qquad [3\text{-}42]$$

The mean square error is

$$E\left[Y_{t+l} - \hat{Y}_t(l)\right]^2 = \left[\sum_{m=0}^{l-1}\left(\widehat{\psi}_{Y,m}\right)^2\right]\sigma^2_\beta + \sum_{i=1}^{3}\left\{\left[\sum_{m=0}^{l-1}\left(\widehat{\psi}_{X_i,m}\right)^2\right]\sigma^2_{\alpha_i}\right\} + \left[\sum_{m=0}^{l-1}\left(\Theta_m\right)^2\right]\sigma^2_a$$

$$+ \left[\sum_{j=0}^{\infty}\left(\widehat{\psi}_{Y,l+j} - \psi^*_{Y,(l+j)}\right)^2\right]\sigma^2_\beta + \sum_{i=1}^{3}\left\{\left[\sum_{j=0}^{\infty}\left(\widehat{\psi}_{X_i,l+j} - \psi^*_{X_i,(l+j)}\right)^2\right]\sigma^2_\alpha\right\}$$

$$+ \left[\sum_{j=0}^{L-l}\left(\Theta_{l+j} - \Theta^*_{l+j}\right)^2 + \sum_{j=L-l+1}^{\infty}\left(\Theta^*_{l+j}\right)^2\right]\sigma^2_a, \quad \text{for } l < L. \qquad [3\text{-}43]$$

$$E\left[Y_{t+l} - \hat{Y}_t(l)\right]^2 = \left[\sum_{m=0}^{l-1}(\widehat{\psi}_{Y,m})^2\right]\sigma_\beta^2 + \sum_{i=1}^{3}\left\{\left[\sum_{m=0}^{l-1}(\widehat{\psi}_{X_i,m})^2\right]\sigma_{\alpha_i}^2\right\} + \left[\sum_{m=0}^{L}(\Theta_m)^2\right]\sigma_a^2$$

$$+ \left[\sum_{j=0}^{\infty}(\widehat{\psi}_{Y,l+j} - \psi^*_{Y,(l+j)})^2\right]\sigma_\beta^2 + \sum_{i=1}^{3}\left\{\left[\sum_{j=0}^{\infty}(\widehat{\psi}_{X_i,l+j} - \psi^*_{X_i,(l+j)})^2\right]\sigma_\alpha^2\right\}$$

$$+ \left[\sum_{j=L-l+1}^{\infty}(\Theta^*_{l+j})^2\right]\sigma_a^2, \quad \text{for } l \geq L. \tag{3-44}$$

which is minimized only if $\psi^*_{Y,l+j} = \widehat{\psi}_{Y,l+j}$, $\psi^*_{X_i,l+j} = \widehat{\psi}_{X_i,l+j}$, and

$$\Theta^*_{l+j} = \begin{cases} \Theta_{l+j} & \text{for } j \leq L-l \\ 0 & \text{for } j > L-l \end{cases}.$$

If we rewrite [3-39]

$$Y_{t+l} = c + \sum_{j=0}^{l-1}\widehat{\psi}_{Y,j}\beta_{t+l-1-j} + \sum_{i=1}^{3}\left[\sum_{j=0}^{l-1}\widehat{\psi}_{X_i,j}\alpha_{i,t+l-j}\right] + \sum_{j=0}^{l-1}\Theta_j a_{t+l-j}$$

$$+ \sum_{j=l}^{\infty}\widehat{\psi}_{Y,j}\beta_{t+l-1-j} + \sum_{i=1}^{3}\left[\sum_{j=l}^{\infty}\widehat{\psi}_{X_i,j}\alpha_{i,t+l-j}\right] + \sum_{j=l}^{L}\Theta_j a_{t+l-j}, \quad \text{for } l < L \tag{3-45}$$

$$Y_{t+l} = c + \sum_{j=0}^{l-1}\widehat{\psi}_{Y,j}\beta_{t+l-1-j} + \sum_{i=1}^{3}\left[\sum_{j=0}^{l-1}\widehat{\psi}_{X_i,j}\alpha_{i,t+l-j}\right] + \sum_{j=0}^{L}\Theta_j a_{t+l-j}$$

$$+ \sum_{j=l}^{\infty}\widehat{\psi}_{Y,j}\beta_{t+l-1-j} + \sum_{i=1}^{3}\left[\sum_{j=l}^{\infty}\widehat{\psi}_{X_i,j}\alpha_{i,t+l-j}\right], \quad \text{for } l \geq L \tag{3-46}$$

Therefore,

$$E_t\left[Y_{t+l}\right] = c + \sum_{j=l}^{\infty}\widehat{\psi}_{Y,j}\beta_{t+l-1-j} + \sum_{i=1}^{3}\left[\sum_{j=l}^{\infty}\widehat{\psi}_{X_i,j}\alpha_{i,t+l-j}\right] + \sum_{j=l}^{L}\Theta_j a_{t+l-j}$$

$$= \hat{Y}_t(l) \tag{3-47}$$

since the conditional expectations of series β_h's, α_h's and a_h's (for h>t) given the past observations up to t are all zeros. Thus the minimum mean square error forecast $\hat{Y}_t(l)$ of Y_{t+l} at origin t is given by the conditional expectation of Y_{t+l} at time t. The forecast is actually conditional on the knowledge of the series from the past up to the present origin t.

Now, using $\langle \ \rangle$ to denote conditional expectations at time t, we can compute the forecast as

$$\hat{Y}_t(l) = \langle Y_{t+l} \rangle = c + \left(\sum_{j=0}^{M_0} \Omega_{0j} \langle Y_{t+l-1-j} \rangle \right) + \sum_{i=1}^{3} \left[\left(\sum_{j=0}^{M_i} \Omega_{ij} \langle X_{i,t+l-j} \rangle \right) \right]$$

$$+ \left(\sum_{j=0}^{L} \Theta_j \langle a_{t+l-j} \rangle \right) \qquad \text{[3-48]}$$

$$\text{where } \langle Y_{t+j} \rangle = \begin{cases} Y_{t+j} & j \le 0 \\ \hat{Y}_t(j) & j > 0 \end{cases} \qquad \text{[3-49]}$$

$$\langle X_{i,t+j} \rangle = \begin{cases} X_{i,t+j} & j \le 0 \\ \hat{X}_{i,t}(j) & j > 0 \end{cases} \qquad \text{[3-50]}$$

$$\langle a_{t+j} \rangle = \begin{cases} a_{t+j} = \dfrac{1}{\Theta_0} \left\{ Y_{t+j} - \left(\sum_{k=0}^{M_0} \Omega_{0k} Y_{t+j-1-k} \right) \right. \\ \qquad - \sum_{i=1}^{3} \left[\left(\sum_{k=0}^{M_i} \Omega_{ik} X_{i,t+j-k} \right) \right] \\ \qquad \left. - \left(\sum_{k=1}^{L} \Theta_k a_{t+j-k} \right) \right\} & j \le 0 \\ 0 & j > 0 \end{cases} \qquad \text{[3-51]}$$

64

Hence, the minimum mean square error forecast $\hat{Y}_t(l)$ of Y_{t+l} at origin t is based on the past observations for the inputs and output up to time t as well as the previous (yet after the current time t) forecasts.

Meanwhile, the variance of the forecast error for $\hat{Y}_t(l)$ is

$$Variance\left(\hat{Y}_t(l)\right) = E\left[Y_{t+l} - \hat{Y}_t(l)\right]^2$$

$$= \left[\sum_{m=0}^{l-1}(\widehat{\psi}_{Y,m})^2\right]\sigma_\beta^2 + \sum_{i=1}^{3}\left\{\left[\sum_{m=0}^{l-1}(\widehat{\psi}_{X_i,m})\right]^2\sigma_{\alpha_i}^2\right\} + \left[\sum_{m=0}^{l-1 \text{ or } L}(\Theta_m)^2\right]\sigma_a^2 \qquad [3\text{-}50]$$

It is apparent to see that the larger the forecast horizon l, the higher the variance would be.

3.3.5 Adaptive calibration and forecasting

In the context of online application, the parameters in the dynamic speed-density model based on the transfer function could be recalibrated using the above ordinary least square based approach when fresh real-time traffic data become available. Such an adaptive mechanism would provide a systematic way to maintain an updated speed-density relation that is consistent with the most recent traffic states.

A rolling horizon scheme is designed for the proposed adaptive calibration and forecasting. In this scheme, the observed data in one finite time period (the horizon) are chosen at each time to obtain estimates of the coefficients. The derived coefficients are applied instantly to the subsequent traffic flow simulation in order to predict speeds. Virtually, the estimated model could be applied to predict speeds for any future time beyond the end of calibration horizon. As time advances, the horizon

moves forward by one roll period and a new calibration is performed. So whenever the new model with the newly calibrated coefficients is obtained, speeds are to be updated adaptively no matter what are the previous predictions. Figure 3-3 shows the data processing pattern for adaptive calibration of speed-density relationship and speed forecasting.
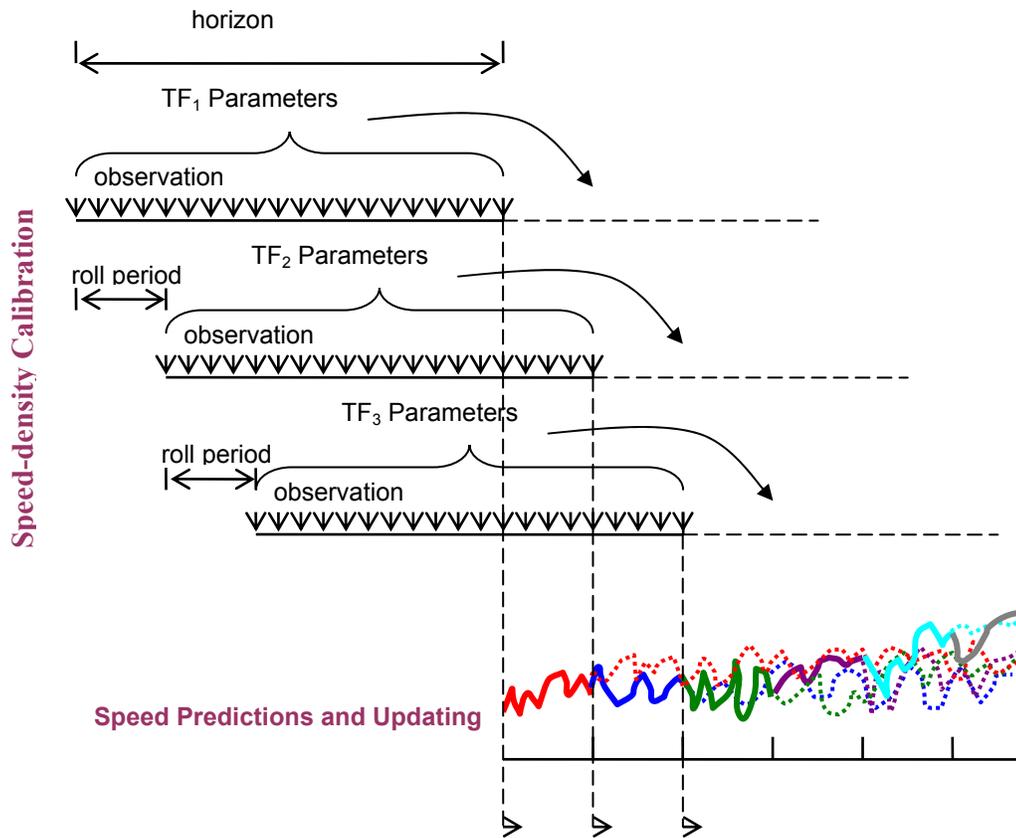

Figure 3-3 Data Processing Pattern for Adaptive Calibration and Forecasting

As shown in the figure, after every roll period, new obtained pairs of measured traffic data (density and speed) are included in the calibration process, while the same amount of the most outdated traffic data in the previous horizon is

discarded. The amount of data being processed each time is fixed for a given horizon duration, which results in nearly constant computation cost each time. Another benefit of implementing the rolling horizon scheme is that it can reflect the most recent traffic dynamics and ignore the effect of the more remote past situations on the current and future system. Especially when traffic undergoes an unusual pattern, the rolling horizon scheme could adjust the speed-density relation in a timely manner to a different pattern to provide more realistic estimation.

It should be pointed out that the adaptation is very critical to the application of the dynamic speed-density model. The ability of the model to capture the inner dynamic evolution of traffic flow is greatly relied on the acquisition of the latest traffic data. It is found in our experiments that, without the adaptation mechanism, the coefficients in the model are time insensitive and the model performance in speed prediction cannot even match the conventional static model.

In the following chapter, a series of tests for adaptive calibration and forecasting are conducted to illustrate the benefit of the proposed dynamic speed-density relation against the conventional static model for speed prediction in the context of real time operation.

3.4   *Summary*

In this chapter, the detailed formulations of speed forecasting from real-time traffic measurements are presented. By extending the higher-order continuum traffic flow theory and exploring the time-series approaches, a dynamic speed-density relation is proposed. The model explicitly includes the speed relaxation, speed

convection and density anticipation which are found to have impacts on dynamics of speeds. Besides these factors, the model also recognizes that the associated noise series is auto-correlated. By applying the approach in the transfer function model, the procedures for the model estimation and speed prediction using the real data are also presented. For online application, the rolling horizon framework is proposed for the adaptive calibration and prediction.

Chapter 4: Standalone Evaluation Numerical Test

To evaluate the model performance for online application, a series of standalone tests are intended to adaptively calibrate the dynamic speed-density model and predict speed using the rolling horizon framework which is already outlined in the Chapter 3. In the standalone tests, we mimic online calibration and speed prediction by assuming densities are known values at any time interval t which are equivalent to the observed density series. The standalone test aims to investigate the performance of the proposed model in replicating the real output given the real inputs. However, in a *real* online network traffic estimation and prediction, densities are usually derived by a certain traffic network modeling, e.g. the simulation-based DTA model. The approaches presented in Chapter 5 and the experiments described in Chapter 6 are to incorporate the proposed model into the online simulation-based DTA model by taking the place of the static model, where densities used are simulated values.

4.1   Test description

The data used in this test were collected from freeway sections on US I-405 in Irvine, Orange County, CA. The dataset is composed of the 5-day density and speed data at 30-second sampling interval from 4:00 am to 10:00 am on three continuous freeway sections. Figure 4-1 shows the location of sampled sections, section A, section B, and section C.

Figure 4-1 Data Sampling Locations

Among the three freeway sections, the model calibration and speed prediction are performed for section A. The archived data from section A provide the information of speed relaxation, while the information of downstream anticipation and upstream convection are obtained via the archived data from the neighboring section B and C, respectively.

In a real transportation network, not all traffic sections are under monitoring and data-sampling. The most informative case occurs when data are collected at the section under study as well as the upstream and downstream sections. If data collection is geometrically isolated, the speed prediction via real-time data can only count on the single data stream which is much less informative than the former case. The in-between case is when data are collected locally plus either upstream or downstream. Therefore, to accommodate the realistic data availability in the real-world, four variants of the dynamic speed-density model are to be tested.

Model type 1 has the simplest formulation where only the relaxation term is included as the dynamic driving force.

$$u_{t+1}^i = u_t^i + \delta_1^{-1}(B)\omega_1(B)\cdot\left[u^e\left(k_t^i\right)-u_t^i\right]+\varphi^{-1}(B)\theta(B)a_t \qquad [4\text{-}1]$$

70

Model type 2 contains effects of the relaxation and convection.

$$u_{t+1}^i = u_t^i + \delta_1^{-1}(B)\omega_1(B)\cdot\left[u^e\left(k_t^i\right)-u_t^i\right]+\delta_2^{-1}(B)\omega_2(B)\cdot\left[u_t^i\left(u_t^{i-1}-u_t^i\right)\right]$$

$$+\,\varphi^{-1}(B)\theta(B)a_t \qquad\qquad\qquad\qquad [4\text{-}2]$$

Model type 3 contains effects of the relaxation and anticipation.

$$u_{t+1}^i = u_t^i + \delta_1^{-1}(B)\omega_1(B)\cdot\left[u^e\left(k_t^i\right)-u_t^i\right]+\delta_3^{-1}(B)\omega_3(B)\cdot\left(\frac{k_t^{i+1}-k_t^i}{k_t^i}\right)$$

$$+\,\varphi^{-1}(B)\theta(B)a_t \qquad\qquad\qquad\qquad [4\text{-}3]$$

Model type 4 incorporates all of the three most important factors.

$$u_{t+1}^i = u_t^i + \delta_1^{-1}(B)\omega_1(B)\cdot\left[u^e\left(k_t^i\right)-u_t^i\right]+\delta_2^{-1}(B)\omega_2(B)\cdot\left[u_t^i\left(u_t^{i-1}-u_t^i\right)\right]$$

$$+\,\delta_3^{-1}(B)\omega_3(B)\cdot\left(\frac{k_t^{i+1}-k_t^i}{k_t^i}\right)+\varphi^{-1}(B)\theta(B)a_t \qquad\qquad [4\text{-}4]$$

As we can see in these model variants 1 through 4, we assume that data are definitely collected from the traffic section under study so that their speed relaxation information is always available. When fewer triggering forces are considered, as in Model type 1, it is expected that the noise would be higher since more uncontrolled factors exist but the model estimation is much simpler and quicker. Our concern is which model variant has a good combination of prediction accuracy and computation efficiency. Hence, comparison of the four model variants is one of the spotlights in the tests.

The effectiveness of the model will be evaluated by comparing MOEs generated in the static model and the dynamic model. The most straightforward MOE is the average root mean squared error (RMSE) of the predicted speed against the actual value over the whole prediction period. To be more sensible, RMSE will be

weighted by the number of vehicles in each observation interval. Using Δ denote the weighted RMSE, then

$$\Delta = \sqrt{\frac{\sum_t \left[k_t \cdot (u_t - \hat{u}_t)^2 \right]}{\sum_t k_t}} \qquad \text{for } t = 1, \cdots, T_p \qquad [4\text{-}5]$$

where t is the time step at which speed prediction is made; there are totally T_p of time intervals.

Using the archived traffic data containing density and speed in Day 1, static speed-density relations specified by the modified Greenshields' model are calibrated. The calibrated static models serve for two purposes. First, the calibrated static models are used as the equilibrium relation to calculate the equilibrium speed in the dynamic model at each time interval given a known density value. Second, the speeds predicted by the static model are about to be compared with the dynamic model, so the calibrated static relation is also a benchmark. In all the tests hereafter, the modified Greenshields' model is adopted as the static model form.

The modified Greenshields' model has the general form

$$u = \left(u^f - u^0 \right) \left(1 - \frac{k}{k^j} \right)^\alpha + u^0 \qquad [4\text{-}6]$$

where

u is the mean speed;

u^f is the mean free speed;

u^0 is the minimum speed;

k is the density;

k^j is the jam density;

α is the power term used to capture the sensitivity of speed to the density.

If a range with a constant speed (e.g. when $0 \leq k \leq k^o$, where k^o is the optimum density) is particularly modeled, a dual-regime modified Greenshields' model emerges in which the constant free-flow speed is specified for the free-flow conditions and a modified Greenshields' model is specified for congested-flow conditions (Figure 4-2 (a)). In contrast, a single-regime model applies the modified Greenshields' model for both free- and congested-flow conditions (Figure 4-2 (a)).

Dual-regime models are generally applicable to freeways, whereas single-regime models apply to arterials [University of Maryland DTA Group, 2003]. The reason why a dual-regime model is applicable for freeways in particular is that freeways have typically more capacity (higher service flow rates)than arterials, and can accommodate dense traffic (up to 2300 pc/hour/lane) at near free-flow speeds. On the other hand, arterials have signalized intersections, meaning that such a phenomenon may be short-lived, if present at all. Hence, a slight increase in traffic would elicit more deterioration in prevailing speeds than in the case of freeways. Therefore, arterial traffic relations are better explained using a single-regime model. Since the sections to be tested are freeway, the dual-regime model is more applicable. Figure 4-3 shows the calibrated static model against the actual data points for the freeway section A.
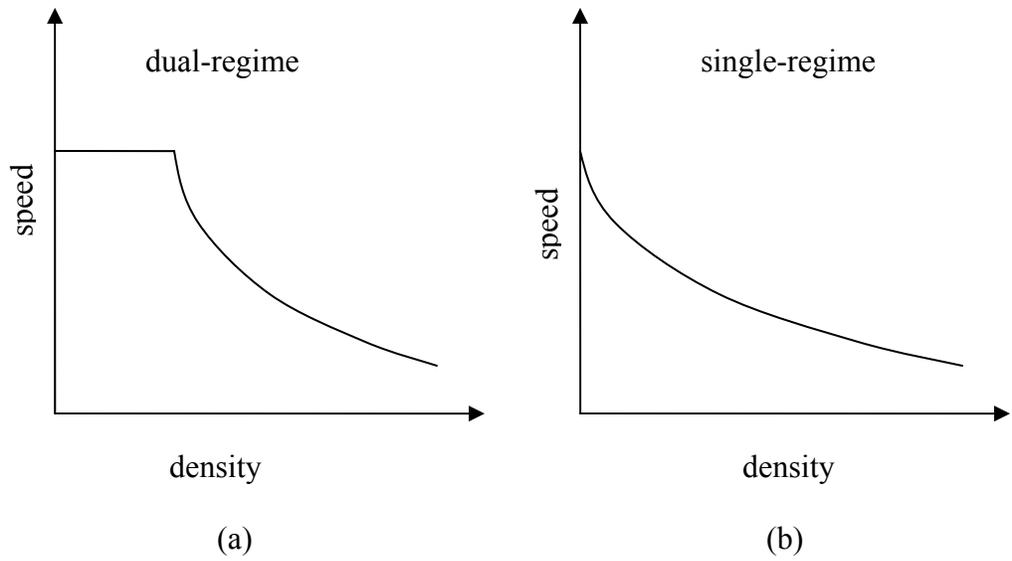
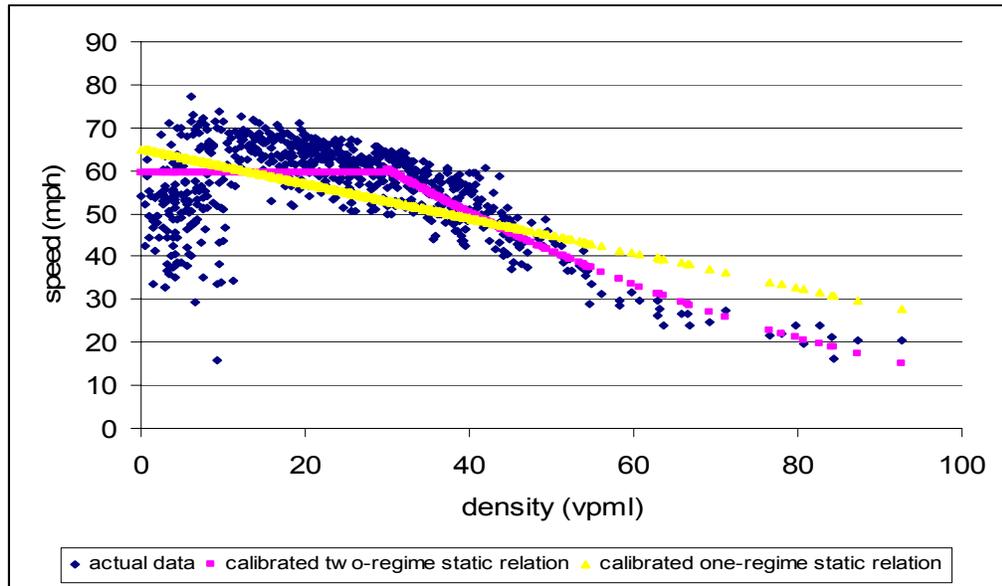**Figure 4-2 Two Types of the Modified Greenshields's model**

Figure 4-3 Illustration of Calibrated Modified Greenshields' Models against Actual Data on Freeway Section A

Two groups of tests are devised to examine the dynamic model in the adaptive mode. In the first set of tests, the rolling horizon scheme is fixed with the calibration horizon = 60-minute and the roll period = 2.5-minute. Performances of the four variants of the dynamic model are to be evaluated combined with different static models (calibrated or borrowed). In the second set of tests, the model sensitivity to the rolling horizon scheme is analyzed.

All the adaptive calibrations and predictions are made using Day 2 ~ 5 data since Day 1 data are completely contributed for the static relation calibration. The following section shows the test results and the related findings.

4.2 *Results and analysis*

4.2.1 Model effectiveness and robustness

First, we use the calibrated dual-regime modified Greenshields' model as the equilibrium relation and the one to be compared with the dynamic model which is applied adaptively. Figure 4-4 to Figure 4-7 are the time series plots for the computed speed verse the actual data of Day 2 in these tests.
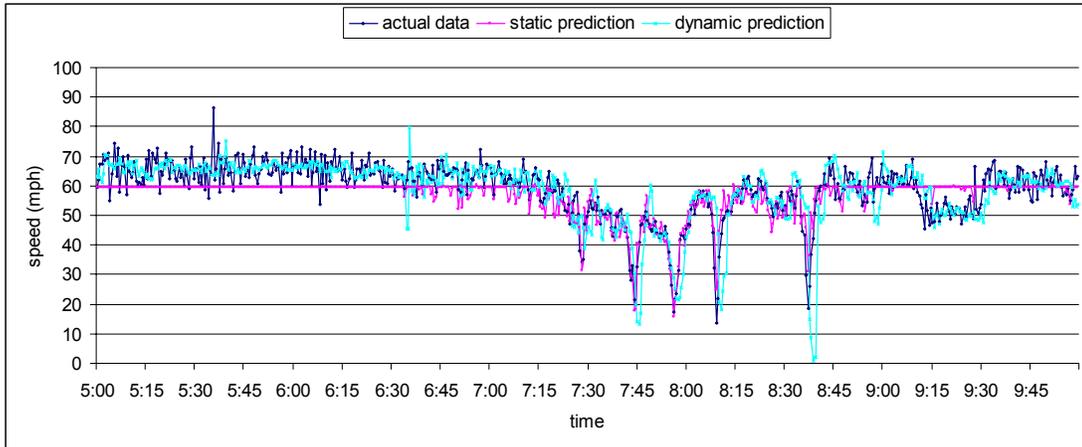
Figure 4-4 Dynamic Model Type 1 + Calibrated Dual-regime MG Model

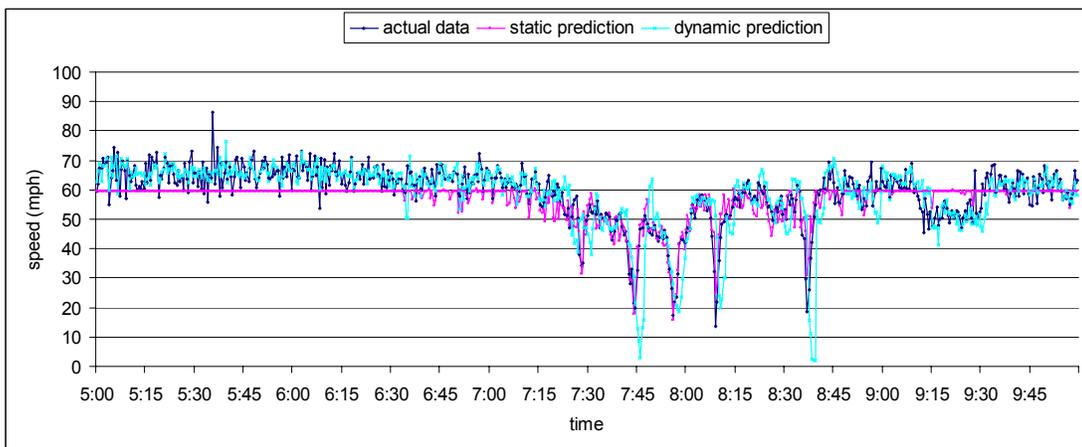**(RMSE: static prediction — 5.47 mph, dynamic prediction — 8.90 mph)**

Figure 4-5 Dynamic Model Type 2 + Calibrated Dual-regime MG Model

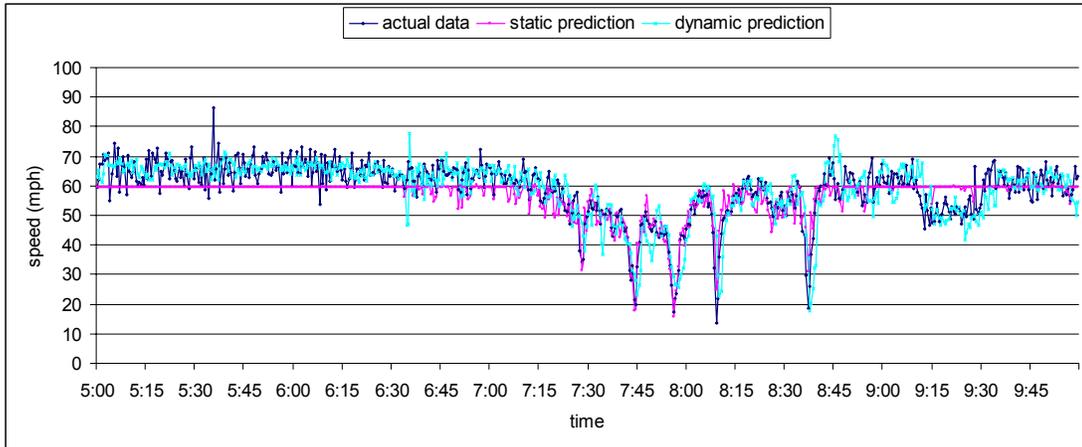**(RMSE: static prediction — 5.47, dynamic prediction — 9.24)**

76

Figure 4-6 Dynamic Model Type 3 + Calibrated Dual-regime MG Model

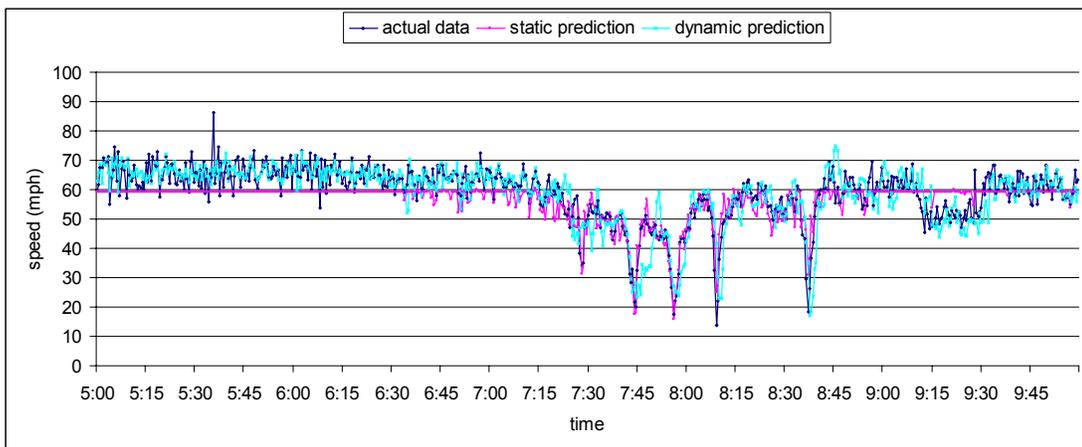**(RMSE: static prediction — 5.47, dynamic prediction — 6.74)**

Figure 4-7 Dynamic Model Type 4 + Calibrated Dual-regime MG Model

(RMSE: static prediction — 5.47, dynamic prediction — 6.66)

Then, we use the calibrated single-regime modified Greenshields' model as the equilibrium relation and the one to be compared with the dynamic model which is

applied adaptively. Figure 4-8 to Figure 4-11 are the time series plots for the computed speed verse the actual data of Day 2 in these tests.
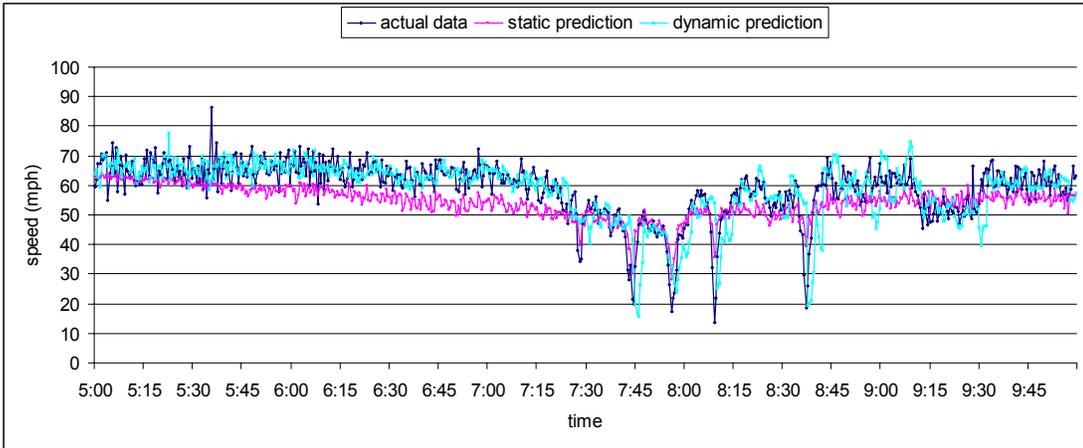

Figure 4-8 Dynamic Model Type 1 + Calibrated Single-regime MG Model

(RMSE: static prediction — 7.64, dynamic prediction — 8.44)

Figure 4-9 Dynamic Model Type 2 + Calibrated Single-regime MG Model

(RMSE: static prediction — 7.64, dynamic prediction — 8.39)

Figure 4-10 Dynamic Model Type 3 + Calibrated Single-regime MG Model

(RMSE: static prediction — 7.64, dynamic prediction — 7.62)

Figure 4-11 Dynamic Model Type 4 + Calibrated Single-regime MG Model

(RMSE: static prediction — 7.64, dynamic prediction — 7.24)

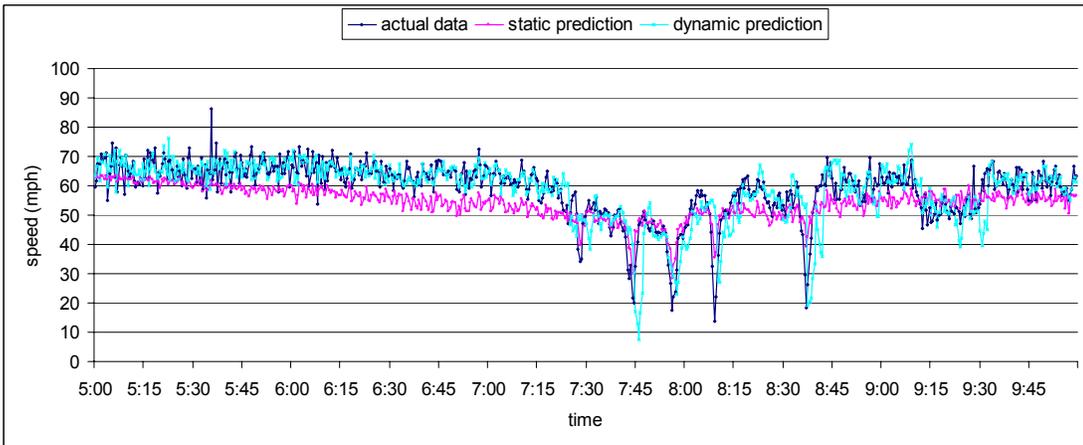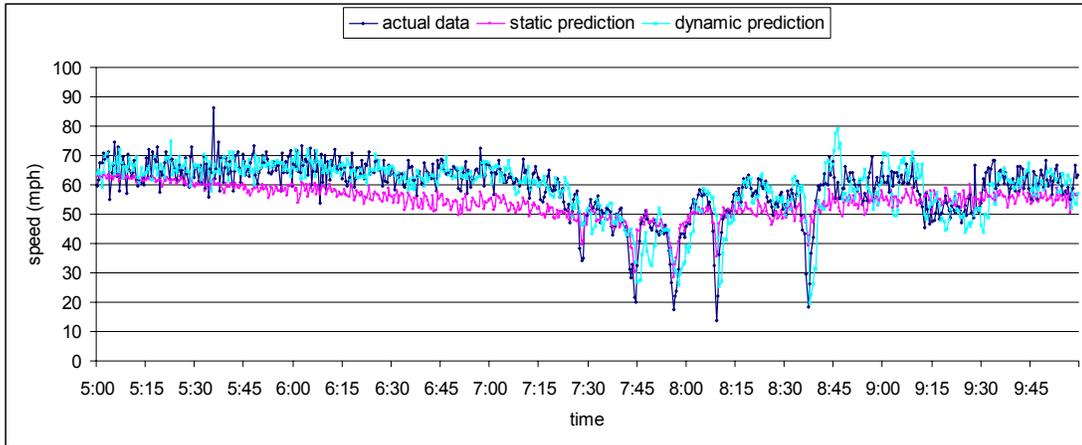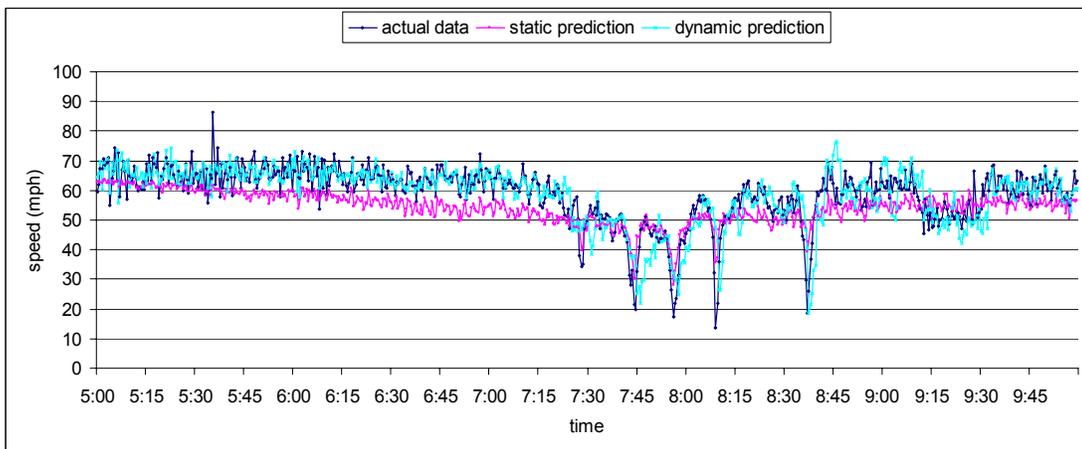The results show that the adaptively calibrated dynamic model performs comparably both with the single-regime static relation and the dual-regime static relation. However the identification of the single-regime static relation would be

79

easier, which would simplify the whole procedure. In addition, all of the four variants of the dynamic models are working well in predicting speed. The dynamic model Type 4 with three inputs has the lowest prediction error over the other three models, but more computation efforts are engaged. By comparing the results form the Type 1 and the Type 2, or the Type 3 and the Type 4, we found that the impact of convection is not as apparent as the anticipation. The model Type 3 which includes the effects of relaxation and anticipation shows the best trade-off between accuracy and computation efficiency.

Now, instead of using the calibrated static relation with the data from the section A under study, the static relations are calibrated for another similar road facility (e.g. a different freeway section but with similar characteristic) and borrowed to the section A to be used as its static relations in the tests. When applied to the section A, these relations either overestimate or underestimate speeds on average. Figures 4-12 to 4-15 are the time series plots for the computed speed with the overestimated equilibrium relation verse the actual data of Day 2 in these tests. To restrain the length of the dissertation, only the time series plots with the single-regime MG model are shown. The results of the dual-regime MG model can be found in a summary plot (Figure 4-20) presented later.

Figure 4-12 Dynamic Model Type 1 + Overestimated Single-regime MG Model

(RMSE: static prediction — 9.64, dynamic prediction — 8.36)

Figure 4-13 Dynamic Model Type 2 + Overestimated Single-regime MG Model

(RMSE: static prediction — 9.64, dynamic prediction — 8.34)

Figure 4-14 Dynamic Model Type 3 + Overestimated Single-regime MG Model

(RMSE: static prediction — 9.64, dynamic prediction — 7.49)

Figure 4-15 Dynamic Model Type 4 + Overestimated Single-regime MG Model

(RMSE: static prediction — 9.64, dynamic prediction — 7.11)

Figure 4-16 to Figure 4-19 are the time series plots for the computed speed with the underestimated equilibrium relation verse the actual data of Day 2.
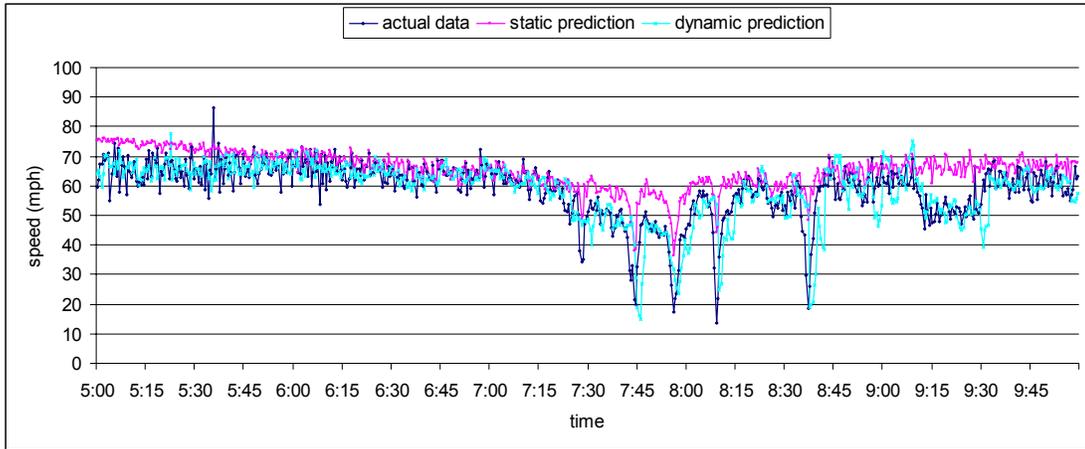
82
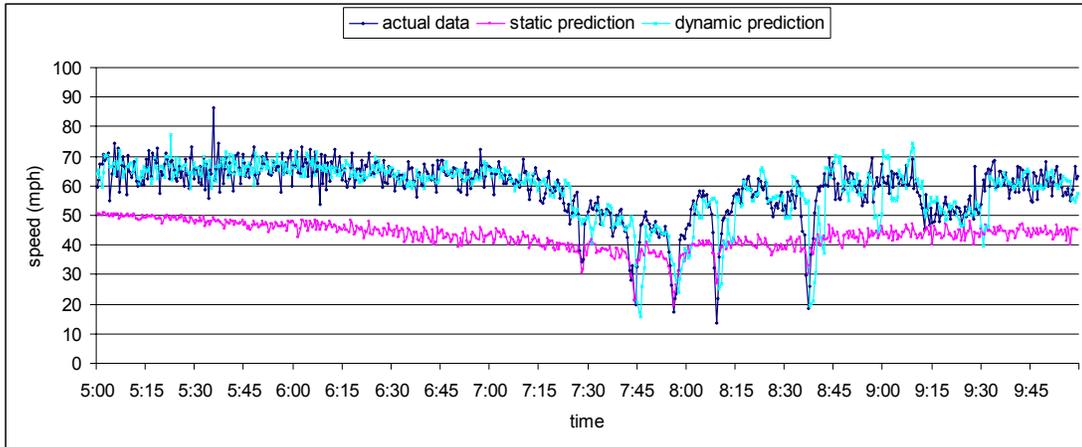
Figure 4-16 Dynamic Model Type 1 + Underestimated Single-regime MG Model

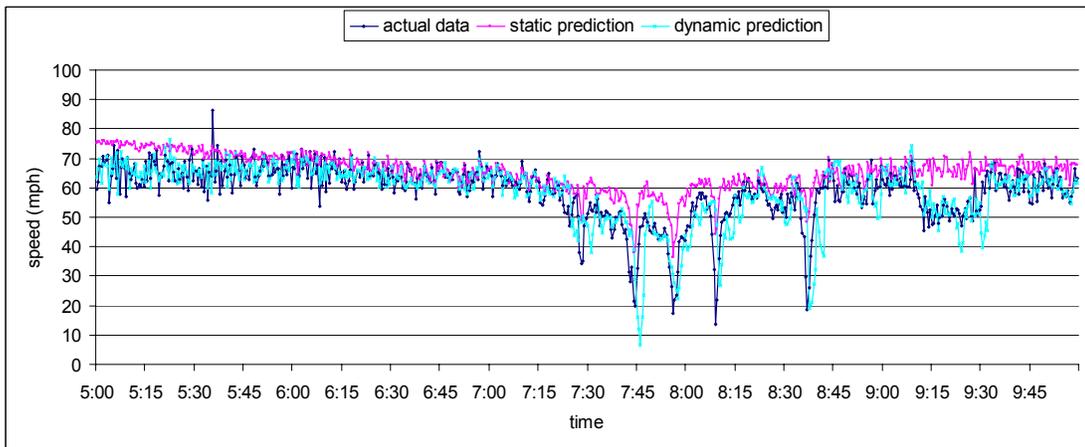**(RMSE: static prediction — 15.93, dynamic prediction — 8.54)**
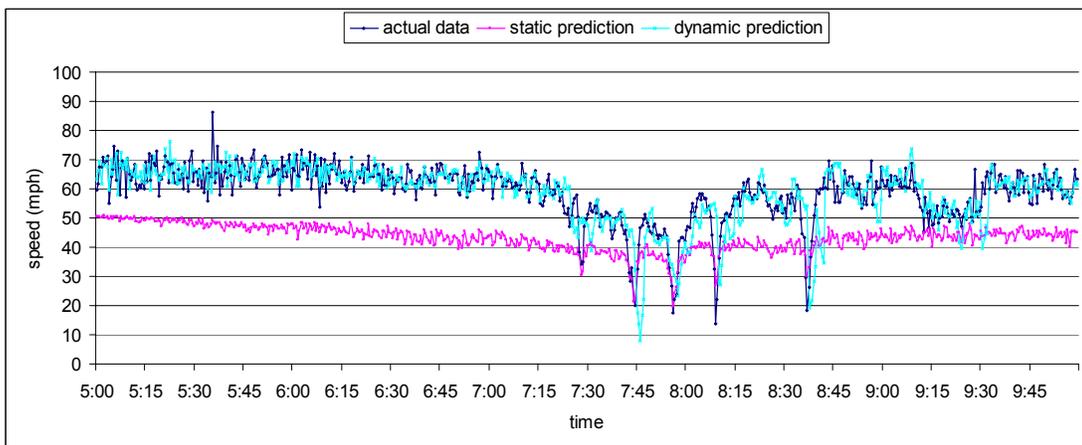
Figure 4-17 Dynamic Model Type 2 + Underestimated Single-regime MG Model

(RMSE: static prediction — 15.93, dynamic prediction — 8.47)
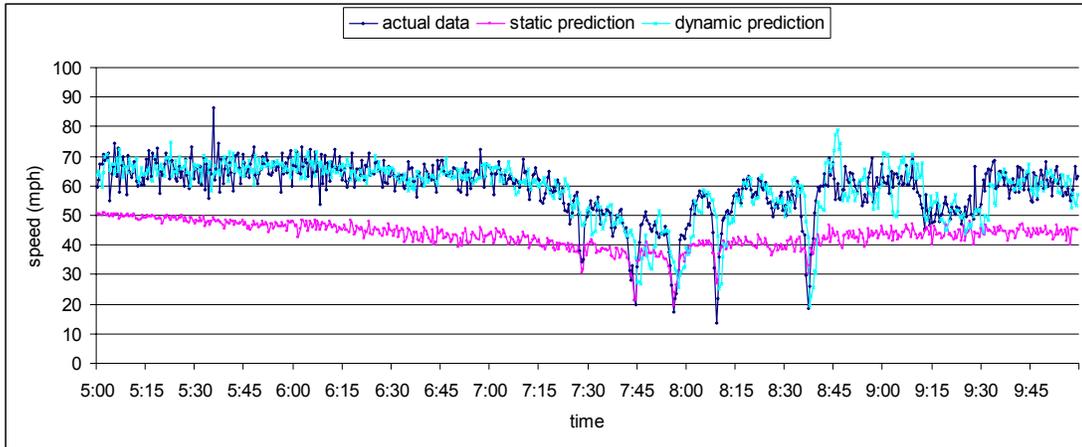
Figure 4-18 Dynamic Model Type 3 + Underestimated Single-regime MG Model

(RMSE: static prediction — 15.93, dynamic prediction — 7.75)

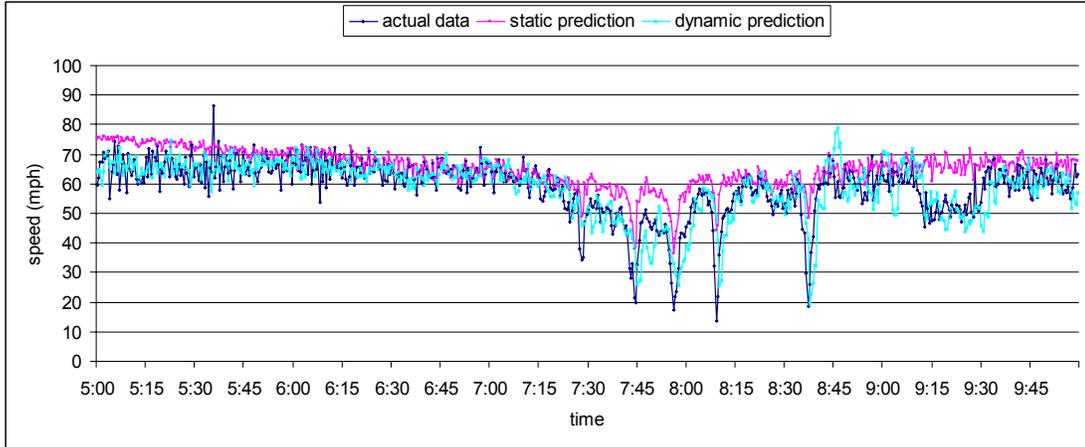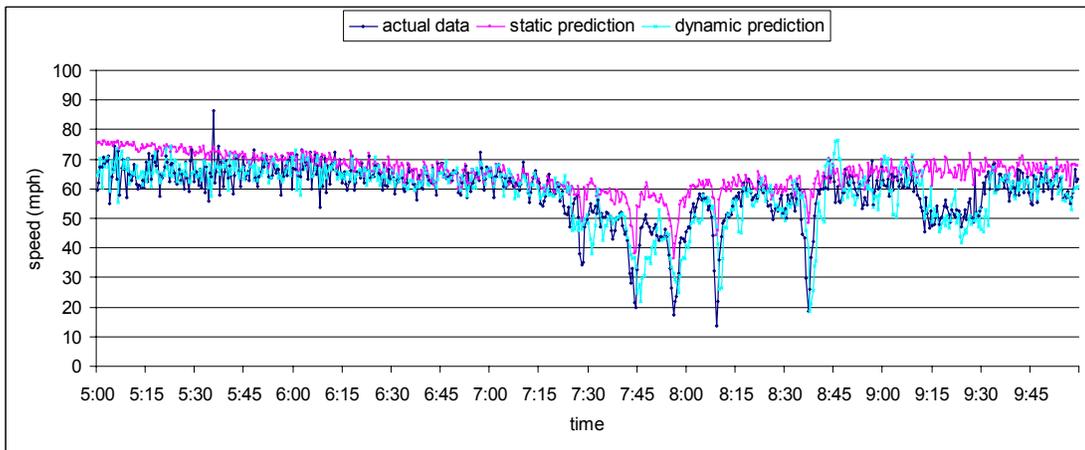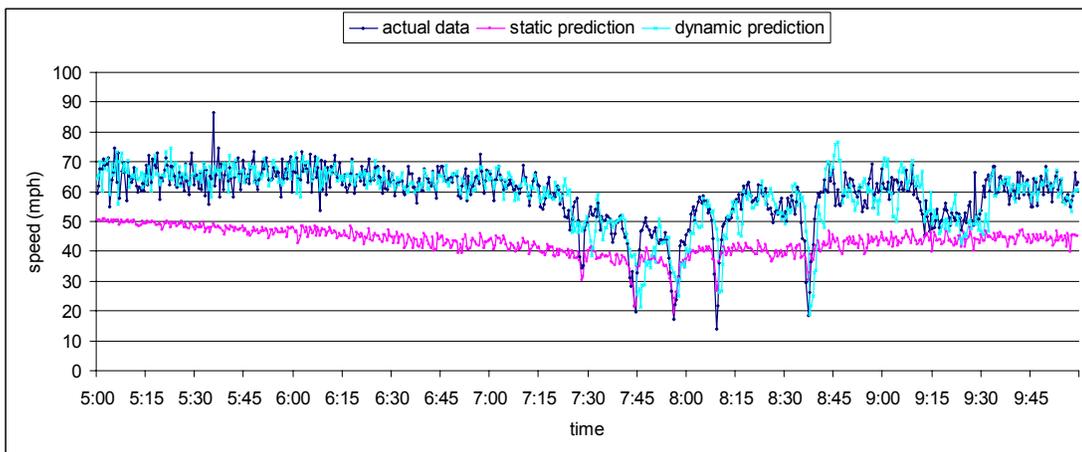**Figure 4-19 Dynamic Model Type 4 + Underestimated Single-regime MG Model**

(RMSE: static prediction — 15.93, dynamic prediction — 7.36)

The above tests show that all four variants of the TF-based dynamic models are not sensitive to the preciseness of the static relation. The performance down-grades are not significant with the quality of the borrowed static models. Similar to

the previous finding, the complex dynamic model 4 with three-input still outperforms the other three and the model type 3 is more efficient and has enough accuracy.

To summarize the above tests, Figure 4-20 and Figure 4-21 show the performances of the dynamic models versus the static models for Day 2 and Day 2 - Day 5 on average. A similar pattern can be viewed from these two plots, indicating that the prediction made by the dynamic model is much more stable than the static model under different quality of the static relation.
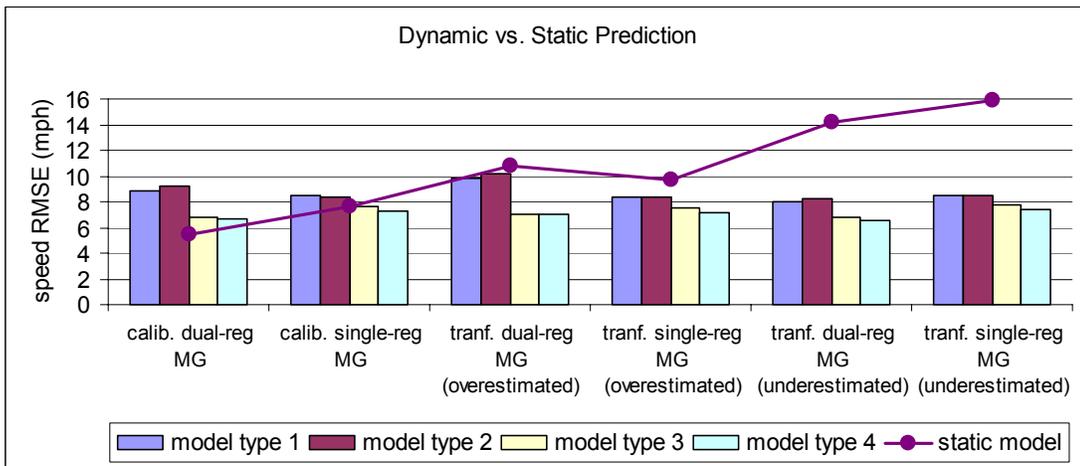
Figure 4-20 Model Performance Comparison for Day 2
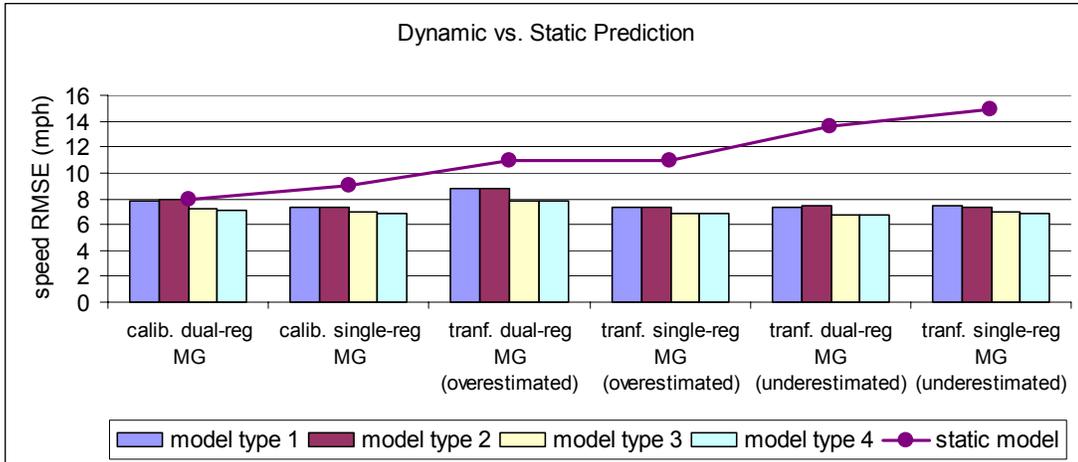
Figure 4-21 Average Model Performance Comparison over Day 2-Day 5

In Figure 4-20, it is clear that the dynamic model is not dominantly better than the static model for Day 2. In the case of using calibrated static model (dual-regime or single regime Modified Greenshields' model) for each dynamic model type, the average errors of the static model is better than (for dual-regime) or similar to (for single-regime) the dynamic models. To look into how the dynamic model does not pick up the speed trend as well as the static model, Figure 4-22 displays the hour-by-hour speed RMSEs for the static model (calibrated for the section A using Day 1 data) and the dynamic model (Type 3) for Day 2. The plot shows that the dynamic model causes lower speed prediction errors than the static model during the off-peak hour periods (5:00am-7:00am and 9:00am-10:00am). However, during the peak hour periods (7:00am-9:00am), the static model becomes more accurate than the dynamic model. The reason for that is because the congestion pattern of the traffic during peak hour in Day 2 is highly close to what happened the same period in Day 1. The one-to-one relation of speed and density during congestion (high density and low speed)

usually exhibits a well identifiable curve with little fluctuation. Thus, the static relation calibrated for the peak hours in Day 1 is quite fit for Day 2, which brings about the good performance during the peak hour for Day 2. On the other hand, the dynamic model is relatively insensitive to the quality of the embedded static relation. No matter a "perfect" static relation or a coarse static relation is used, the dynamic model only passes smaller changes on the prediction quality than the static model.

The point can be demonstrated further in Figure 4-23 and Figure 4-24. In Figure 4-23, the same static model (calibrated from Day 1) is used for Day 4. As we can see, the dynamic model outperforms the static model during the whole prediction horizon. The traffic pattern in Day 4 is different from Day 1. The underlying static models for these two days are different. Higher prediction errors than Day 2 occur when the static model from Day 1 is used for Day 4 to estimate speeds. However, the corresponding dynamic model produces a little error increase compared to Day 2. It is because the equilibrium value is not the only trigger factor to determine the speed dynamics and a series of past as well as neighboring traffic states are taken into account in the dynamic model. It would be more obvious if we revisit the time-series plots (Figure 4-4 through Figure 4-19). During the time between 9:15am to 9:30am, the actual traffic experienced a short slow-down. The dynamic model shows the ability to pick up this dynamics even though the static model can not.

In Figure 4-24, another static model which is calibrated from another location instead of the section A is used for Day 2. It is evident that the static model is not fit for Day 2 so that the prediction errors of the static model in all the study hours get

significant increases.   Similarly, the performance of the dynamic model is not

affected greatly per se.
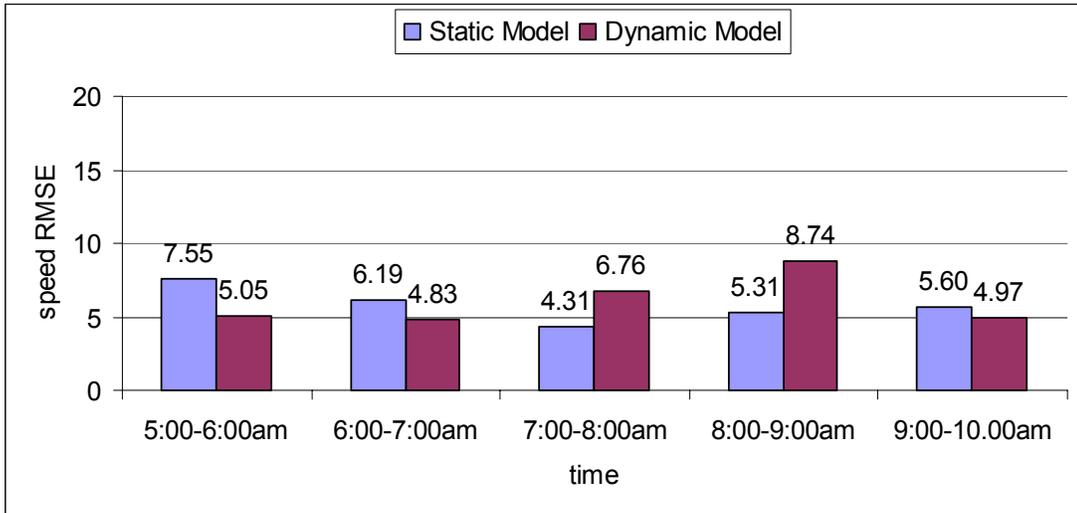
**Figure 4-22 Hour-by-hour Speed RMSEs for Static Model vs. Dynamic Model
(Day 2)**
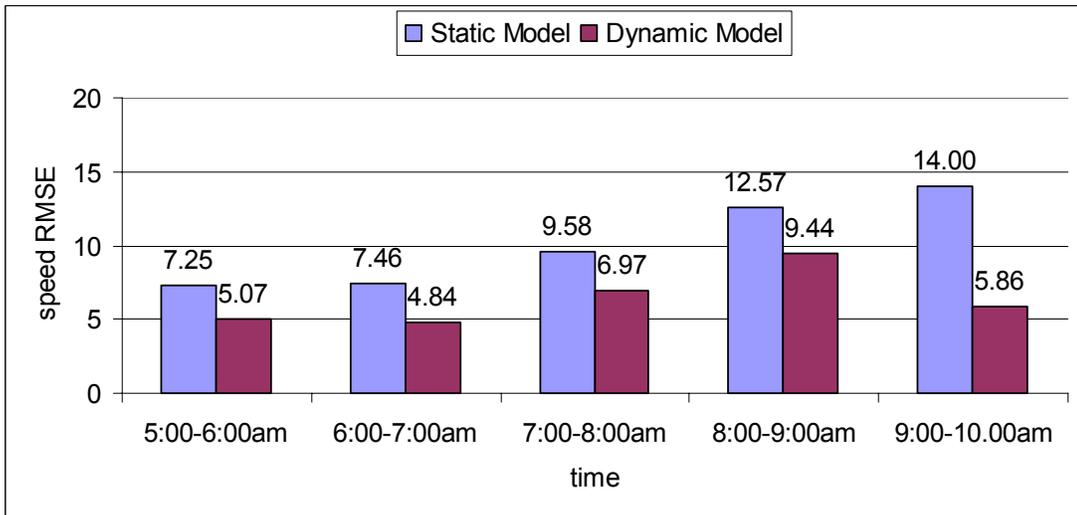
**Figure 4-23 Hour-by-hour Speed RMSE for Static Model vs. Dynamic Model
(Day 4)**
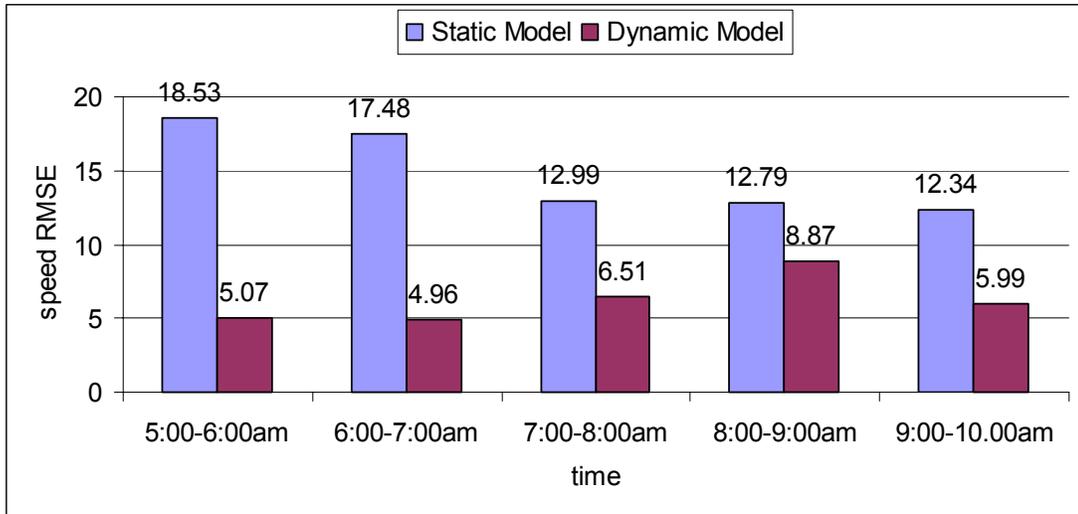
Figure 4-24 Hour-by-hour Speed RMSE for Static Model (borrowed) vs. Dynamic Model (Day 2)

In summary, the robustness of the dynamic model is revealed under different quality of the equilibrium relations. Fully incorporating three factors (relaxation, convection, and anticipation) would bring the most accurate result. However in terms of accuracy and efficiency, the model with relaxation and anticipation performs the best. Using dual-regime MG relation in the dynamic model does not significantly outperform the single-regime MG relation. Another implication from the test results is that if the static model during a particular period (i.e. peak hour) is deterministic and very close to the actual traffic pattern, it might be better than dynamic model. However, the day-to-day variations or location-to-location variations existing in the underlying traffic patterns would not make the static model as reliable as the dynamic model that can adapt to the different equilibrium relations easily without great loss of accuracy.

4.2.2   Sensitivity of the rolling horizon scheme

First, we examine the sensitivity of model performance to the length of roll period. The adaptive model calibration and the speed prediction are performed using the roll period of from 1-minute up to 30-minute and the calibration horizon is fixed at 60-minute. Figure 4-22 and Figure 4-23 are the RMSEs of speed prediction at the different roll period for the model with the dual-regime and the single-regime MG relation, respectively. The result shows that the longer the roll period (which means less frequent model updating), the lower performance the dynamic model tends to have. When the model calibration is frequent, the parameters keep being updated frequently so as to be able to reflect the most recent dynamics in traffic flow, and at same time increase the computation efforts needed for the repeated calibrations. In our tests, 2.5min or 5min is considered as the most appropriate roll period length for the sufficient quality with the reasonable computation resource.
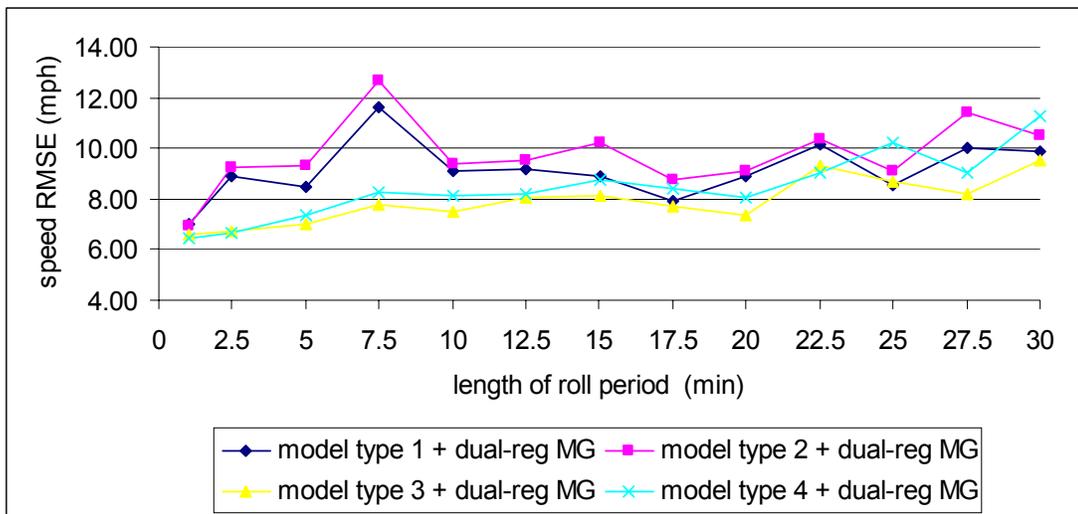
Figure 4-25 Speed RMSEs at Different Roll Periods for Dynamic Models with a Dual-regime MG Relation
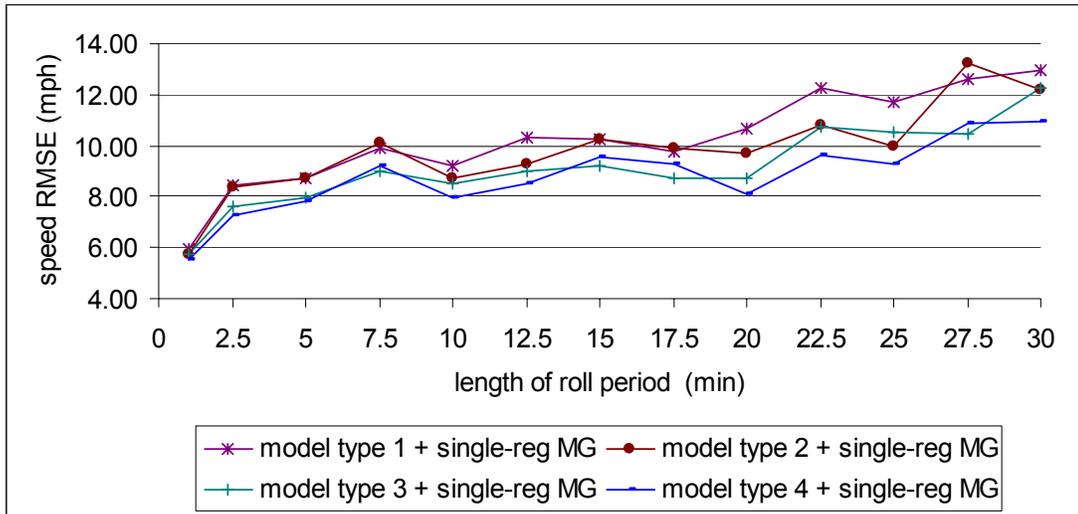
Figure 4-26 Speed RMSEs at Different Roll Periods for Dynamic Models with a Single-regime MG Relation

Then, we examine the sensitivity of model performance to the length of calibration horizon. The adaptive model calibration and speed prediction are performed using the calibration horizon of from 20-minute up to 120-minute and the roll period which is fixed at 2.5-minute. Figure 4-24 and Figure 4-25 are the RMSEs of speed prediction at the different calibration horizon for the model with the Dual-regime and the Single-regime MG relation, respectively. The result shows that the shorter the calibration horizon, the lower performance the dynamic model tends to have. Shorter calibration horizon provides insufficient data points to reveal the underlying dynamics which might be the cause of the lower performance. Longer calibration horizon requires longer computation time. The performances are virtually comparable beyond 60 minutes. However, due to the limitation of finite amount of data in the test, the longer calibration horizon means that fewer data are evaluated and

lower the reliability of the result. Take into account the above aspects, 60-minute is

chosen as the most appropriate length of calibration horizon in the test.
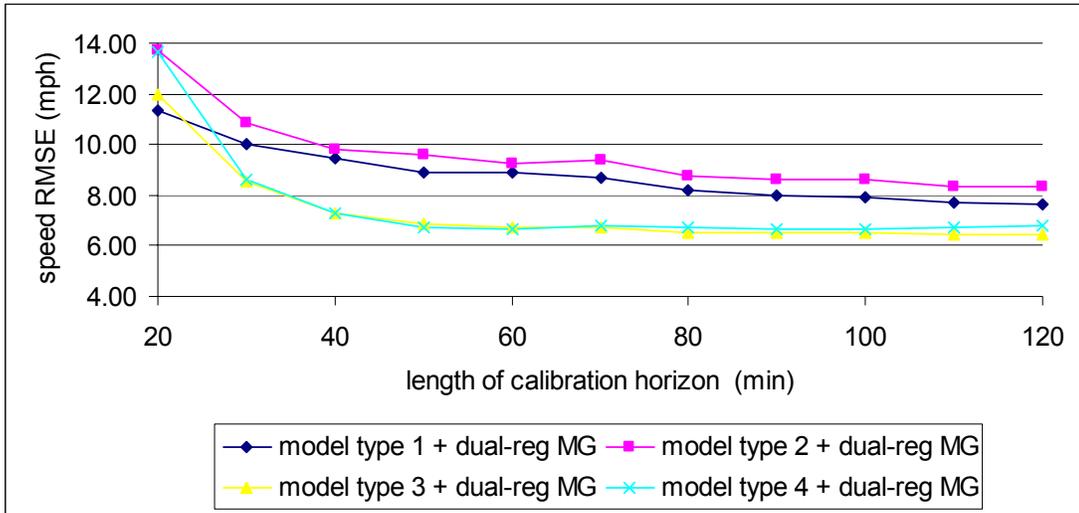
Figure 4-27 Speed RMSEs at Different Calibration Horizons for Dynamic Models with a Dual-regime MG Relation
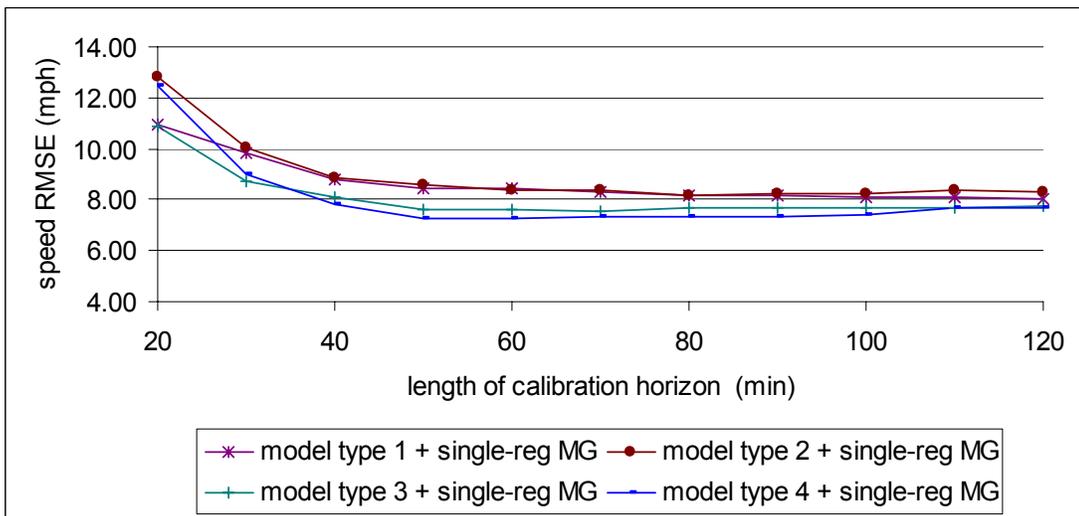
Figure 4-28 Speed RMSEs at Different Calibration Horizons for Dynamic Models with a Single-regime MG Relation

4.3   *<u>Summary</u>*

In this chapter, a set of standalone numerical experiments are designed to evaluate the performance of the model and the adaptive mechanism presented in Chapter 3 under various scenarios. The reliability and robustness of the model as well as the sensitivity analysis of the rolling horizon scheme are comprehensively discussed. The results indicate that the proposed model is robust under different quality of the equilibrium relations. Speed dynamics are mainly affected by speed relaxation and density anticipation for the section A in these standalone tests. Fine-tuned rolling horizon scheme is another important factor to ensure the higher performance of the dynamic model over the static model.

Chapter 5: Traffic Flow Modeling in Real-Time DTA

5.1   *Introduction*

The overall success of the proposed dynamic model is ultimately dependent on the integration with the traffic simulation based system. In the dissertation research, the dynamic model is intended to enhance the real-time modeling capability of the advanced TrEPS, DYNASMART-X. As stated in Chapter 2, the traffic flow modeling in the core simulation module of the DYNASMART relies on the static speed-density relation, namely the Modified Greenshields' model. For the long term planning, the Modified Greenshields' model has been sufficiently satisfactory by reflecting the main trend of traffic flows, as shown in the system evaluation of DYNASMART-P [University of Maryland DTA Group, 2003].   However, DYNASMART-X, with the same core simulation module with DYNASMART-P, is oriented to traffic management applications where short term and real-time traffic monitoring and operational controls are required. In such cases, the static speed-density relation would lack in response to the actual traffic occurring. Traffic flow modeling capability would be improved when quasi-continuous real-time traffic data are available. Under the motivation, the dynamic model proposed in the dissertation research aims to enhance the capability through an adaptive learning process from the real-time data.

One of the problems in implementing the dynamic model in the traffic simulation model is how to deal with time resolution differences in the observation (data sampling) interval and the simulation interval. The usual observation interval

ranges from 20 seconds to 5 minutes, while the mesoscopic type traffic simulation model uses a much finer time scale to track the vehicle trajectories, like 0.1 minute in DYNASMART. Thus, the direct application of the dynamic model based on relatively longer time intervals to the traffic simulation based on shorter updating steps would be inappropriate. Section 5.2 will provide a solution which is based on the connection between continuous dynamic system and discrete dynamic system.

As mentioned in Section 2.3.3.2, the algorithmic component which is the main entity of DYNASMART-X includes the traffic state estimation module and the traffic state prediction module. The purpose of the state estimation module is to estimate the prevailing traffic state in the network. The state prediction module, on the other hand, periodically provides future network traffic states with the descriptive and the normative capabilities for a pre-defined horizon through the rolling horizon operational mode. At the start of each state prediction stage, the predictor reads the current network conditions from the real time estimator and uses the predicted time-varying origin-destination traffic demand values to predict network conditions over the next stage. The reliability of the traffic state estimation is a critical issue for the real-time traffic management and control. If the traffic network state can not be estimated with a reasonable accuracy, any predictions produced based on that can be wrong. This can result in erroneous traffic advisory or control strategies that may cause the network to perform worse than the one without a real-time system in place.

Although a great deal of dedicated endeavors is always made to remove causes of inconsistency, the traffic simulation may not fully replicate the real-world. Despite the best efforts, there remains a tendency for the simulation system to wander

95

off actual situation. For instance, the speed, which is used to move vehicles and determine shortest paths, and the density, which are the outcome of vehicle movements and path selection in the simulation, are usually different from the observed speed and density measured by sensors. The reasons could be unknown or uncontrolled factors in the entire traffic network models and algorithms. In such circumstances, some process adjustment or regulation might be necessary to compensate for the deviation. Otherwise, it could occur that if left to itself, the entropy or disorganization of any system can never decrease and will usually increase, which is implied by one of the fundamental physical laws — the second law of thermodynamics [Kuhn, 1978]. A procedure called short term correction is therefore presented to identify discrepancy between real-time simulation and real-world and use it to direct the reduction of inconsistency. The procedure is based on feedback control theory and will be described in Section 5.3.

5.2   *Different time scales*

5.2.1   Continuous dynamic system vs. discrete dynamic system

In the research, the dynamic model proposed in Chapter 3 is in the discrete form. The discretisation is made based on the momentum equation in the higher order continuum model. The momentum equation in fact is a continuous differential equation consisting of the derivatives with respect to space and time. In many cases, a discrete dynamic system corresponds to an underlying continuous dynamic system. Therefore, for a discrete transfer-function equation in a general form of

$$Y_t = \sum_{j=0}^{\infty} v_j X_{t-j}, \qquad\qquad\qquad [5\text{-}1]$$

the corresponding continuous transfer-function equation can be written in the form

$$Y(t) = \int_0^{\infty} v(u)X(t-u)du . \qquad\qquad [5\text{-}2]$$

where,

Y_t is the discrete system output. Its corresponding continuous variable is $Y(t)$.

X_t is the discrete system input. Its corresponding continuous variable is $X(t)$.

v_j is the discrete impulse response of Y_t to one unit of X_t. Its corresponding continuous variable is $v(u)$.

Typically, the discrete observations are a stepwise approximation of the underlying continuous variable (Figure 5-1). If the record is read at a sufficiently small sampling interval so that sudden changes do not occur between the sampled points, the approximation will be very close to the continuous one. In this case, the different techniques, such as middle point approximation, mean value approximation and instantaneous point approximation, will result in comparable approximations.

It is reasonable to assume that the discretised impulse responses (v_j) estimated from the discretised observations (Y_t and X_t) are also the approximation of the actual continuous impulse responses $v(u)$.
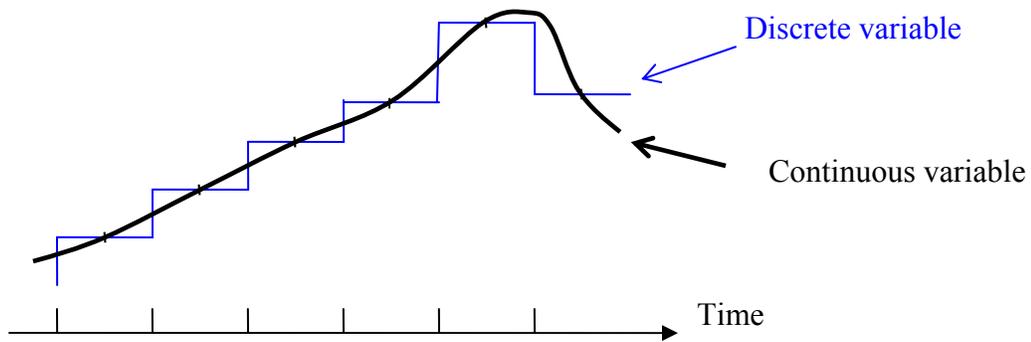
Figure 5-29 Replacement of Continuous Variable by Discrete Variable

5.2.2   Model implementation in DTA-type traffic simulation

In Figure 5-1, the time step length could be varied depending on the system under study. In the typical traffic surveillance system, the step length ranges from 20 seconds to 5 minutes. In the traffic simulation system, the step length is a small value (e.g. 0.1 minute in DYNASMART). To apply the dynamic traffic flow model, we seek to infer the stepwise impulse responses at the simulation interval from the stepwise impulse responses obtained at the observation interval. Based on the discussion of continuous and discrete systems in Section 5.2.1, a straightforward solution is proposed and involves a simple procedure including:

1. approximating the continuous function by connecting the middle points of the discrete function at all steps; then

2. re-discretising the resulting function into a new stepwise function with a smaller step length.

Obviously, a piecewise linear function is resulted in the first step and new stepwise impulse responses are obtained in the second step (Figure 5-2).

Therefore, the new stepwise impulse responses can be computed given the values of the prior stepwise impulse responses and the ratio of two time units.
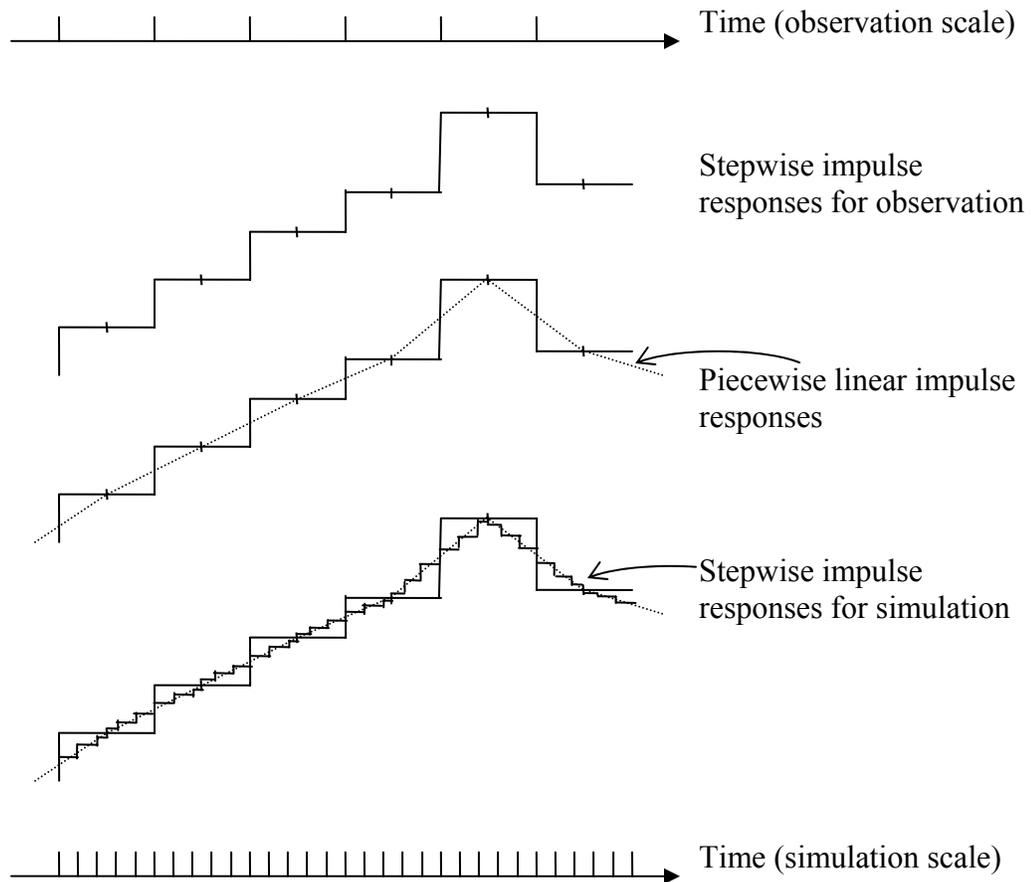
Figure 5-30 Impulse Response Conversion

For example, in the Irvine network which is the sample network in the dissertation, the freeway observation interval is 30 seconds; while the simulation interval in DYNASMART is 6 seconds. Therefore, the ratio is 5. By using the procedure stated, the following converting equations are used to obtain the new

99

impulse responses, v_{s-4}, v_{s-3}, v_{s-2}, v_{s-1}, and v_s, given the prior impulse responses,

v_h, v_{h-1}, and v_{h+1}, as shown in Figure 5-3.

$$v_{s-4} = (3v_h + 2v_{h-1})/5 \qquad \text{[5-3]}$$

$$v_{s-3} = (4v_h + v_{h-1})/5 \qquad \text{[5-4]}$$

$$v_{s-2} = (38v_h + v_{h-1} + v_{h+1})/40 \qquad \text{[5-5]}$$

$$v_{s-1} = (4v_h + v_{h+1})/5 \qquad \text{[5-6]}$$

$$v_s = (3v_h + 2v_{h+1})/5 \qquad \text{[5-7]}$$

where

H is the time scale for observations;

S is the time scale for simulations;

h is a specific time point on H;

s is a specific time point on S; and

h and s represent the exactly same time point.
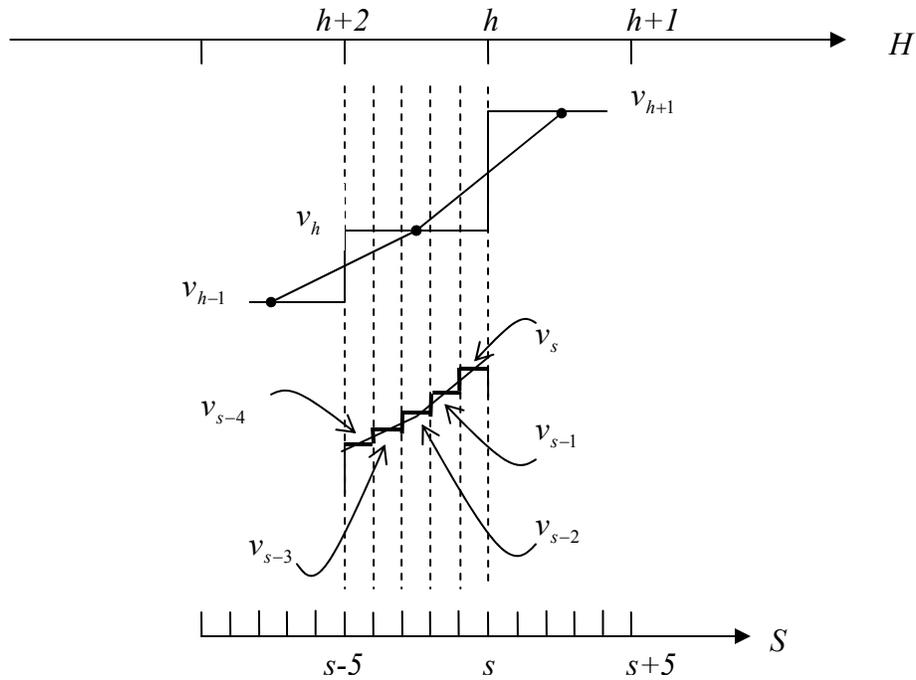
Figure 5-31 An Example for Impulse Response Conversion

5.3   *Short term correction*

5.3.1   Feedback control

The principle of feedback in control systems is to compare the signal to be controlled to a desired reference signal and the discrepancy is used to compute corrective control action [Doyle et.al. 1992]. A general system of feedback control is shown in Figure 5-4. The process is affected by a disturbance which in the absence of corrective control action would cause the output quality characteristic to deviate from target by an amount Δ_t. Thus $\{\Delta_t\}$ is a time series exemplifying what would happen at the output if no control were applied. In fact, a corrective control action C_t

can be manipulated to cancel out this disturbance as far as is possible. When C_{t-1} is

applied onto the system at time t-1, a compensation e_t will be produced at time t.

Hence, the final deviation is $\varepsilon_t = \Delta_t + e_t$ (definition of ε_t is the difference of output

and target value), which is the part of errors that controller fails to cancel out. The

subsequent control action C_t is dependent on present and past errors $(\varepsilon_t, \varepsilon_{t-1}, \cdots)$.
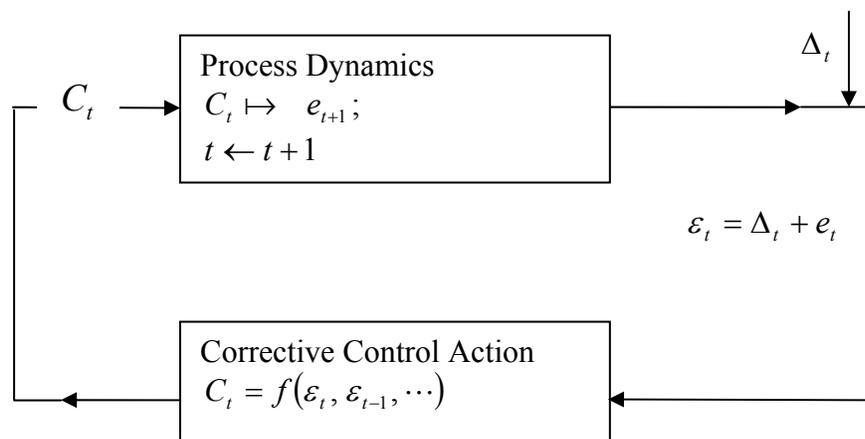
Figure 5-32 Feedback Control Loop

Among feedback control techniques, PID (proportional-integral-derivative)

control has been used in the process industries for many years. Controller of this kind

is a three term controller, which means that, if ε_t is the error at the output at time t,

the control action could be made proportional to ε_t itself, to its integral with respect

to time, or to its derivative with respect to time. A PID controller uses a linear

combination of these modes of control action, say

$$C_t = k_0 + k_D \frac{d\varepsilon_t}{dt} + k_P \varepsilon_t + k_I \int \varepsilon_t dt \qquad [5\text{-}8]$$

Its associated discrete analog is

$$C_t = k_0 + k_D \nabla \varepsilon_t + k_P \varepsilon_t + k_I \sum_{i=1}^{t} \varepsilon_i \qquad\qquad \text{[5-9]}$$

where k_0, k_D, k_P, and k_I are constants.

Usually only one or two of these three modes of action are used, such as integral control and proportion-integral (PI) control.

The principle of PI (proportional-integral) control mechanism is used for the short term correction for online traffic estimation and prediction. It is in fact one of the special cases of [5-9] if only k_P and k_I are nonzero ($k_D = 0$). Since no correction is made at the very beginning stage $t = 0$, $k_0 = 0$. Hence, we have the following correction term C_t

$$C_t = k_P \varepsilon_t + k_I \sum_{i=1}^{t} \varepsilon_i \qquad\qquad \text{[5-10]}$$

Here, C_t is intended to make adjustment to the speed in the upcoming simulation interval right after the observation.

Before continuing with the description of the consistency checking in online application, Figure 5-5 depicts the labels on the time axis in terms of observation interval τ and simulation interval κ. m is the ratio of the observation interval length to the simulation interval length.

103

Observation interval τ

t   t+1

Time

Simulation interval κ

(t-1)m+1 (t-1)m+2 (t-1)m+ m-1 (t-1)m+ m t·m+1 t·m +2 t·m + m-1 t·m + m
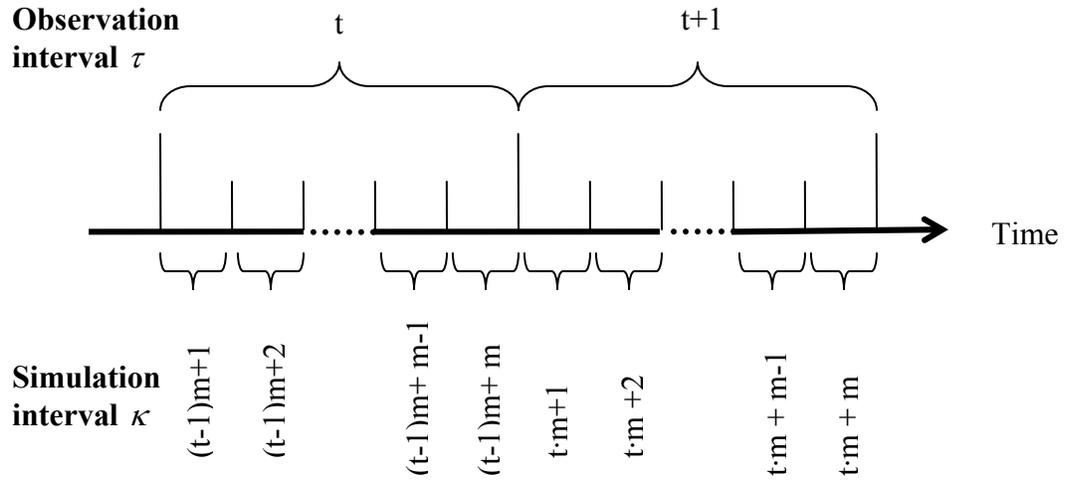
Figure 5-33 Observation Interval vs. Simulation Interval

5.3.2 Speed-deviation-triggered CCU

For an online simulation-based traffic estimation and prediction system, the speed-triggered short term correction aims to compute the discrepancies between simulated speeds and observed speeds and direct the *correction to speed* by the latest and past discrepancies.

Suppose, at the end of the observation time t, an actual traffic speed $\bar{u}_t^{obs}$ is observed and a speed $\bar{u}_t^{sim}$ can be computed as the average value of the simulated speeds u_k^{sim}'s (where $k = (t-1) \cdot m + 1, (t-1) \cdot m + 2, \cdots\cdots, (t-1) \cdot m + m$) in the simulation intervals included in that observation interval t. So, the deviation for time t is $\varepsilon_t = \bar{u}_t^{obs} - \bar{u}_t^{sim}$. The series $\{\varepsilon_t\}$ forms the basis for corrective action C_t as in [5-10]. To be able to be implemented in the online operation, the integral part in [5-10] is reformulated to a truncated summation from t-h to t as in [5-4], where h represents the number of the significant past intervals.

$$C_t = k_P \varepsilon_t + k_I \sum_{i=t-h}^{t} \varepsilon_i \qquad\qquad\qquad \text{[5-11]}$$

The corrective adjustment is applied to the simulated speed in the simulation interval right after the latest observation, say $k = t \cdot m + 1$.

$$u_{tm+1}^{sim} \leftarrow u_{tm+1}^{sim} + C_t \qquad\qquad\qquad \text{[5-12]}$$

The corrected speed u_{tm+1}^{sim} is expected to serve as a better starting point after time t compared to the original speed without adjustment. The speeds afterwards (u_k^{sim}'s, where $k = t \cdot m + 2, t \cdot m + 3, \cdots\cdots, t \cdot m + m$) before the next adjustment will be the outcome of the applied traffic simulation and traffic flow models.

5.3.3   Density-speed-deviation-triggered CCU

In contrast to the speed-deviation-triggered short term correction, the discrepancies under study is not only for speeds but also for densities in the density-speed-deviation-triggered short term correction. The combined discrepancies are used to determine the *correction to speed*.

Suppose, at the end of the observation time t, an actual traffic density $\bar{k}_t^{obs}$ and speed $\bar{u}_t^{obs}$ are observed. A density $\bar{k}_t^{sim}$ and a speed $\bar{u}_t^{sim}$ can be computed as the average value of the simulated densities k_k^{sim}'s and speeds u_k^{sim}'s (where $k = (t-1) \cdot m + 1, (t-1) \cdot m + 2, \cdots\cdots, (t-1) \cdot m + m$) in the simulation intervals included in that particular observation interval t. So, the deviation of density for time t is $\varepsilon_{k,t} = \bar{k}_t^{obs} - \bar{k}_t^{sim}$ and the deviation of speed for time t is $\varepsilon_{u,t} = \bar{u}_t^{obs} - \bar{u}_t^{sim}$. The

series $\{\varepsilon_{k,t}\}$ and $\{\varepsilon_{u,t}\}$ form the basis for corrective action C_t. Similar to the previous case, a truncated summation from t-h to t is used for the integral control term, where h represents the number of the significant past intervals [5-13].

$$C_t = k_{k,P}\varepsilon_{k,t} + k_{k,I}\sum_{i=t-h}^{t}\varepsilon_{k,i} + k_{u,P}\varepsilon_{u,t} + k_{u,I}\sum_{i=t-h}^{t}\varepsilon_{u,i} \qquad [5\text{-}13]$$

Obviously, the corrective action C_t considers the deviations on both density and speed and totally four factors $k_{k,p}$, $k_{k,I}$, $k_{u,p}$, and $k_{u,I}$ are required. The corrective adjustment is applied to the simulated speed in the simulation interval right after the latest observation, say $k = t \cdot m + 1$.

$$u_{tm+1}^{sim} \leftarrow u_{tm+1}^{sim} + C_t \qquad [5\text{-}14]$$

The corrected speed u_{tm+1}^{sim} is expected to serve as a better start point after time t compared to the original speed without adjustment. The speeds afterwards (u_k^{sim}'s, where $k = t \cdot m + 2, t \cdot m + 3, \cdots\cdots, t \cdot m + m$) before the next adjustment will be the outcome of the applied traffic simulation and traffic flow models.

5.3.4 Adaptive estimation of control factors

To determine the best estimates of the control factors k_p and k_I in [5-11] or $k_{k,p}$, $k_{k,I}$, $k_{u,p}$, and $k_{u,I}$ in [5-13], adaptive calibration is a preferred approach. The parameter estimation is performed using the least-squared-error criterion, which seeks to minimize the discrepancies between the estimated speeds from the model and the measured speeds. The optimization problem can be expressed as follows:

$$\min_{K_{k,P},K_{k,I},K_{u,P},K_{u,I}} w\left(\frac{\overline{u}_t^{sim} - \overline{u}_t^{obs}}{\overline{u}_t^{obs}}\right)^2 + (1-w)\left(\frac{\overline{k}_t^{sim} - \overline{k}_t^{obs}}{\overline{k}_t^{obs}}\right)^2 \qquad [5\text{-}15]$$

$$\text{s.t. } u_{(t-1)m+1}^{sim} \leftarrow u_{(t-1)m+1}^{sim} + C_{t-1} \qquad [5\text{-}16]$$

$$C_{t-1} = K_P \varepsilon_{t-1} + K_I \sum_{i=t-1-h}^{t-1} \varepsilon_i \qquad [5\text{-}17a]$$

$$\text{or } C_{t-1} = K_{k,p}\varepsilon_{k,t-1} + K_{k,I}\sum_{i=t-1-h}^{t-1}\varepsilon_{k,i} + K_{u,p}\varepsilon_{u,t-1} + K_{u,I}\sum_{i=t-1-h}^{t-1}\varepsilon_{u,i} \qquad [5\text{-}17b]$$

$$\left(u_{(t-1)m+2}^{sim}, \cdots\cdots, u_{(t-1)m+m}^{sim}\right) \leftarrow \text{DTA simulation} \qquad [5\text{-}18a]$$

$$\text{or } \begin{pmatrix} k_{(t-1)m+1}^{sim}, k_{(t-1)m+2}^{sim}, \cdots\cdots, k_{(t-1)m+m}^{sim} \\ u_{(t-1)m+2}^{sim}, \cdots\cdots, u_{(t-1)m+m}^{sim} \end{pmatrix} \leftarrow \text{DTA simulation} \qquad [5\text{-}18b]$$

$$\overline{u}_t^{sim} = \text{average}\left(u_{(t-1)m+1}^{sim}, u_{(t-1)m+2}^{sim}, \cdots\cdots, u_{(t-1)m+m}^{sim}\right) \qquad [5\text{-}19a]$$

or

$$\overline{u}_t^{sim} = \text{average}\left(u_{(t-1)m+1}^{sim}, u_{(t-1)m+2}^{sim}, \cdots\cdots, u_{(t-1)m+m}^{sim}\right)$$
$$\text{and} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad [5\text{-}19b]$$
$$\overline{k}_t^{sim} = \text{average}\left(k_{(t-1)m+1}^{sim}, k_{(t-1)m+2}^{sim}, \cdots\cdots, k_{(t-1)m+m}^{sim}\right)$$

where [5-17a], [5-18a] and [5-19a] are for the speed-deviation-triggered consistent checking, and [5-17b], [5-18b] and [5-19b] are for the density-speed-deviation-triggered consistent checking. w and $(1-w)$ in [5-15] are the weights applied to deviation of speed and density respectively, and $0 \leq w \leq 1$.

In the optimization problem, the parameters (K_P, K_I) or $(K_{k,P}, K_{k,I}, K_{u,P}, K_{u,I})$ are viewed as decision variables given the observation at time t. [5-16] are used to adjust the speed $u_{(t-1)m+1}^{sim}$ in the simulation interval right after the observation interval t-1, whereby the adjustment C_{t-1} is determined by [5-

17a] or [5-17b]. The speeds afterwards $\left(u_{(t-1)m+2}^{sim}, \cdots\cdots, u_{(t-1)m+m}^{sim}\right)$ before the next adjustment are obtained through DTA simulation [5-18a]. [5-18b] is to obtain $\left(u_{(t-1)m+2}^{sim}, \cdots\cdots, u_{(t-1)m+m}^{sim}\right)$ and $\left(k_{(t-1)m+1}^{sim}, k_{(t-1)m+2}^{sim}, \cdots\cdots, k_{(t-1)m+m}^{sim}\right)$ through DTA simulation. $u_{(t-1)m+1}^{sim}$ together with $\left(u_{(t-1)m+2}^{sim}, \cdots\cdots, u_{(t-1)m+m}^{sim}\right)$ are used to compute an average speed $\bar{u}_t^{sim}$ for time t in [5-19a] and [5-19b]. $k_{(t-1)m+1}^{sim}$ together with $\left(k_{(t-1)m+2}^{sim}, \cdots\cdots, k_{(t-1)m+m}^{sim}\right)$ are used to compute an average speed $\bar{k}_t^{sim}$ for time t in [5-19b].

In the context of online operational application of DTA simulation modeling, the above optimization is performed adaptively when new real-time traffic data become available. One of the constraints in the optimization problem is to satisfy the network dynamic traffic assignment requirements; thereby the problem is more like a bi-level optimization problem which might need iterative evaluations to obtain the best estimates. Figure 5-6 is the algorithmic procedure for the short term correction with the adaptive factor estimation.
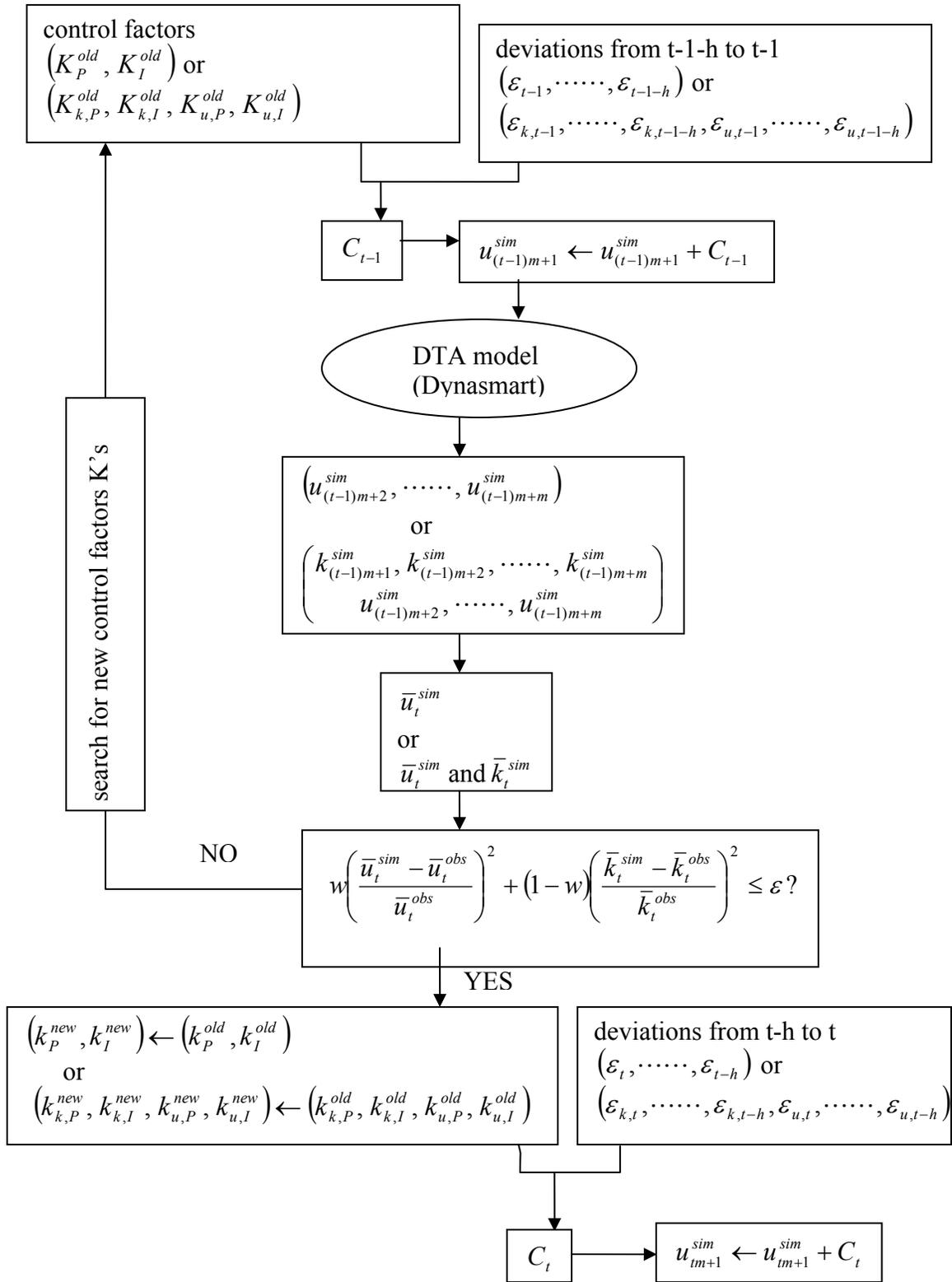
Figure 5-34 Algorithmic Procedure for the Short Term Correction with the Adaptive Factor Estimation

5.4   *<u>Summary</u>*

This chapter basically describes two problems encountered when the proposed dynamic traffic flow model is applied in the mesoscopic type DTA system. The first problem is how to deal with the different time scales between the data sampling interval and the simulation updating step. It is solved by assuming that two discrete dynamic systems are related to the same underlying continuous dynamic system. The second problem is how to lower the inconsistency caused by unknown and uncontrolled reasons. To solve the problem, the short term correction procedure is formulated to identify discrepancies between real-time simulated values and real-world observations and adjust speeds periodically. The adjustment could be triggered by either speed-deviation or density-speed-deviation.   The associated tuning parameters can be estimated adaptively using the least-square-error method which is constrained by DTA simulation model to reach the internal and external consistency. The model performance in real-time DTA operation will be evaluated through the comprehensive laboratorial tests based on real data in the following chapter.

Chapter 6: Performance Analysis of Real-Time Traffic Flow Model

6.1   *Introduction*

Chapter 6 discusses the numerical experiments conducted to evaluate and analyze the performance of the proposed real-time traffic flow model under various operational application scenarios. DYNASMART-X is used as the traffic simulation platform within which the proposed model is implemented.

As mentioned in Chapter 2, the current (default) traffic flow model in DYNASMART-X is based on a static speed-density relation, the Modified Greenshields' model. The dynamic model proposed in the dissertation is intended to replace the static model to produce link speed predictions for every simulation updating interval. The calibration of the proposed dynamic model requires actual traffic sensor data to be available quasi-continuously for links. It is unlikely that all the links in a network would be under surveillance; therefore the adaptive model calibration and speed prediction are only possible for those observed links. Accordingly, the first step of the performance analysis of the dynamic model is to evaluate system performance, particularly estimation and prediction accuracy, under scenarios of partial implementation of the dynamic model. The evaluation is meant to study the compatibility between the two kinds of models in determining the estimation and prediction quality for a transportation network where the dynamic model is applied to those links with real-time observations, while the static model is still used for those links without observations in the DTA simulation.

Although they lack quasi-continuous surveillance data, the unobserved links essentially share a similar traffic pattern (under consistent traffic facility environment and travel behavior) as the neighboring links with observations. Therefore, a second important focus is the model transferability, namely the extent to which the adaptively calibrated model from the observed links can be effectively borrowed to those unobserved links to predict speeds, instead of using the static models, in the DTA simulation.

Under real-time DTA operation, traffic simulation is performed for both traffic static estimation and prediction which are the most essential functions of the DYNASMART-X prototype. The traffic estimation function synchronizes its step with the real clock and estimates the concurrent traffic states. The traffic estimation function uses pre-calibrated or online calibrated OD demand desires over the network in a one-shot simulation-assignment mode. A robust and reliable traffic flow model is expected to improve the estimation capability. Moreover, the concurrency between the traffic estimation function and the real world makes possible the short term correction function, which would boost the estimation capability to a higher level. Observed deviations of the simulated values from the actual values provide a basis for the correction to the next link speeds. The short term correction procedure described in Chapter 5 is to be implemented jointly with the proposed dynamic traffic flow model to minimize the deviations between the estimated and the real system state variables. The effect of the short term correction on the model performance will be tested and discussed.

The traffic state prediction provides the network traffic states for a pre-defined prediction horizon. Every time it executes, it produces the current state of the network at the start of the new stage and travel demand predictions for this new stage. It can operate in either descriptive or normative mode. When operating in the descriptive mode, it presents only the traffic evolutions consistent with users' behaviors for the next prediction horizon. When operating in the normative mode, it utilizes the iterative Multiple User Class (MUC) algorithm and solves for optimal routing policies and other possible network management strategies.

The traffic state prediction is implemented in DYNASMART-X based on a rolling horizon approach. In this framework (Figure 6-1), the planning horizon is subdivided into several overlapping *stages*. The consecutive stages overlap at fixed intervals, the length of each is referred to as the *roll period*. The *stage length* (or *horizon*) is denoted by h and the *roll period* is denoted by l. In the example of Figure 6-1, h is set to 20 minutes and l is set to 5 minutes. The roll period l is the short-term future duration for which the forecasts are considered to be reliable. In the remaining part of the stage, $(h–l)$, forecasts become less reliable. In the DYNASMART-X prototype, the predictions for a stage are made at the beginning of the stage. The traffic simulation for the predictions is intended to support the online traffic management decision-making by operators under various control strategies, information dissemination strategies or travel behaviors. As time advances, the traffic state predictions will be repeated periodically. The previous predictions for the overlapping part of the stage, $(h–l)$, will be replaced by the new predictions made at the new active stage. For example, in Figure 6-1, for a duration A, the traffic status is

predicted first at Stage *m*, a second time at Stage *m*+1, third time at Stage *m*+2, and fourth time at Stage *m*+3. The prediction made for time period A in Stage *m* is expected to be less accurate than the prediction made in Stage *m*+3. In the experiments, a comparison between the static model and the dynamic model in terms of prediction capability will be conducted.
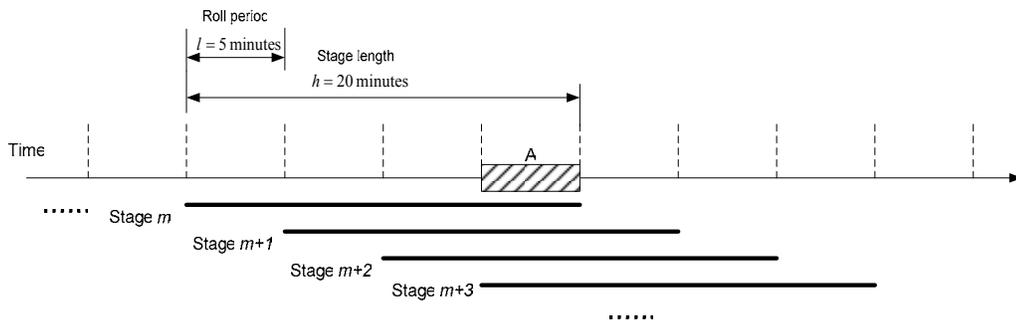
Figure 6-35 Rolling Horizon Procedure

As presented in Chapter 3, the proposed dynamic model is extended from a discretised version of the higher order continuum model. The discretisation scheme of time and space is of importance to the model performance. The rougher the discretisation, the lower the approximation precision is, but the faster the numerical computation could be. Hence, the impacts of the temporal scale of the measurements and the spatial scale of the link segments with observations on the model performance are to be tested and analyzed.

In the following sections, the network and associated data will be described first. Following presentation of the experimental design, the results of the experiments will be discussed.

6.2   Network overview and data description

The Irvine network (Orange County, California) includes two interstate freeways, namely the I-5 and the I-405, as well as part of the state highway 133. The rest of the network consists of arterials and ramps. Figure 6-2 depicts the Irvine network with 326 nodes, 626 links and 61 traffic analysis zones, shown in the GUI of DYNASMART-X.
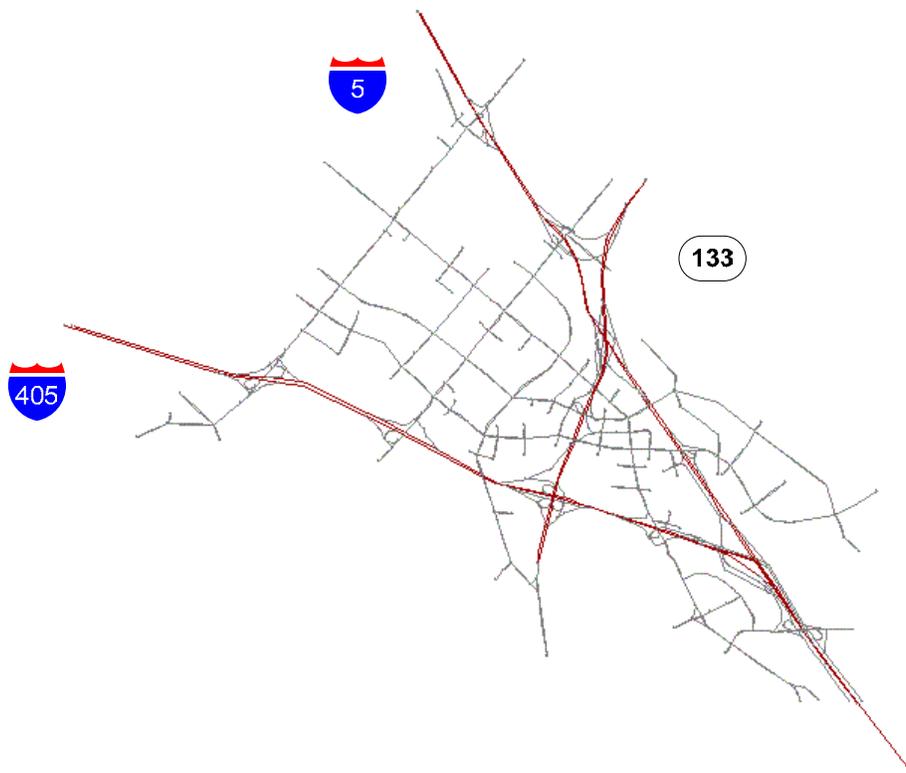
Figure 6-36 Irvine Network Displayed in the DYNASMART-X GUI

Real-time traffic data from the Irvine network has been collected by loop detectors of the Freeway Performance Measurement System (PeMS) in California. The data include freeway traffic volume and occupancy at 30-second intervals. The sampling dates are May 22, 24, 30, June 1 and 5 in 2003. The specific sampling period each day is from 4:00 am to 10:00 am, which basically covers the typical morning rush hours. After the proper data clean-up and checking, 13 freeway links with reliable real-time data are identified, each of which is named and highlighted in Figure 6-3.

The 13 freeway links become the study links in the experiments described hereafter and are divided into five groups according to the freeway location and the traffic direction, see Table 6-1. The table also lists the set(s) of the connected neighboring links for each group. The information is provided to assist in designing the experiments when testing the performance of the four different model types (see Chapter 4 for the specification of the model types). As seen in Table 6-1, there are two sets of two sequential links and one set of three sequential links.
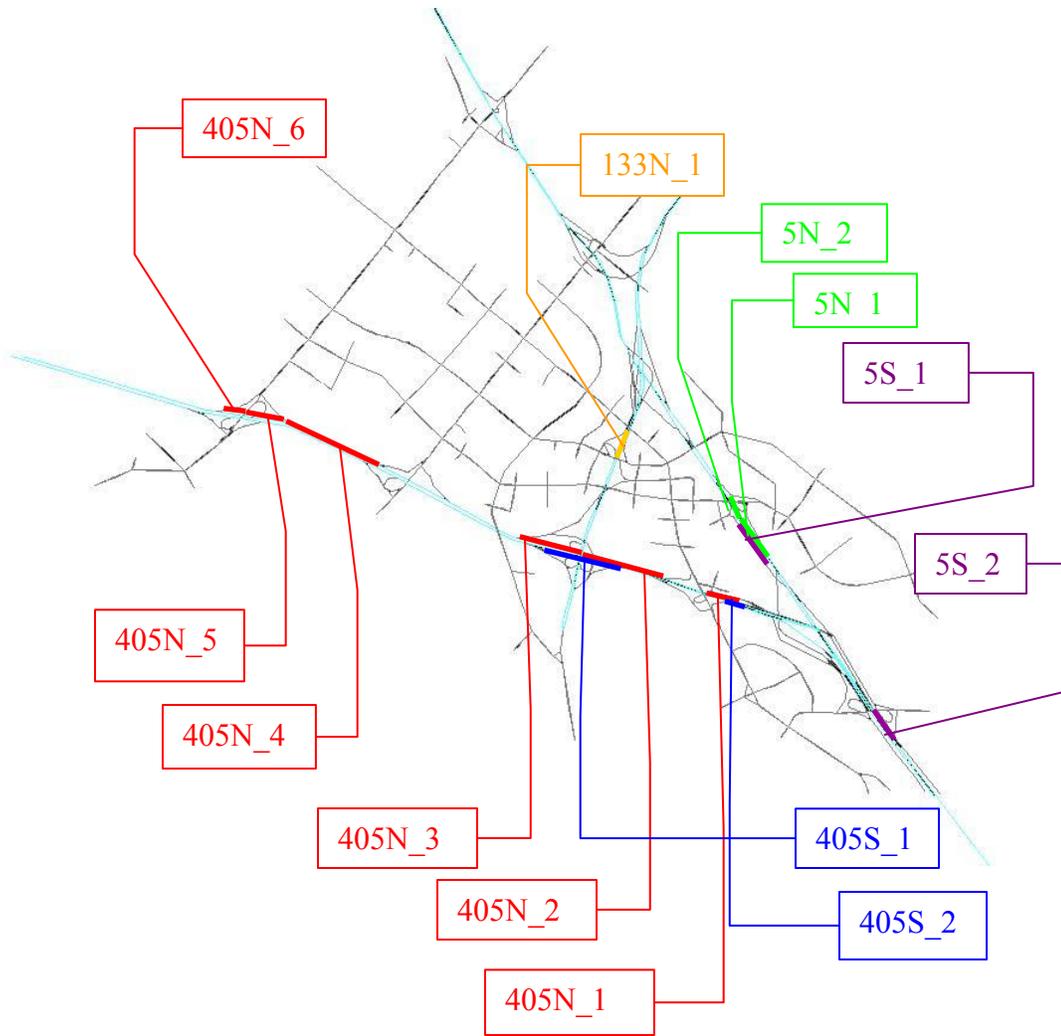
Figure 6-37 Detector Coverage over the Irvine Network

Table 6-1 13 Observed Freeway Links in 5 Groups

Group #	Freeway	Direction	Link names	Connected links
1	I-405	North bound	405N_1, 405N_2, 405N_3, 405N_4,	405N_2 — 405N_3; 405N_4 — 405N_5 — 405N_6

			405N_5, 405N_6	
2	I-405	South bound	405S_1, 405S_2	NA
3	I-5	North bound	5N_1, 5N_2	5N_1 — 5N_2
4	I-5	South bound	5S_1, 5S_2	NA
5	133	North bound	133N_1	NA

Furthermore, the densities are converted from the occupancy data using the following relationship:

$$K = \frac{52.8}{L_v + L_s} * occ(\%)$$

where

K : lane density (in vehicles/lane-mile);

L_v : average vehicle length (in feet);

L_s : average sensor length (in feet); and

occ: occupancy (%).

L_v was assumed to be 5 meters (approximately 16.4 feet), and L_s was set to 2 meters (approximately 6.5 feet).

To obtain the time-series data of the speeds, the flow and density measurements are used to compute the average speeds as $V = Q/K$, where V is the

speed (in mile/hour), Q is the flow (vehicles/lane-hour), and K is the density (in vehicles/lane-mile).

6.3   *Experimental settings, results and discussion*

6.3.1   Experimental settings

Unlike the standalone experiments presented in Chapter 4, the experiments conducted here are intended to investigate the effectiveness of the proposed dynamic traffic flow model and the associated algorithms in the context of real-time DTA operation.   The integration of the dynamic model with the DTA system is accomplished within the DYNASMART-X prototype (see Chapter 2 for the overview of the DYNASMART-X).   The full execution of DYNASMART-X includes six algorithmic modules, i.e., traffic estimation, traffic prediction, OD demand estimation, OD demand prediction, short term consistency checking and long term consistency checking.   However, to rule out the effect of possible modeling imperfections in the modules which are not directly relevant to the traffic flow modeling, the experiments here are primarily conducted by only activating the traffic estimation and the traffic prediction modules, and the short term consistency checking when required by the experiment design.   In other words, the time-varying OD demands are not estimated and predicted online (so the OD estimation and prediction modules are disabled), but calibrated a priori; and no online correction on OD matrices is made (so the long term consistency checking module is disabled).   So, at the beginning of the simulation, the network is empty, with no vehicle occupying any

link. As time goes on, vehicles are generated in the network based on the time-dependent OD demand information available in the data broker. Table 6-2 summarizes the pertinent scheduling parameters applied in the experiment. It defines the module execution frequency and duration as well as observation sampling frequency. The parameter settings could be accomplished through the DYNASMART-X GUI, according to the instructions provided in the user's guide [University of Maryland DTA Group, 2004]

Table 6-2 Scheduling Parameters in DYNASMART-X

Parameter	Value
Assignment Interval	5 min
Observation Interval[*1]	0.5 min
Traffic State Estimation Period	0.5 min
Traffic State Prediction Roll Period	5 min
Traffic State Prediction Horizon	20 min
Traffic Flow Model Updating Period	2.5 min
Traffic Flow Model Calibration Horizon	60 min
Short Term Correction Period[*2]	0.5 min

*1: the observation interval will be changed to be longer than 0.5 min when the impact of temporal scale is tested.

*2: the short term correction module is activated upon request. The period should be same as the length of the observation interval.

As in Chapter 4, the MOE selected is the average root mean squared error (RMSE) of the estimation against the actual observations over the specified

estimation period. The RMSE will be weighted by the number of vehicles in each observation interval. In Chapter 4, given the known and correct values of link densities, we only consider the RMSE of link speeds which are the product of the real-time prediction. However, in the DTA system, the speed is predicted through the traffic flow model which requires the information such as the prevailing density. The predicted speed becomes the factor to determine how fast and along which path to the destination the vehicles should move in the current simulation interval. As a result, the new prevailing density is the outcome of these vehicle movements. Then the speed prediction starts again. The simulated link density will exhibit some discrepancy relative to the observed link density to some extent. This would introduce additional errors in the subsequent prediction of link speeds. Such errors can be kept to a minimum level by a reliable and robust traffic flow model, and the error propagation could be limited by introducing the short term correction process. Therefore, in the experiments conducted here, RMSE is computed for both speeds and densities of the study links.

Again similarly to Chapter 4, the static relation serves two purposes. One is to provide the equilibrium value for the relaxation term in the dynamic model; the other is to provide a standard of comparison in judging the experimental effects of the dynamic model. The static speed-density relations used for links in the Irvine network were previously calibrated for use in an operational planning application of the DYNASMART system.

In the following sections, 6.3.2 through 6.3.7, the experiments, based on the available data, focus on the model compatibility, the model transferability, the traffic

estimation capability with short term correction, the traffic prediction capability, and the impacts of the temporal scale and the spatial scale. The experiment descriptions and results will be presented together with the pertinent discussion and concluding remarks.

6.3.2 Model compatibility

The objectives of the evaluation performed here are 1) to assess whether the partial application of the dynamic traffic flow model over the network provides benefits to the overall traffic state estimation capability of the real-time DTA system; and 2) to assess the relative effect of the different model specifications, with regard to including one or more of the higher-order dynamic effects (i.e., relaxation, convection, and anticipation) based on the partial observations.

The 13 links with the archived traffic data are treated as the study links in the experiments. Either the static model or the dynamic model could be applied to any one of them, depending on the particular experiment design. After completion of the DTA simulation, the RMSEs are calculated for all 13 links, using their actual data. Results from the study links with the static models (called "SM links" hereafter) are generalized to the links without observations in the network; in contrast, the study links that use the dynamic models calibrated from their own real-time data (called "DM links" hereafter) represent the links with observations in the network.

First, the number of DM links in the network is a factor of interest, and Table 6-3 lists the pertinent experiments. Experiment 1 uses the static model for each link in the network, and provides a benchmark against which the subsequent experiments

are compared. Experiments 2 ~ 5 vary the number of DM links from 1 to 13; these links are distributed across the freeways. Experiments 6 ~ 9 vary the number of DM links from 1 to 6, all located on the I-405 freeway northbound. Hence, Experiments 2 ~ 5 and Experiments 6 ~ 9 allow investigating the effect of number of DM links from two perspectives. The former is increasing the number of DM links "latitudinally" (i.e., across the network), while the latter increases them "longitudinally" (i.e., along a particular corridor). In Experiments 2 ~ 9, the Type I dynamic model, which contains the local speed relaxation as the driving force, is used. The selection of the DM links in Experiments 10 ~ 12 is the same as in Experiment 5, where all the study links are DM links. Experiments 10 ~ 12 mainly differ from Experiment 5 in the model specification. Experiment 10 includes the convection effects in the dynamic models for the DM links (i.e., link 405N_3, 405N_5, 405N_6, and 5N_2) whose upstream links are also DM links. Experiment 11 includes the anticipation effects in the dynamic models for the DM links (i.e., link 405N_2, 405N_4, 405N_5, and 5N_1) whose downstream links are also DM links. Finally, Experiment 12 includes the convection or/and the anticipation effects in the dynamic models for the DM links (i.e., link 405N_2, 405N_3, 405N_4, 405N_5, 405N_6, 5N_1, and 5N_2) whose upstream or/and downstream links are also DM links.

Table 6-3 Details of Experiments 1 ~ 12

Experiment #	Feature	No. of DM links	Location(s) of DM links	No. of tests performed
1	SM across the board	0	NA	1
2	DM for 1 link	1	on one of the freeway (I-405N, I-405S, I-5N, I-5S, or 133).	13
3	DM for 5 links distributed in the network	5	one link on each freeway (I-405N, I-405S, I-5N, I-5S, and 133).	48
4	DM for 9 links distributed in the network	9	two links on the freeways I-405N, I-405S, I-5N, I-5S; and one link on 133.	15
5	DM for 13 links distributed in the network	13 (all the study links)	six links on the freeways I-405N; two links on the freeways I-405S, I-5N, I-5S; and one link on 133.	1
6	DM for 1 link on a freeway	1	on the freeway I-405N	6
7	DM for 3 link on a freeway	3	on the freeway I-405N	20

8	DM for 5 link on a freeway	5	on the freeway I-405N	6
9	DM for 6 link on a freeway	6 (all the study links on the freeway)	on the freeway I-405N	1
10	Same as Exp.5, but including the convection effect	Same as Experiment 5	Same as Experiment 5	Same as Exp.5
11	Same as Exp.5, but including the anticipation effect	Same as Exp.5	Same as Exp.5	Same as Exp.5
12	Same as Exp.5, but including the convection and anticipation effect	Same as Exp.5	Same as Exp.5	Same as Exp.5

Note: SM stands for the Static Model; DM stands for the Dynamic Model.

First, the test results correspond to the static model performance. Figure 6-4 depicts the hour-by-hour and overall RMSEs of speeds and densities averaged for the 13 study links (all of which use the static models) in Experiment 1. Figure 6-5 shows the frequency histograms of the overall RMSEs of speeds and densities for the 13 study links.

In the remaining experiments involving the dynamic models, comparison with the static models (Experiment 1) is of interest. It is done by calculating the percentage error reduction for speed and density. If the error reduction is positive, the

estimation quality is improved compared to Experiment 1; otherwise, if the error reduction is negative, the estimation quality is worse than Experiment 1.
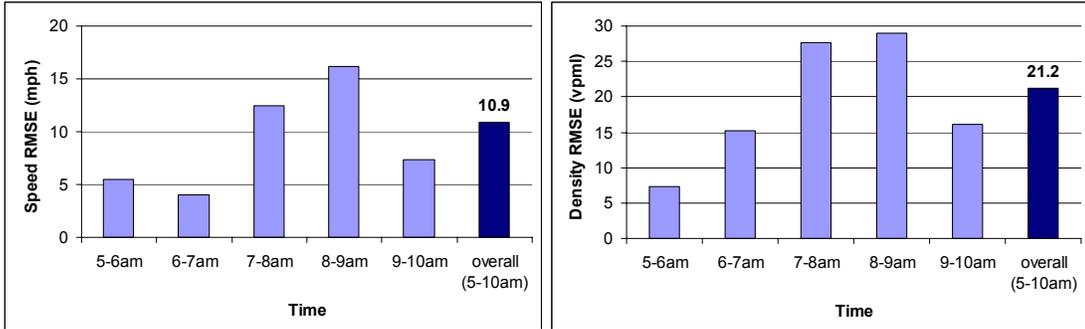

Figure 6-38 Static Models: Average Hour-by-Hour and Overall Errors (Speed and Density) in Experiment 1
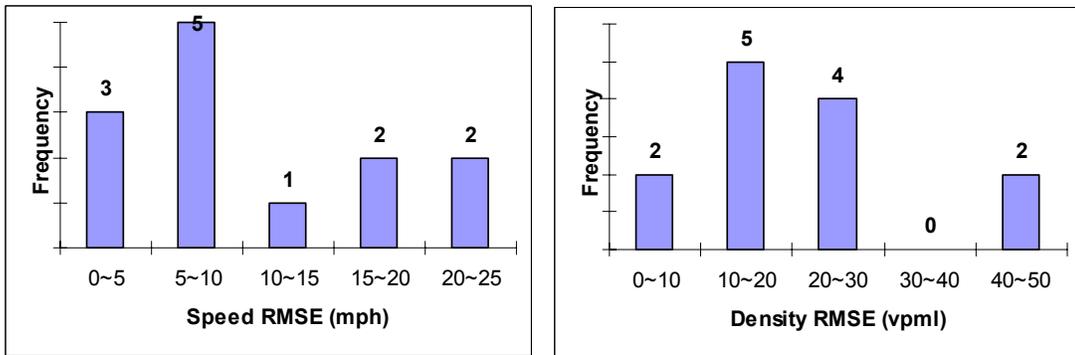

Figure 6-39 Static Models: Frequency Histograms of Overall Errors (Speed and Density) for Study Links in Experiment 1

Figure 6-6 and Figure 6-7 plot the error reductions of speed and density vs. the number of DM links over the network. The study links are divided into DM links and SM links. Introducing a few DM links could affect not only those DM links

themselves but also the other SM links. Therefore, in the figures, besides the average error reduction over all the study links, the error reduction on the SM links and DM links are also shown respectively. Note that in the figures, there is no value for the error reduction for the SM links for Experiment 5, in which all 13 links used the dynamic model specification.

The results reveal error reductions for the estimated speeds and densities of the DM links; these reductions are more apparent as the number of DM links increases. However, when the dynamic model is applied to one link only, the errors on the remaining SM links appear to be worse than under Experiment 1. When the number of DM links increases, such deterioration shows a decreasing trend in the figure. Assuming the trend maintains, there would exist a threshold for the number of DM links in the network, beyond which a positive error reduction for SM links would occur. Due to the data limitation (2% data coverage) in the current study, these experiments did not offer the opportunity to determine the exact number. Considering the average errors for all the study links (SM links and DM links), the network-wide modeling capability is improved with the increasing number of DM links in the network. It should further be noted that the O-D demand was calibrated off-line using a version of the DTA simulator that relies on the static model specification only.

Figure 6-8 depicts frequency histograms of the overall RMSEs of speeds and densities for the 13 study links in Experiment 5. By comparing the results in Figure 6-4 against Figure 6-8 and Figure 6-5 against Figure 6-9, we learn that replacing the

static models with the dynamic models lowers the estimation errors of link speeds and

densities. The errors during the rush-hour period are significantly reduced.
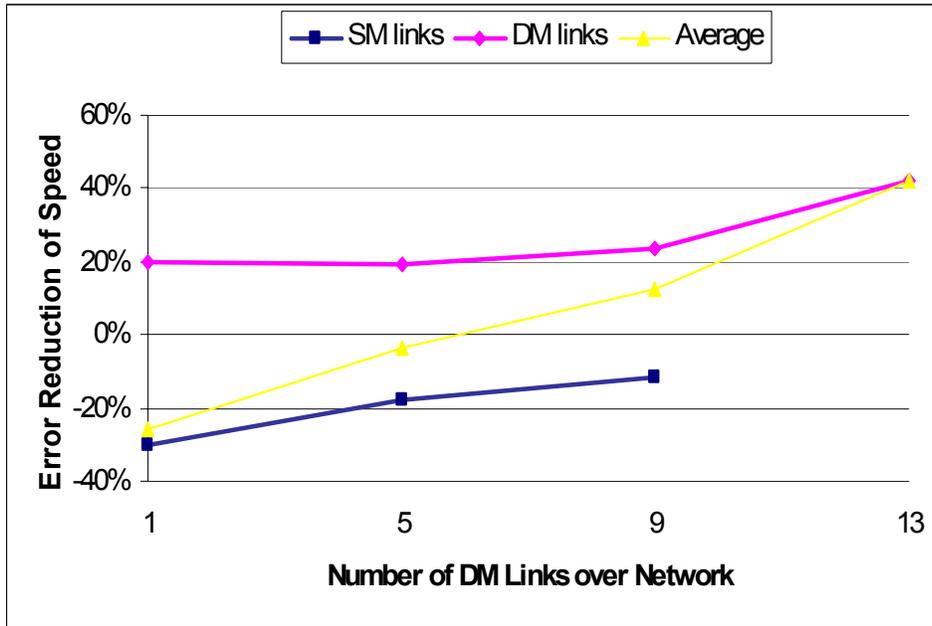

Figure 6-40 Reduction of Speed Error vs. No. of DM Links over Network in Experiments 2 ~ 5
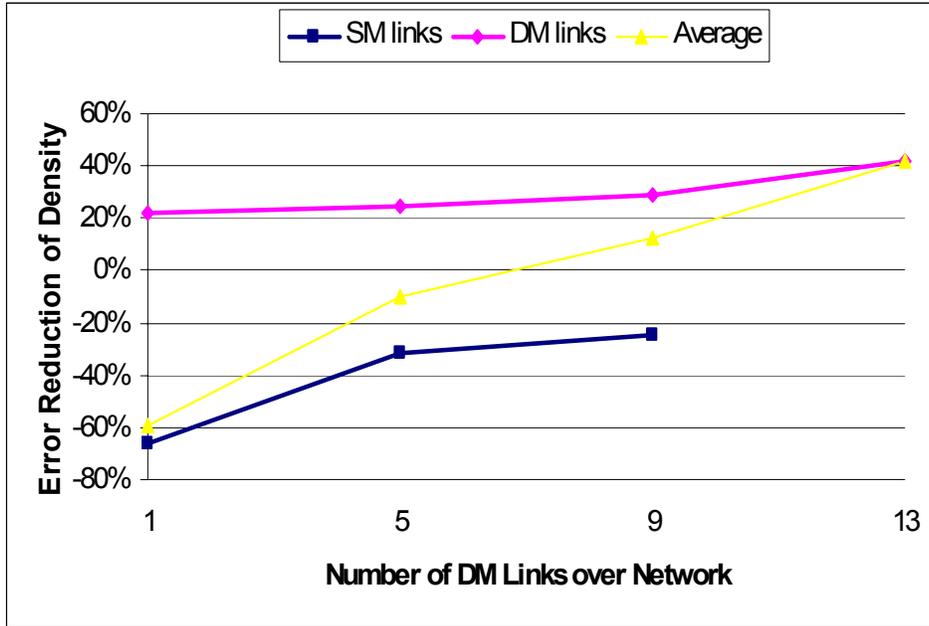
Figure 6-41 Reduction of Density Error vs. No. of DM Links over Network in Experiments 2 ~ 5
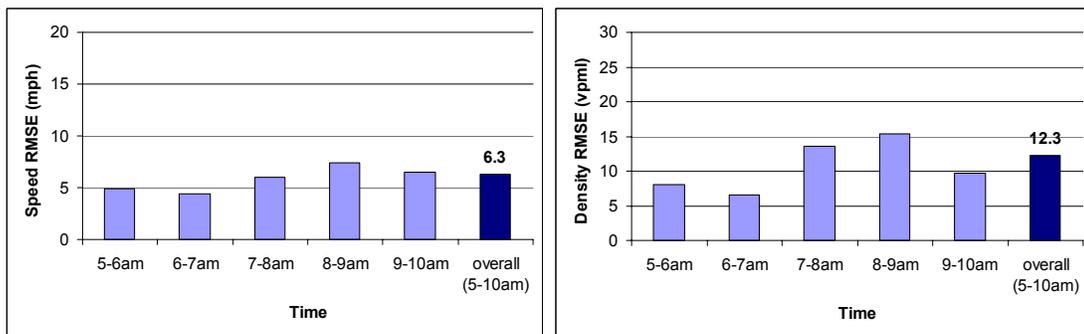

Figure 6-42 Dynamic Models: Average Hour-by-Hour and Overall Errors (Speed and Density) in Experiment 5
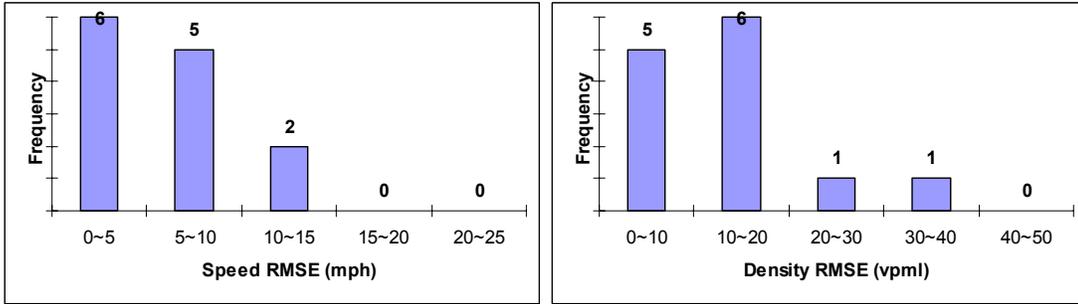
Figure 6-43 Dynamic Models: Frequency Histograms of Overall Errors (Speed and Density) for Study Links in Experiment 5

The plots in Figure 6-10 and Figure 6-11 show the error reduction of speed and density vs. the number of DM links on the I-405 freeway northbound. Error reductions for SM links on I-405N, SM links on other freeways and DM links on I-405N are shown in the figures, together with the average error reduction. Note that in the figures, there is no value of error reduction for SM links on I-405N for Experiment 9 (with 6 DM links), because all the study links on I-405N have been used as DM links. These results confirm that using the dynamic model improves the estimation precision for DM links. The improvement is reinforced with the number of DM links along the freeway. SM links both on I-405N and the other freeways experience increased errors, but the inclusion of more DM links along the same corridor lessens the negative impact. The trends of the error reduction for SM links and all the links with the increasing number of DM links on I-405N are evident in the figures. Existence of a threshold for the number of DM links along a certain corridor, beyond which SM links have reduced errors, is highly possible.
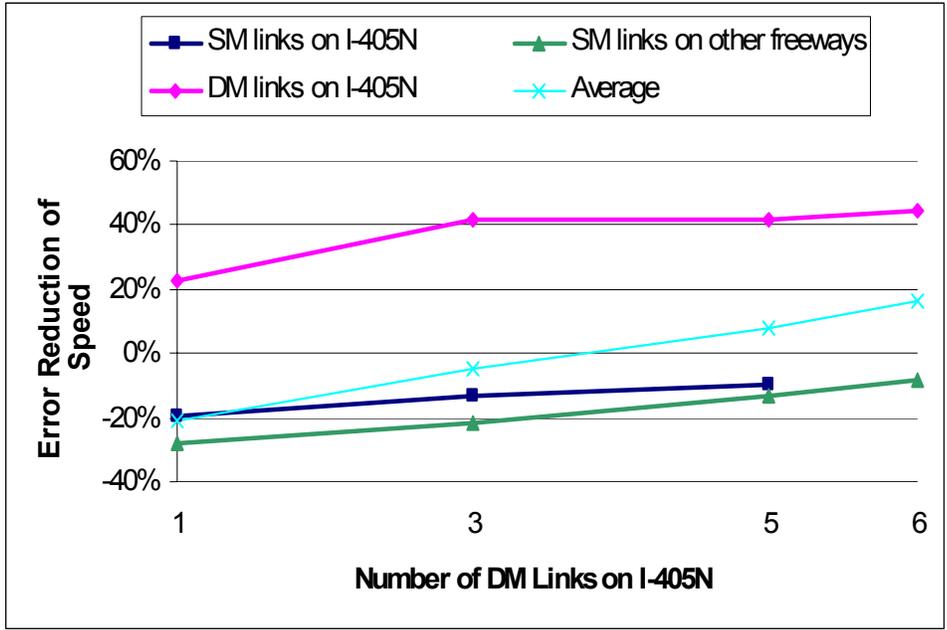
Figure 6-44 Error Reduction of Speed vs. No. of DM Links on I-405N in Experiments 6 ~ 9
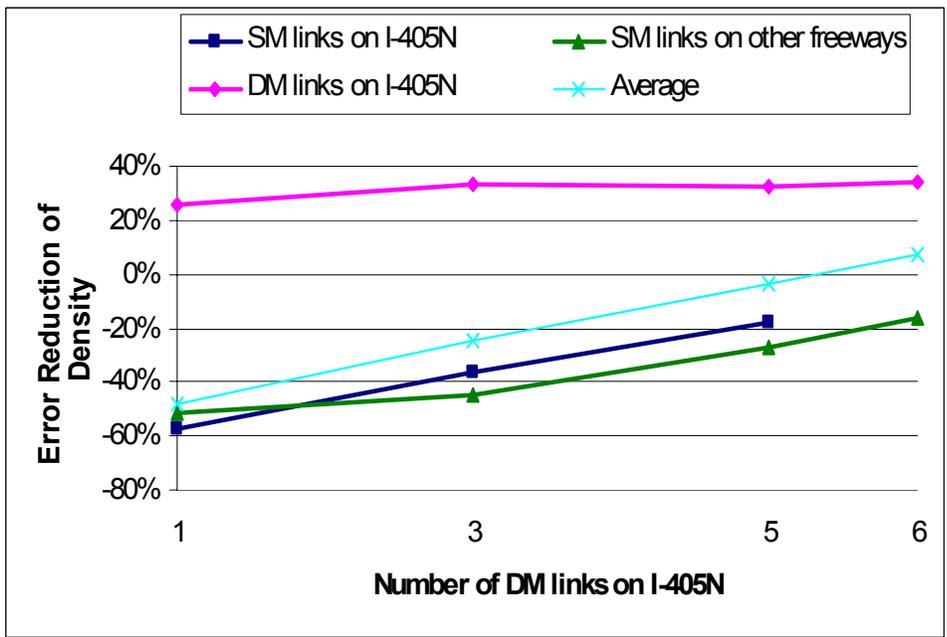

Figure 6-45 Error Reduction of Density vs. No. of DM Links on I-405N in Experiments 6 ~ 9

The above experiment results indicate that the detector coverage in the network is an important factor for the effective application of the dynamic models. An inadequate number of detectors, and more generally inadequate detector coverage, may be of no benefit to, and may even impair, the traffic estimation capability. The impairment is primarily due to the incompatibility of the modeled traffic states between neighboring SM links and DM links, which can extend to the remaining links in the network. Increasing the number of detectors either "latitudinally" or "longitudinally" can reduce the errors gradually. Therefore, an ideal distribution of detectors is the one under which all the major corridors have relatively dense and balanced detector coverage.

Experiments 2 ~ 9 evaluate the model compatibility in terms of the number and distribution of the DM links, while Experiments 10 ~ 12 are meant to assess the compatibility of the different dynamic model specification types, including the different combinations of the higher-order dynamic effects (as defined in Chapter 4).

The plots in Figure 6-12 and Figure 6-13 show the average errors of speeds and densities for different density ranges with the static model and the dynamic models, respectively. "DM I" ~ "DM IV" refer to Experiments 9 through 12 respectively. The results reveal that the application of any of the dynamic model types is beneficial to the traffic estimation capability, because the errors of the speeds and densities are less than under the static model. Experiment 5, in which the relaxation-only model specification is used, exhibits the best error reduction at different congestion levels. When the additional terms (convection or/and

anticipation) are added to the dynamic model of the eligible links in Experiments 10 ~ 12, the estimation quality is not as good as the relaxation-only dynamic model. A likely reason is that there are not many connected links (as indicated in Table 6-1) for which the more complex model specification can be applied extensively, due to the sensor distribution in the Irvine test bed (Figure 6-3). The mixed utilization of the model types does not appear to be especially beneficial to the traffic estimation capability, though the results in this regard are relatively limited. Among model types II, III and IV, the model type III specification, with relaxation and anticipation effects is most effective for speed estimation (Figure 6-12), which is a consistent finding with the standalone tests in Chapter 4. If the majority of the observed links in a network are sequential, it is likely that introducing the anticipation factor would be more effective in improving the estimation quality than the relaxation only scenario.
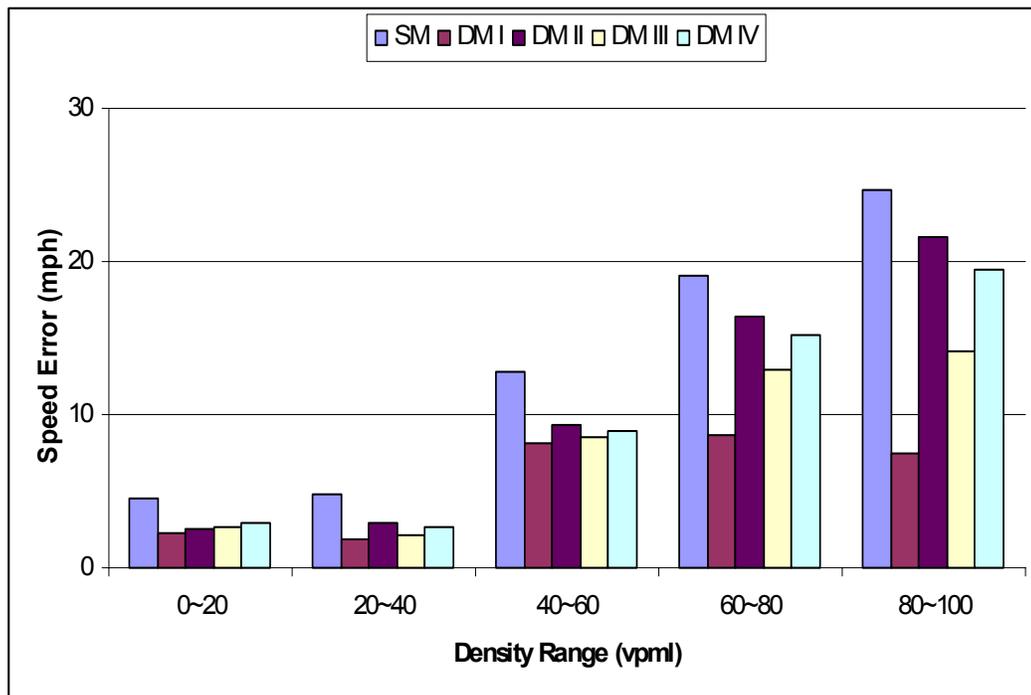
Figure 6-46 Average Speed Error vs. Actual Density Range for Various Models
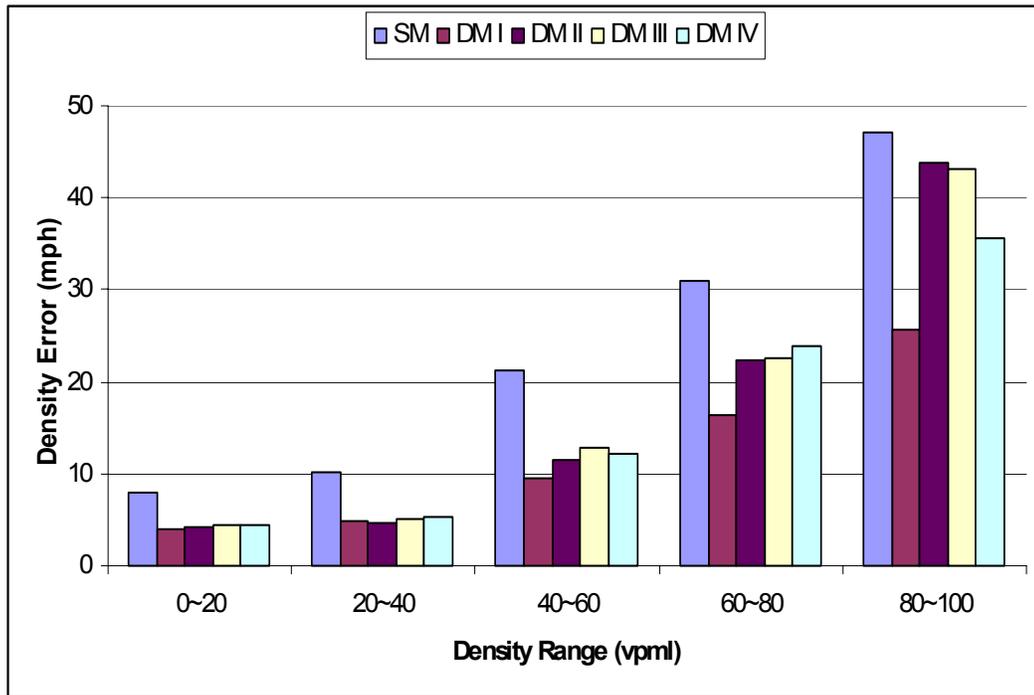
Figure 6-47 Average Density Error vs. Actual Density Range for Various Models

According to the above findings, it is recommended to review the amount, spatial distribution, and the interconnectedness of sensors when deciding on an appropriate traffic flow model strategy for a specific network.

6.3.3   Model transferability

The objective of the evaluation performed in this section is to assess the quality of the traffic estimation when the adaptively calibrated dynamic models are applied to links for which there are no real-time sensor measurements.

Table 6-4 lists the details of the relevant experiments.   For each of the experiments, the DM links are either the own-DM links or the borrowed-DM links.

The own-DM links use the dynamic models adaptively calibrated with their own observed data; while the borrowed-DM links use the dynamic models "borrowed" from some other own-DM links. In Experiment 13, the dynamic model is borrowed from one own-DM link on I-405N to another two study links along the same corridor. Experiment 14 transfers the dynamic model of one own-DM link on I-405N to the remaining five study links along the same corridor. In Experiment 15, three study links are chosen to be own-DM links and three neighboring (upstream or downstream) links to them are chosen to be borrowed-DM links. In Experiment 16, the five own-DM links are distributed over I-405N, I-405S and I-5N and the five borrowed-DM links are located next to the own-DM links. Experiment 17 uses all 13 study links as their own-DM links, so the borrowed-DM links are selected from those non-study links which do not have any archived data but are geographically close to the own-DM links (Figure 6-14).

Table 6-4 Details of Experiments 13 ~ 17

Experiment #	Own-DM link(s)		Borrowed-DM links		No. of tests performed
	number	Location	number	Location	
13	1	I-405N	2	I-405N	6
14	1	I-405N	5	I-405N	6
15	3	I-405N	3	I-405N	6
16	5	I-405N, I-405S, I-5N	5	I-405N, I-405S, I-5N	24
17	13	I-405N, I-405S, I-	17	I-405N, I-405S, I-	1

		5N, I-5S, 133N		5N, I-5S, 133N	
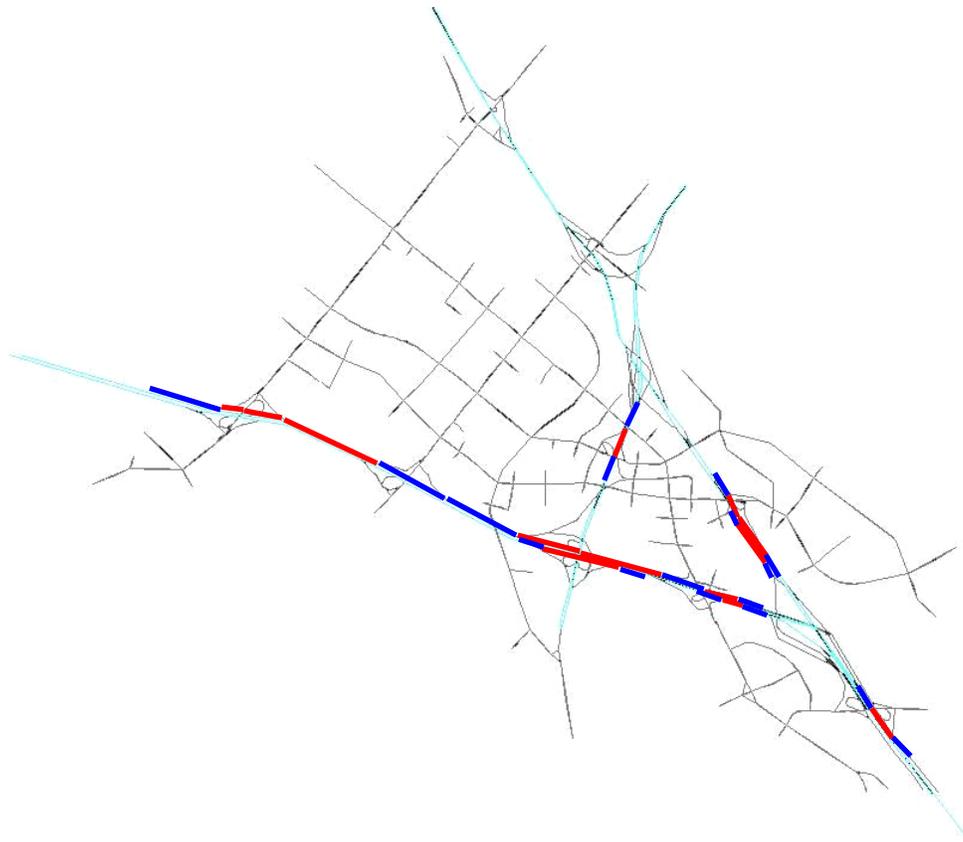
Figure 6-48 Distribution of Own-DM Links (in Red) and Borrowed-DM (in Blue) Links in Experiment 17

The plots in Figure 6-15 and Figure 6-16 show the error reductions of the speed and density estimates in Experiments13 ~ 17, with respect to the average on all the links, the own-DM links, the borrowed-DM links and the SM links, respectively. It is found that when 1) the number of own-DM links and borrowed-DM links are limited (in Experiment 13) or 2) the borrowed-DM links are not dominantly adjacent to where the dynamic models are originated (as in Experiment 14), the overall errors could be worse than those obtained with the static models. With more own-DM links

and nearby borrowed-DM links in the network, the entire traffic estimation capability is improved significantly (in Experiments 15 ~ 17). For instance, in Experiment 17, the traffic estimation exhibits slightly better error reduction when compared to Experiment 5, in which the dynamic models are only applied to the study links. These results suggest that the value of even a small number of sensors can be increased by judiciously transferring the dynamic model properties to adjacent links.
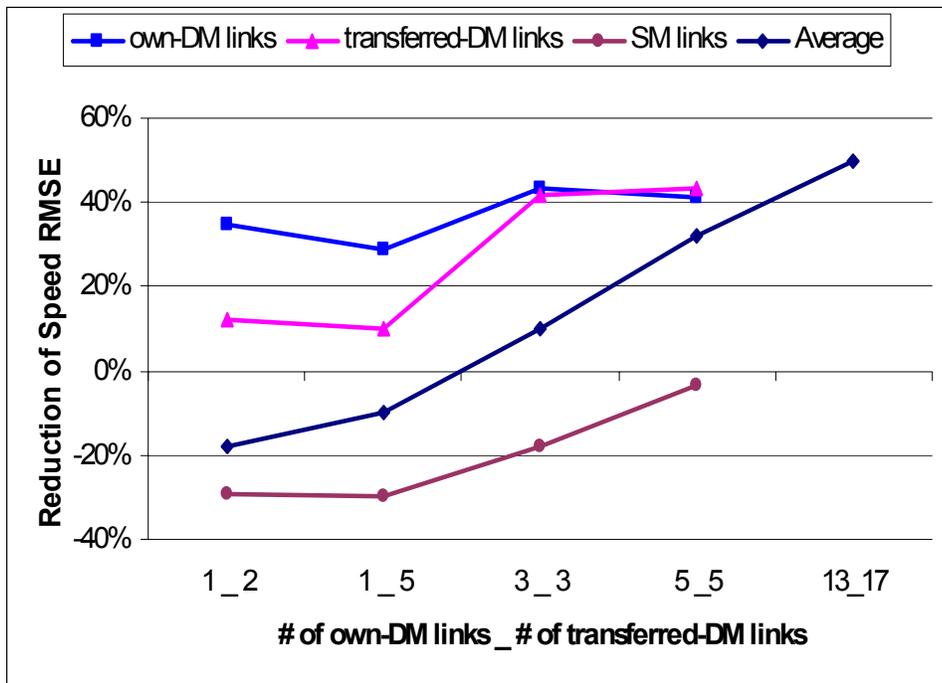
Figure 6-49 Error Reduction of Speed under Various DM Configurations in Experiments 13 ~ 17
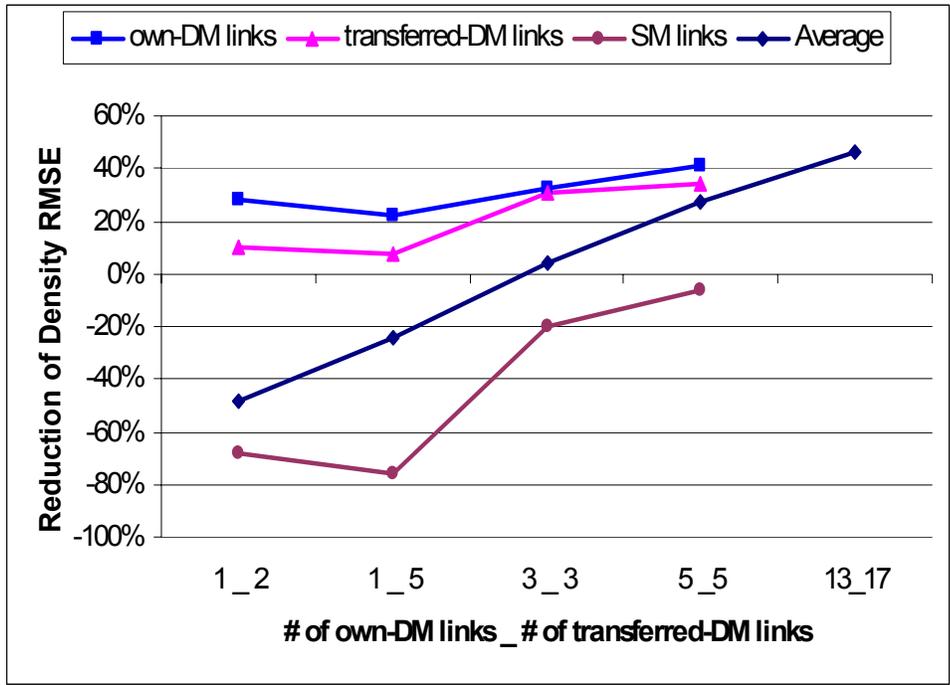
Figure 6-50 Error Reduction of Density under Various DM Configurations in Experiments 13 ~ 17

6.3.4   Traffic estimation capability with short term correction

The objective of the evaluation performed here is to evaluate the functionality of the short term correction proposed in Chapter 5. The main purpose of the short term correction is to adjust the simulated values to be more consistent with the observations. In DYNASMART-X, the variable to be corrected periodically is the speed predicted for the simulation interval whose start point is coincident with the receipt of the latest field observation. The discrepancies observed will be used for the correction of the next prediction of the speed. The experiment (Experiment 18) performed in this section is extended from Experiment 5, defined in Table 6-3. The DM links selected are the 13 study links, each of which has a dynamic model from its

138

own time series observations. The short term correction is applied for those links in Experiment 18; to provide a basis for comparison, Experiment 19 is conducted, in which the short term correction is applied to the static model (hence extended from Experiment 1).

Figure 6-17 shows the frequency histograms of the overall RMSEs of speed and density for the study links in Experiment 18. Comparing Figure 6-17 with Figure 6-9, the short term correction is found to provide an improvement in the quality of the predictions, as more links shift from the high-error regions to the low-error ones.

Figure 6-18 compares the hour-by-hour and overall errors for the study links using different modeling approaches, i.e. the static model without or with the short term correction (STC) in Experiments 1 and 19, and the dynamic model without or with the STC in Experiments 5 and 18. It is seen that the short term correction is useful to reduce the simulation errors for both the static model and the dynamic model. The static model produces larger improvement when the short term correction is employed, whereas the improvement for the dynamic model is not as pronounced as for the static model. Therefore, the value of the short term correction is in inverse proportion to the modeling quality of the original traffic flow model without correction algorithms.

Figure 6-19 and Figure 6-20 show an example of the time series of speeds and densities for one of the study links, 405N_5. In the plots, the actual observed data are compared with four different modeling approaches, in terms of which traffic flow model is used and whether the short term correction is activated. The speeds and densities generated by the static model have the lowest precision and tend to be less

responsive. The static model in conjunction with the short term correction becomes more responsive but still not as good as the dynamic model. The latter captures the underlying interrelation between speeds and densities and produces simulation results that are more in line with the observations. The short term correction can help improve the performance of the dynamic model, but not as much as in the case of the static model application.
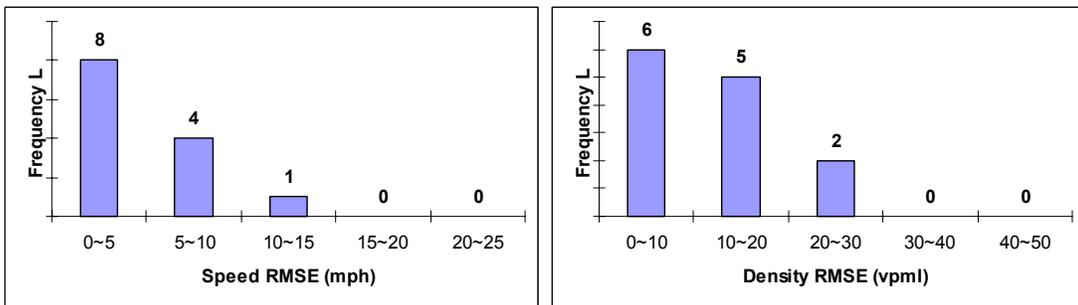
Figure 6-51 Dynamic Models with Short Term Correction: Frequency Histograms of Overall Errors (Speed and Density) for Study Links in Experiment 18
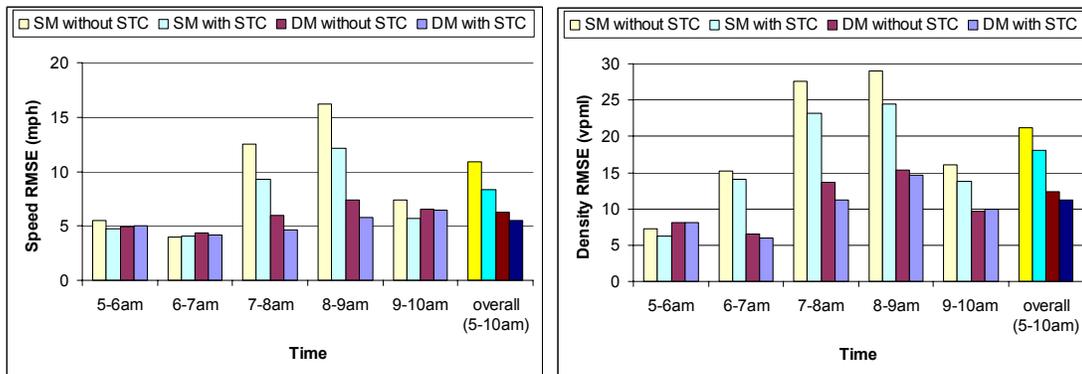
Figure 6-52 Comparison of Hour-by-Hour and Overall Errors (Speed and Density) Averaged for Study Links with Various Modeling Approaches
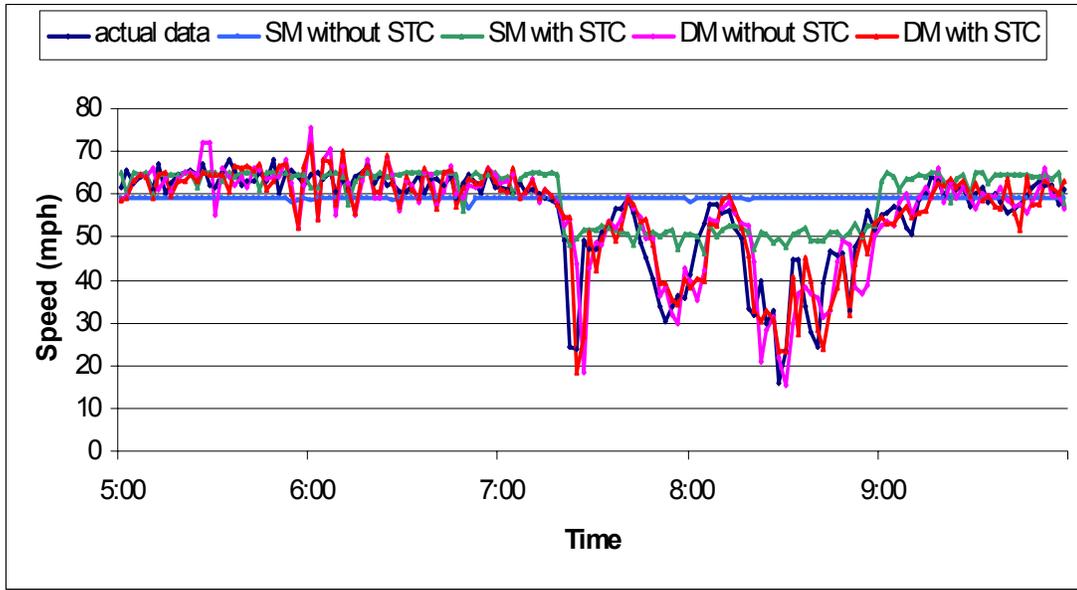
Figure 6-53 Comparison of Time Series of Speeds on Link 405N_5
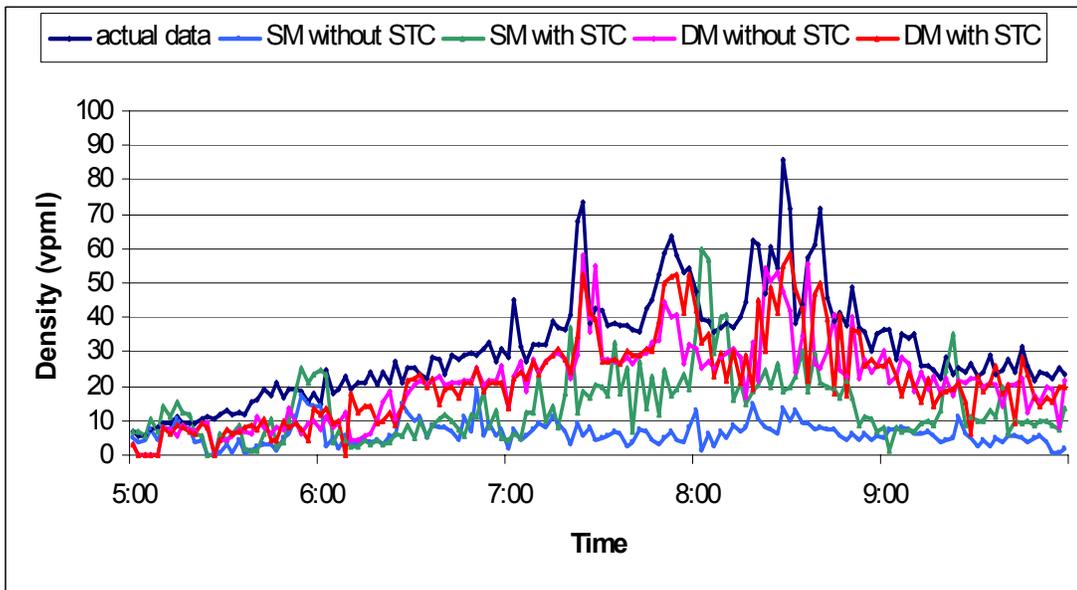
Figure 6-54 Comparison of Time Series of Densities on Link 405N_5

6.3.5   Traffic prediction capability

The objective of the evaluation performed in this section is to assess the dynamic model performance for traffic prediction over a pre-defined projection

141

horizon; naturally, it is not possible in this case to use sensor data beyond the projection horizon start time.

To examine the accuracy of prediction, link performance (density or speed) is considered through four groups that correspond to predictions obtained at four different times (corresponding to the consecutive stages m to $m+3$ in Figure 6-1). In other words, the first prediction for period A is obtained from prediction stage m; the second prediction is from stage $m+1$; the third prediction is from stage $m+2$; and the fourth prediction is from the nearest prediction stage $m+3$. The following experiment results are extracted from Experiments 1 and 5 (described in section 6.3.2) since each execution of the DYNASMART-X system includes both the traffic estimation and the traffic prediction results.

Figures 6-21 and 6-22 depict the prediction errors of speed and density generated at the different prediction stages for the static model and the dynamic model, respectively. The results conform to the expectation that the discrepancy in the first time prediction is in general larger than the second, third and fourth times. The fourth and latest prediction is closest to the actual observed value. Therefore, predictions in the near future are more reliable than predictions in the farther future, which is one of the motivating observations underlying use of the rolling horizon approach in stochastic dynamic systems. Overall, the dynamic model exhibits better modeling performance than the static model at each prediction stage. Figure 6-23 and Figure 6-24 provide an illustrative result of one of the study links, 5S_2. The time series of predicted speeds and densities at the different prediction times against the observed data are plotted. It is evident that the prediction of traffic 20 minutes ahead

is not as accurate as the one 5 minutes ahead. In sum, the dynamic traffic flow model improves the traffic state prediction capability progressively, and is more accurate than the static model.
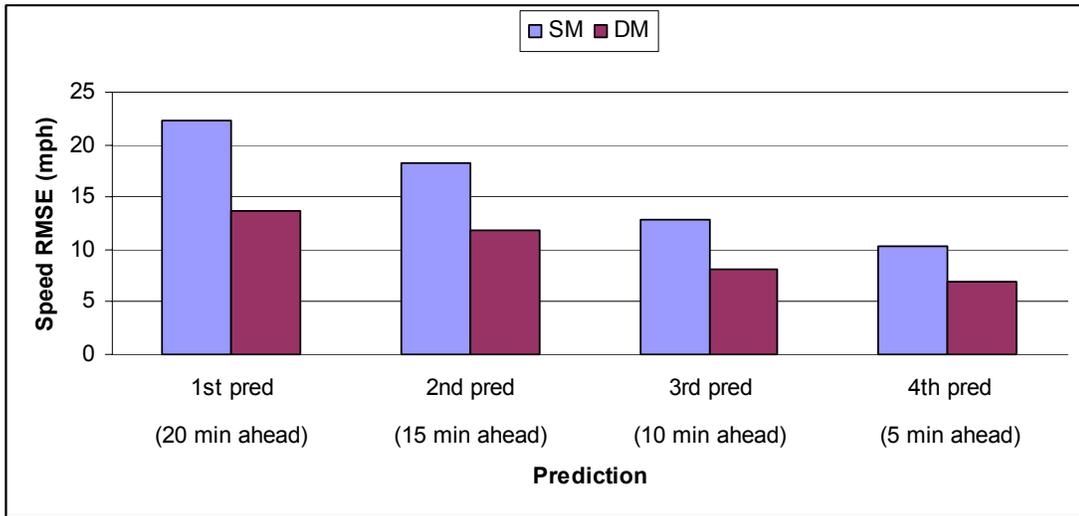
Figure 6-55 Average Speed Errors at Different Prediction Horizons for Static Model and Dynamic Model
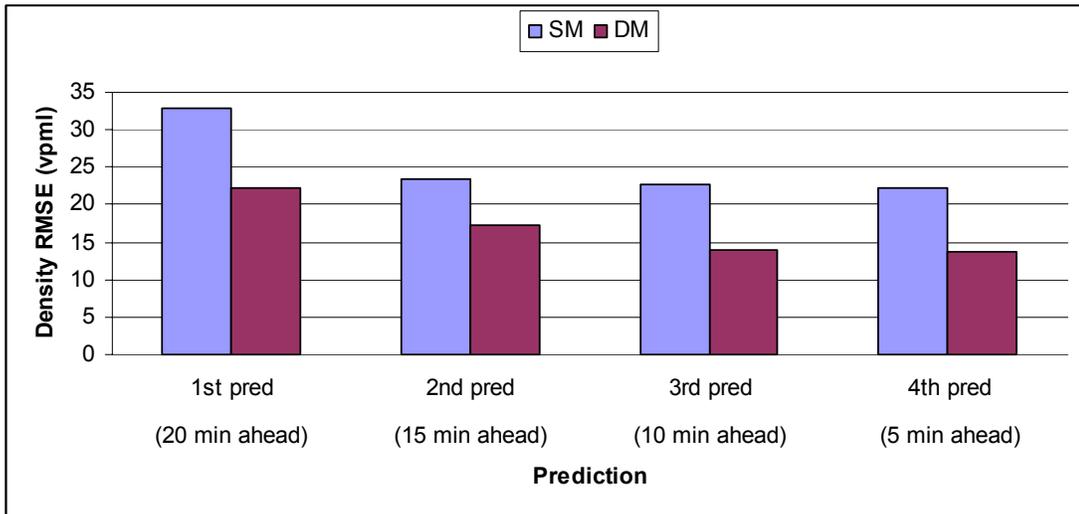
Figure 6-56 Average Density Errors at Different Prediction Horizons for Static Model and Dynamic Model
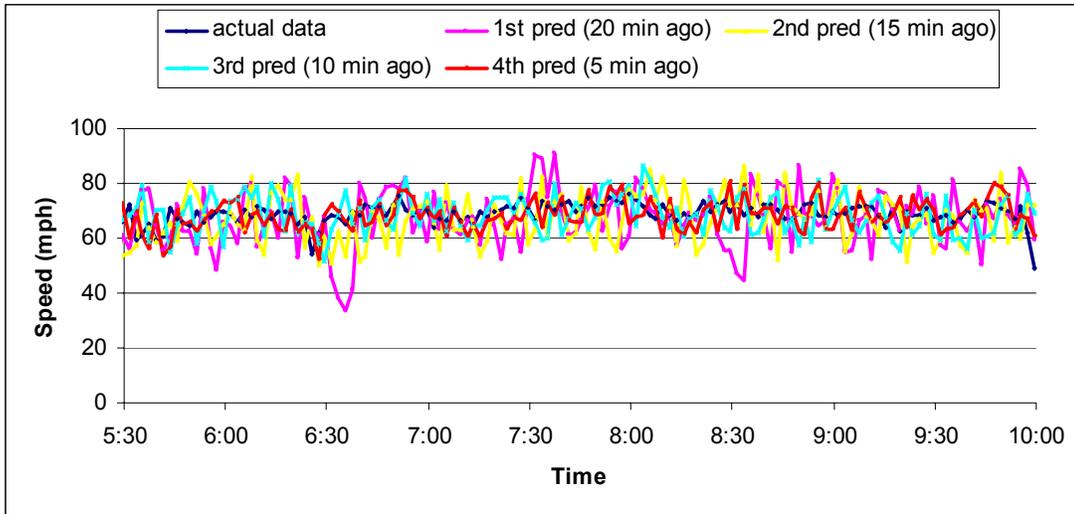
Figure 6-57 Comparison of Time Series of Speeds at Different Prediction Horizons on Link 133N_1
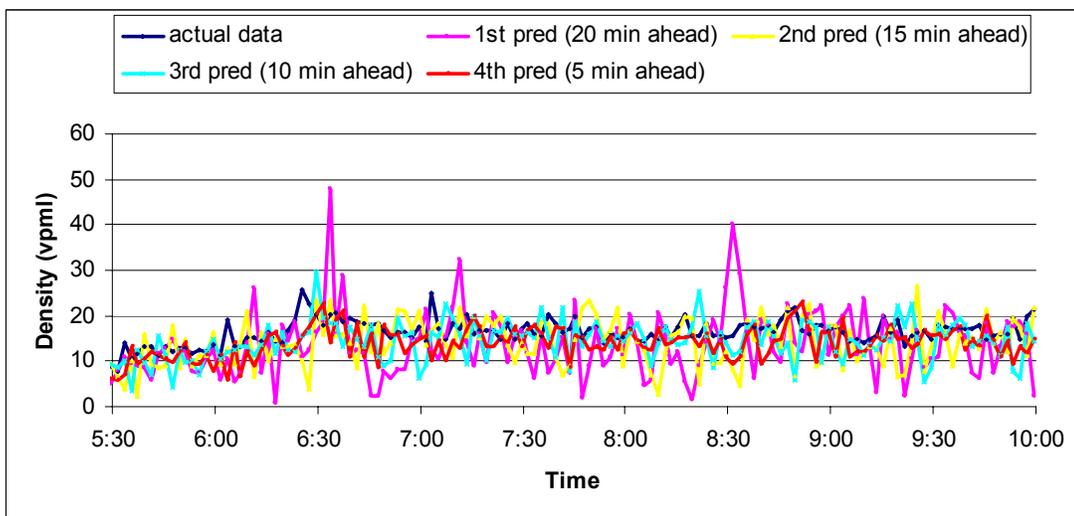
Figure 6-58 Comparison of Time Series of Densities at Different Prediction Horizons on Link 133N_1

6.3.6   Impact of temporal scale

The objective of the evaluation performed in this session is to analyze the impact of the temporal scale of the field observations on the performance of the dynamic model.

144

The experiments conducted here include all the 13 study links as the DM links. Experiment 5 (performed in session 6.3.2) provides the results for the case of the 30-second observation interval. By aggregating the original archived real-time data at 30-second interval, we perform additional three experiments (Experiment 20, 21, and 22) to have the results for the 1-minute, 3-minute, and 5-minute observation intervals respectively. The Type I dynamic model is used throughout the experiments.

Figure 6-25 and 6-26 are the plots of the error reduction of speeds and densities under different observation intervals. At the first glance, the experiment results show that, the 30-second observation interval leads to the most accurate estimation, 1-minute interval is the second best one; 3-minute or 5-minute interval cause a lower level of accuracy on speed and/or density than the static model. However, it should be noted that the comparison is made by using the original data interval (30-sec) in the post-processing computation, which means, even if the adaptive model calibration is based a long observation interval (i.e. greater than 30-sec), the resulted simulated values are aggregated into 30-sec intervals and compared with the original data series. The way is to retain a maximum credibility in the actual data. But it may also bring in a lot of fluctuations in the data series and swamp the underlying trend. The situation is especially worse when the calibrated model is based on a long observation interval. It is because the long interval is too aggregate to allow the calibrated model to predict the inner dynamic evolution of traffic flows from a perspective of a much shorter time interval. So, it is reasonable to simultaneously inspect the estimation quality over a longer time period. The

simulated values of the experiments together with the original data series are to be aggregated into 1-min through 5-min. Figure 6-27 and Figure 6-28 display the RMSEs of speeds and densities vs. data aggregation length for both static model and dynamic model with different observation intervals. The results tell us that, when aggregated for a longer interval, the simulated values from the dynamic model exhibit lower errors. The dynamic model with 3-min or 5-min observation interval turns out to be more desirable than the static model if the comparison is made at the aggregation interval greater than 3-min. The finding could sort of ease up the concerns towards the efficient usage of the real-time data at a longer observation interval.

Therefore, it is desirable to have a small time scale for observations to keep up a satisfying precision when the dynamic modeling with observations is employed in the traffic estimation and prediction system. However, it should be kept in mind that a long observation interval could be also valuable in the real-time modeling, not only due to the less demanding computation efforts, but also due to the ability to characterize the traffic pattern on a longer term.
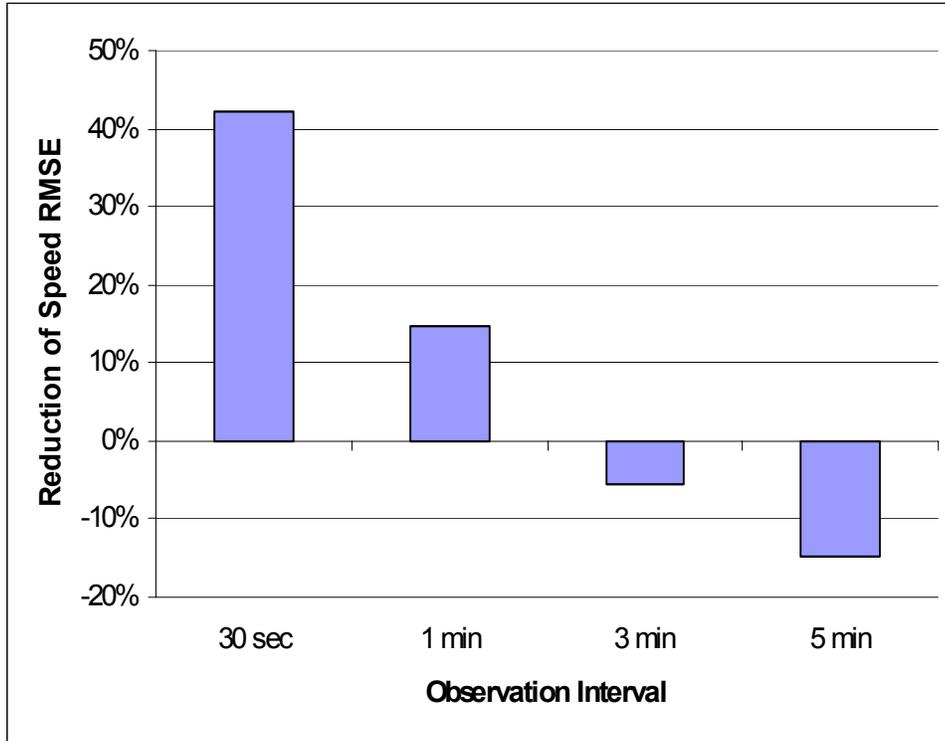
Figure 6-59 Reduction of Speed Errors with Length of Observation Interval
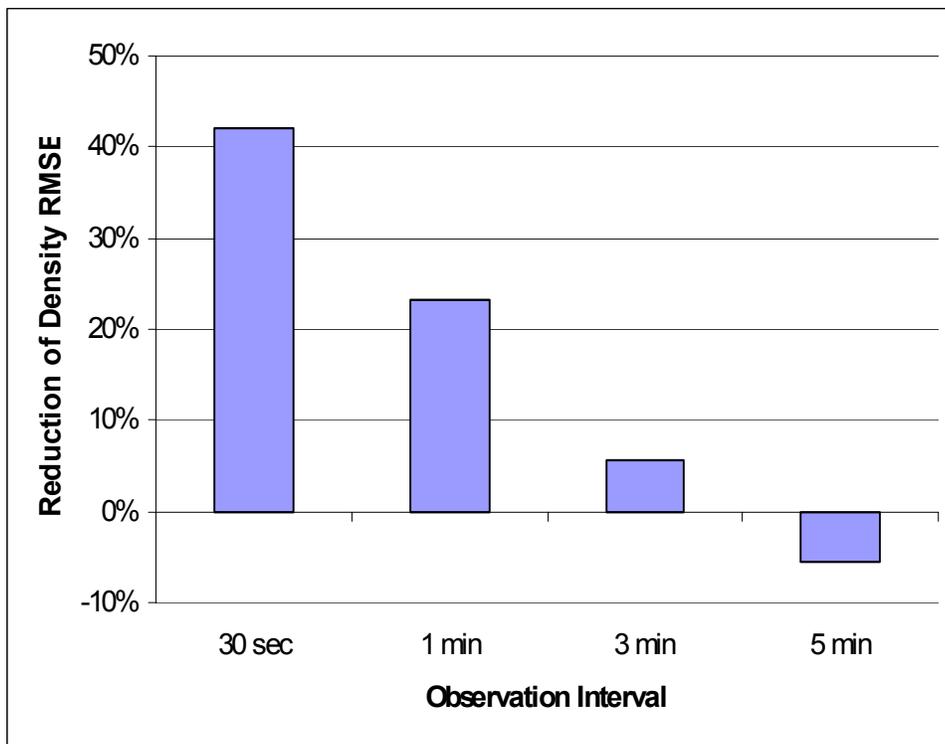
Figure 6-60 Reduction of Density Errors with Length of Observation Interval
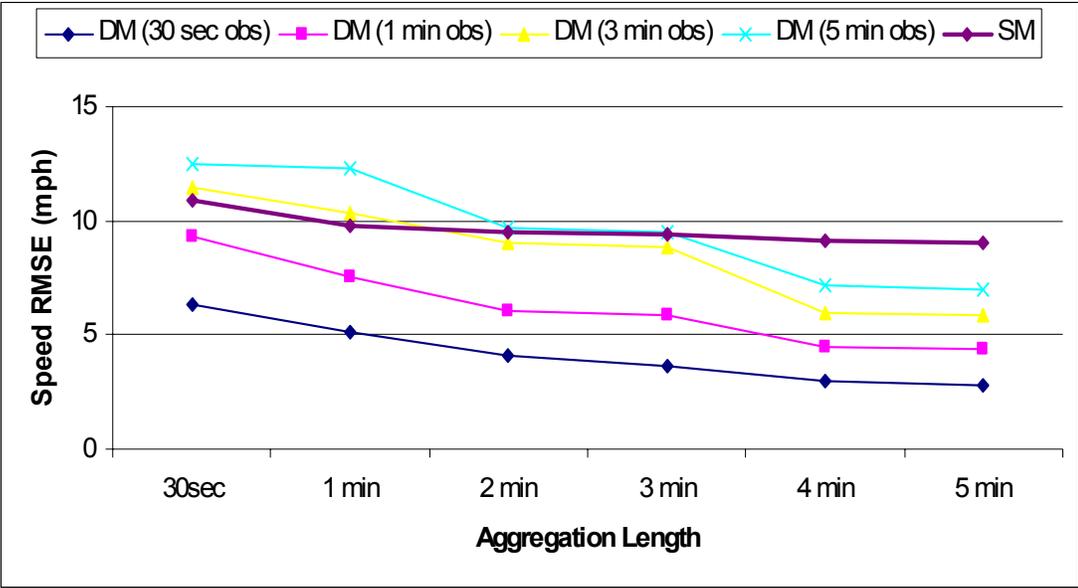
Figure 6-61 Speed RMSEs vs. Data Aggregation Length for Static Model and Dynamic Model with Different Observation Intervals
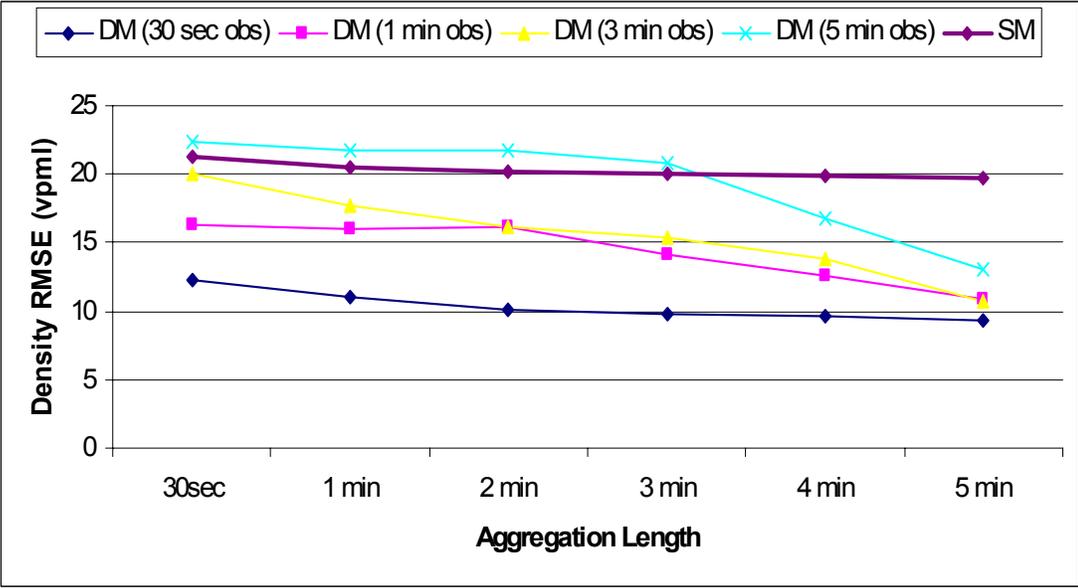
Figure 6-62 Density RMSEs vs. Data Aggregation Length for Static Model and Dynamic Model with Different Observation Intervals

6.3.7   Impact of spatial scale

The objective of the evaluation performed in this section is to analyze the impact of the spatial scale on the performance of the dynamic model.

Usually, a minimum link length is required for tracking the individual vehicles effectively in the traffic simulation. For instance, the minimum link length for the link i, $L_m(i)$, in DYNASMART is dependent on the average free-flow speed for the that link. If the free-flow speed of i is low, L_m is short; and vice versa. In the original Irvine network, the study links are in different sizes with respect to the minimum link length (see Table 6-5). A link is split into two or more short links if it is longer than the designated link length range in the experiments performed here. Traffic on a shorter link segment is conceptually more uniform. The new network with split links is modeled with either the static model or the dynamic model specification. For the dynamic model, the archived data are maintained for one of the split links depending on the sensor location and the Type I dynamic model (relaxation-only term) is used.   Table 6-5 lists the details of the conducted experiments.

Table 6-5 Details of Experiments 20 ~ 25

Original link length l	1~2 L_m	2~3 L_m	3~4 L_m	5~6 L_m	7~8 L_m
No. of study links	4	2	4	2	1

Experiment #	Model	Link length	Split scheme				
20	SM	3~4 L_m	Not use	Not use	Original length	Split: $\frac{2}{3}l+\frac{1}{3}l$	Split: $\frac{1}{2}l+\frac{1}{2}l$
21	DM						
22	SM	2~3 L_m	Not use	Original lengths	Split: $\frac{2}{3}l+\frac{1}{3}l$	Split: $\frac{1}{2}l+\frac{1}{2}l$	Split: $\frac{1}{3}l+\frac{1}{3}l+\frac{1}{3}l$
23	DM						
24	SM	1~2 L_m	Original lengths	Split: $\frac{1}{2}l+\frac{1}{2}l$	Split: $\frac{1}{2}l+\frac{1}{2}l$	Split: $\frac{1}{3}l+\frac{1}{3}l+\frac{1}{3}l$	Split: $\frac{1}{4}l+\frac{1}{4}l+\frac{1}{4}l+\frac{1}{4}l$
25	DM						

Figure 6-29 and 6-30 show the average hour-by-hour error reduction for speeds and densities for two models with various link lengths for the study links. It is seen in the results that the shorter link length leads to slightly better estimation quality when using the dynamic model, especially for density estimation. The performance of the static model is not quite sensitive to the link length ranges in the experiments conducted here, though it is worse than the dynamic model especially during the peak hours.
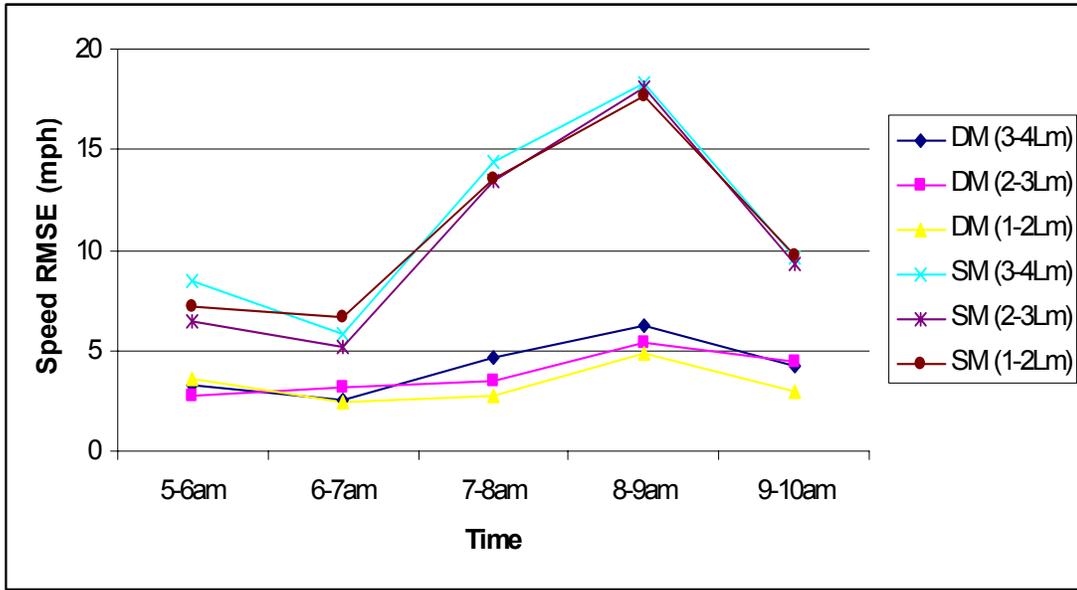
Figure 6-63 Average Hour-by-Hour Error Reduction of Speeds for Dynamic Model and Static Model with Various Link Lengths in Experiments 20 ~ 25
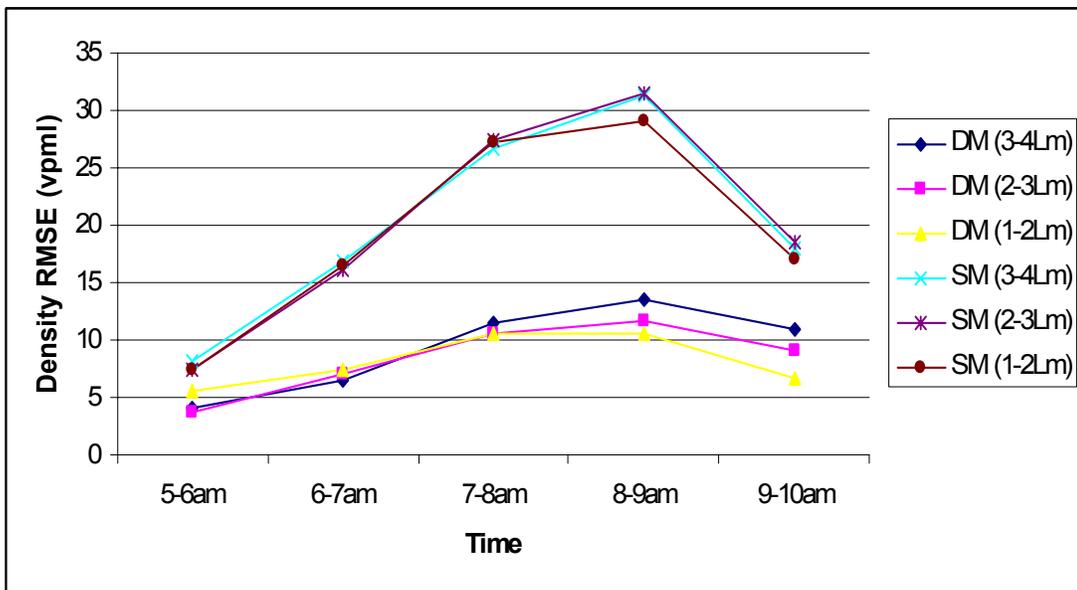
Figure 6-64 Average Hour-by-Hour Error Reduction of Densities for Dynamic Model and Static Model with Various Link Lengths in Experiments 20 ~ 25

6.4   <u>*Summary*</u>

The performance of the dynamic traffic flow model integrated with the real-time network traffic estimation and prediction system has been evaluated and illustrated in this chapter. The results, albeit limited to a single network with limited detector coverage (which is nonetheless still representative of the state of deployment in most regions in the US), are relevant to the practical usefulness of the proposed approach, and overall quite encouraging, although many challenges and difficulties remain for future exploration.

First, the dynamic model compatibility with the conventional static model in the practical case of partial sensor deployment observations is evaluated. The initial results reveal that the extent of the compatibility and effectiveness of the dynamic model is highly related to the number and spatial distribution of the sensors. The different specifications ("types") considered for the dynamic model (with different driving forces) are all effective in the traffic flow modeling, but the ultimate selection of the particular type is again dependent on the number of sensors and their distribution, especially in terms of allowing calculation of the anticipation and convection terms, which require downstream and upstream section detectors, respectively.

Model transferability is evaluated by transferring the adaptively calibrated model from one link to the other links. The results indicate that transferring to the neighboring links with similar facility type is an effective way to improve the modeling capability under limited availability of real-time information.

The third focus of the evaluation is the traffic estimation capability in conjunction with the short term correction function integral to the DYNASMART-X prototype. The short term correction has the ability to improve the estimation accuracy. The extent of the improvement is related to the performance of the original model without correction. The experimental results show that the impact of the short term correction is more apparent for the static model than the dynamic model. However, because of its incorporating the real time data into capturing underlying traffic flow pattern, the dynamic model without any correction still outperforms the static model.

Next, the performance of the dynamic model in terms of prediction capability is evaluated, confirming improved performance relative to the exclusive use of the static model. The results also confirm that, as expected, the near-term prediction is constantly better than prediction for farther intervals.

Finally, the impact of the temporal scale and the spatial scale in the dynamic model are analyzed by aggregating the original archived data and splitting the long links to shorter links. It is desirable to have a small time scale for observations to keep up a satisfying precision when the dynamic modeling with observations is employed in the traffic estimation and prediction system. However, a long observation interval could be also valuable in the real-time modeling, not only due to the less demanding computation efforts, but also due to the ability to result in a fairly good prediction defined for a longer interval. The impact of spatial scale is not as obvious as other factors according to the conducted experiments and needs to be further tested with more data coverage over the network.

The next chapter provides an overall summary and conclusion to this research.

The research contribution and future possible extensions are highlighted.

Chapter 7: Conclusions and Future Research

7.1   *Overall conclusions*

Dynamic traffic flow modeling is a challenging research topic and critical task for online network traffic estimation and prediction in the context of ITS. In this research, the fundamental objectives are to formulate and develop a dynamic traffic flow model driven by real-world observations, which is suitable for mesoscopic type dynamic traffic assignment simulation.

In this research, a dynamic speed-density relation is proposed by incorporating the physical and fundamental concept in continuum and kinetic models, coupled with the structural formulation of the transfer function model. The model explicitly includes phenomena such as speed relaxation, speed convection and density anticipation, which are found to affect the dynamics of the speed. The proposed model recognizes the time-lagged response of speed to the influential factors as well as the potential autocorrelated system noise. The transfer function method is typically used to describe the dynamic relationship between a responsive (output) variable and one or more influential (input) variables. It is useful in forecasting, and is a widely used linear time series model in engineering and other areas. By applying techniques adapted from the transfer function theory, the procedures for the model estimation and speed prediction using the real-time data are presented. The model estimation method is oriented to identifying a multiple-input transfer function which characterizes the dynamic traffic flow model proposed in this research. The speed is predicted by means of minimum mean square error. A rolling-horizon framework is

proposed for the adaptive calibration of the dynamic models upon receipt of the latest series of real-time observations. Such an adaptive mechanism provides a systematic way to maintain an updated traffic flow model that is adapted to the most recent traffic states.

Following the detailed description of the model formulation and the related algorithms, the dynamic model is demonstrated to be valid in its own right. It is evaluated in a series of standalone numerical experiments where link densities are assumed to be known values all the way along the entire study period and link speeds are to be predicted adaptively.

The robustness of the dynamic model with respect to the underlying equilibrium relation is established by examining the quality of resulting precisions using equilibrium relations of varying quality (in terms of goodness of fit to historical local data). Using dual-regime Modified Greenshields (MG) relation in the dynamic model does not significantly outperform the single-regime MG relation, although the former generally provides better representation of freeway traffic than the latter in the static representation. Moreover, the dynamic model is considered robust and reliable in the sense that the embedded equilibrium relation could be based on previous days or other traffic sections. This advantageous property allows the dynamic model to readily adapt to the equilibrium relations, which might be different from prevailing traffic patterns due to day-to-day variation or location-to-location variation, without substantial loss of accuracy. Among the three factors disclosed in the full model form, the relaxation and the anticipation are the most evident contributors to speed

dynamics. In other words, drivers' speeds change primarily with their current speeds relative to average speeds and traffic densities ahead.

The sensitivity analysis of the rolling horizon scheme with regard to the underlying speed density model is also performed in the standalone experiments. The trends of model performance with the different roll periods or calibration horizon are shown. Generally speaking, the prediction errors increase with the length of the roll period and decrease with the length of the calibration horizon. However, such improvement in prediction is usually accompanied by additional computational effort. Based on the numerical results, a rolling horizon scheme with a 2.5 minute roll period and 60 minute calibration horizon is recommended as providing a good compromise between accuracy and computational efficiency.

The principal issues concerning the application of the dynamic traffic flow model into the real-time DTA-based traffic estimation and prediction system are then discussed. In the context of real time DTA simulation operation, the dynamic model faces a problem related to the different time scales in the data observation and the traffic simulation, respectively. To accommodate the shorter simulation interval, at which link performance is updated in the DTA simulation, the impulse response calibrated using the observation interval is post-processed to approximate the impulse response at the simulation interval. The approximation approach is based on the connection between the discrete dynamic systems with different discrete scales due to the common underlying continuous dynamic system. The approach provides a straightforward, practical and efficient way to the real-world application of the dynamic traffic flow model based on the current surveillance characteristics.

To reduce the potential inconsistency due to unknown and uncontrolled factors when the model is integrated with network dynamic traffic assignment, the short term correction procedures are formulated to identify discrepancies between the simulation and the real-world observations, and adjust speed periodically. By applying the PI (proportional-integral) control mechanism adapted from feedback control theory, the speed adjustment is triggered by either speed-deviations or density-speed-deviations. The associated tuning factors can be either pre-specified or estimated adaptively using the least-squared-error method, constrained by the DTA models to reach internal and external consistency. The short term correction aims to minimize the discrepancy of the simulation from the reality and improve the entire estimation and prediction capability of the traffic flow model.

The performance of the dynamic model integrated in the DTA system is evaluated numerically. The Irvine network, with partial observations on a subset of links, provides the test bed for the experiments. With the limited real-time data, it is first investigated whether there is a benefit to apply the static model (for unobserved links) and the dynamic model (for observed links) in parallel. The experiments demonstrate that the overall estimation performance is dependent on the number and geographical distribution of the sensors that provide the real-time data. The dynamic model application with few and sparse sensor coverage could not improve the estimation on average (although it generally led to better results for the impacted links) and could even lead to worse results than the static model applied across the board. Therefore, when a network has sufficient amount of sensors installed, the dynamic model implementation would be able to enhance the entire traffic estimation

158

capability. It is also found that using the homogeneous model forms is preferable to the mixed forms. For instance, if most of the observed links are separated from each other, applying the Type I dynamic model (with the relaxation term as the influential factor) for all the observed links is more effective. Model transferability is assessed by transferring the adaptively calibrated dynamic model to other links for the purpose of speed predictions. The numerical results show that transferring among neighboring links is most effective. Hence, notwithstanding the limited number of sensors, it would still be useful to model the adjacent traffic characteristics in similar traffic environment.

Following evaluation of the model compatibility and transferability, the traffic estimation capability with short term correction and the traffic prediction capability of the dynamic model are studied. The effective short term correction is a supporting and effective approach to improve the performance of modeling traffic flow characteristics because it takes advantage of the real-time data to make comparison and correction. The numerical results reveal that the approach is effective for both the static model and the dynamic model, though the impact on the static model is larger. From the test results, the performance of the static model with the short term correction did not even beat the dynamic model without the short term correction, which gives a strong confirmation of the superiority of the dynamic model. For traffic state prediction, the dynamic model again outperforms the static model in progressively improving the traffic state prediction quality.

Finally, the impact of the temporal scale of the real-time data and the spatial scale of the observed links are examined. The experiments use 30 second, 1 minute,

3 minutes and 5 minutes as the observation intervals and the results show that the dynamic model based on an observation interval of 30 second or 1 minute is good at short term predictions. However, data at 3 or 5-min intervals could be useful as well in providing predictions which have sufficient accuracy for the long term.

7.2   Research contributions

One of the most crucial requirements for advanced real-time traffic simulation tools is to be able to provide a quasi-continuous view of the state of the traffic system over time and space. Such tools as simulation-based Dynamic Traffic Assignment (DTA)-type models are intended to perform real-time system-wide traffic estimation and prediction, based on the existing surveillance system, and meet the information requirements for decision making for operators and users of the traffic network, in the complex ITS environment. Reliable and robust traffic flow models which are capable of representing the dynamic evolution of traffic over space and time ensure that the information provided by DTA systems, generally including descriptive traffic conditions and normative route guidance, is credible and close to reality. This dissertation introduces a new perspective to the specification, calibration and application of ITS-oriented traffic flow models.

The most popular and widely-used models applied to mesoscopic type traffic simulation are in the category of the macroscopic simple order continuum traffic flow theory. The simple order model captures the equilibrium relations among the essential traffic characteristics and provides a static representation which is more suitable for long term operational planning. In the context of real-time DTA

operations, a more responsive and dynamic type of models is desirable to adapt the traffic state estimation/prediction to the real world.

In this dissertation, the underlying physical phenomenon embedded in the proposed dynamic traffic flow model is well grounded due to its strong connection with the higher-order continuum traffic flow theories, which are intended to overcome the "shortcomings" of the simple order ones. However, instead of struggling to explore the exact formulation and estimation of the higher-order models, which is still an inconclusive topic in traffic flow research, the mathematical formulation proposed in the dissertation is adapted from the classical time series theory. Introducing the time-series theory into the traffic flow modeling is totally driven by the availability of the real-time traffic data. To analyze and capture the interrelation of two or more variables of interest, the transfer function method, which is a widely-used modeling approach for dynamic systems, is an appropriate technique to explore.

Applying transfer function methods results in a functional structure of the model formulation to describe the time-dependent interrelation between speed and its influential factors, such as speed relaxation, speed convection and density anticipation. Speed relaxation describes that drivers tend to accelerate or decelerate toward the desired equilibrium speed if the actual speed is lower or higher. Speed convection illustrates that speeds could increase or decrease due to the faster and slower vehicles behind. Density anticipation expresses that drivers determine to accelerate or decelerate according to the lower or higher traffic density ahead. Each influential factor can be quantitated if the collected real-time data are densities and

speeds at each observation interval. The transfer function modeling framework declaims that a dynamic system is identified for the speed dynamics over time under the time-lagged influences of these factors as well as the potential autocorrelated noise pattern.

The model estimation is extended from the approach suitable for the multi-input transfer function. The minimum mean square error forecast is given by the expectation conditional on the knowledge of the series from the past up to the present, which has very practical meaning. The theoretical basis and the algorithmic procedures for estimation and prediction are well described in the dissertation.

The application of the dynamic traffic flow model is designed to be taken as part of an online operational capability for dynamic traffic assignment (DTA) simulation modeling to predict network traffic conditions in real-time, in order to support traffic operations management and information distribution. Therefore, to accommodate the requirement in the real-time DTA operation, three important application details are pointed out and discussed in the dissertation.

First, the model is proposed to be implemented through rolling horizon framework which is very flexible and effective in terms of the calibration of the parameters. The parameters are derived with the most recent traffic data and changing along changes of traffic. Unlike the conventional static traffic flow model, the dynamic model itself is more adaptive to real-world traffic situation.

Second, to deal with different observation intervals and simulation intervals when applying the dynamic model in the simulation system, a simple but effective approach is proposed. The solution is a result of recognizing that a common

underlying continuous dynamic system can be approximated by different discrete dynamic systems with different time scales. Although approximation errors could exist in the procedure suggested, the entire modeling framework and approaches are practical and effective as indicated by the experimental results.

Third, the concept and framework of short term correction is introduced. Its operation conjunct with the traffic flow model plays an essential role for real-time DTA operation. Short term correction intends to capture the critical deviations between the DTA-simulator and the actual transportation system and prescribes the treatment to compensate the discrepancies. In other words, more reliable and consistent traffic state estimation is highly possible with the support of the short term correction. The effect is attributed to the thorough utilization of the real-time data, so the real-time data are used not only for model estimation and speed prediction but also for consistency checking.

In sum, the dissertation proposes a dynamic traffic flow model with real-time traffic sensor data for the purpose of online traffic estimation and prediction to support ATMS/ATIS in an urban transportation network. The model is meaningful due to the powerful theoretical background, practical due to the well-designed approaches, and worthy of further application and exploration due to the encouraging and interesting numerical results.

7.3   *Future research*

As an initial effort in the challenging area of real-time traffic flow modeling, several aspects of the theory and methods proposed in the dissertation remain to be improved and explored. Several examples are given as follows.

The first example would be to extend the existing model to include seasonal patterns. The seasonal pattern here refers to the periodic behavior of traffic data series with period s. Similarities in the series occur after s basic time intervals. At first, we can focus on day-by-day similarities; then we can further move onto week-by-week similarities. So, a seasonal time series can be decomposed into trend, seasonal, and noise components. The trend and noise components have been represented in the dissertation and the seasonal component is left for further investigation. The intrinsic implication from models with seasonal patterns is to include the historical information (e.g., one day before or one week before) in the current forecasting. The resulting forecasts would take into account both ordinary behaviors which are summarized from the previous similar periods, and possible behaviors which are exclusive for that day.

The second example of further research is to explore other forms of the higher order continuum models. The research in the dissertation is primarily based on the original higher order continuum model form proposed by Payne [1971] and Whitham [1974]. As stated in Chapter 2, some efforts have been made to extend and improve the original model. Some new or adjusted terms, such as an adjusted anticipation term, a viscosity term, a friction term, and so on, are introduced in these efforts. By using a similar modeling framework as in the dissertation, the performances of these

new model specifications can be evaluated and compared with each other in the context of real-time traffic flow modeling.

The third example is inspired by the test results in section 6.3.4, where the traffic estimation capability with short term correction is assessed. It is found that the performance of the dynamic model outperforms the static model with short term correction function. Short term correction does not have significant influence on the dynamic model because the adaptive traffic information has been absorbed into the dynamic model, which is updated quasi-continuously. Therefore, other than its use to model the traffic flow directly in the DTA system, the dynamic model proposed in this research could also be used for the purpose of short term correction to support the static traffic flow model. This approach could provide a more reliable mechanism that originates from physical background for short term correction.

The fourth example of further investigation is to evaluate the proposed dynamic model with a larger amount and more diverse types of data. The conclusions reached in the previous chapters are based on tests with data from a limited number of locations. If more link sections in the network are covered by sensors, the evaluation of the dynamic model would be more comprehensive and conclusive. In addition to using data from loop detectors, other real-time data sources could be considered to improve the entire traffic flow modeling performance. For example, travel time extracted from AVI (automated vehicle identification) equipments, or traffic densities obtained from processed video imaging data, could be used to conduct consistency checking on estimation of link performance.

Bibliography

1. Ahmed, M.S. and A.R. Cook. Analysis of Motorway Traffic Time Series Data by Using Box-Jenkins Techniques. Transportation Research Record 722, 1-9, 1979.

2. Annunziato, M., I. Bertini, A. Pannicelli, and S. Pizzuti. Evolutionary feed-forward neural networks for traffic prediction. EUROGEN2003, Barcelona, Spain, Sept. 2003.

3. Ben-Akiva, M., Koutsopoulos, H.N. Mishalani, R. "DynaMIT: A Simulation-Based System for Traffic Prediction". Paper presented at the DACCORD Short Term Forecasting Workshop, Delft, The Netherlands. 1998.

4. Box, G.E.P., and G.M. Jenkins. Time Series Analysis: Forecasting and Control. Holden-Day, Inc., San Francisco, CA, 1976.

5. Brilon, W., and N. Wu. Evaluation of cellular automata for traffic flow simulation on freeway and urban streets. Brilon et al., 63-180, 1999.

6. Bureau of Transportation Statistics, *National Transportation Statistics 2004*, Washington, D.C.: U.S. Department of Transportation, 2004.

7. Chang, G.L., Mahmassani, H.S. and Herman, R. A Macroparticle Traffic Simulation Model to Investigate Peak-Period Commuter Decision Dynamics, Transportation Research Record 1005, 107-120, 1985.

8. Cremer, M., and M. Papageorgiou. Parameter identification for a traffic flow model, Automatica 17, 837-843, 1981.

9. Daganzo, C. F. Requiem for Second-Order Fluid Approximations of Traffic Flow. Transportation Research B, Vol. 29B, No. 4, 277-286, 1995.

10. Derzko, N.A., A.J. Ugge, E.R. Case. Evaluation of a dynamic freeway model using field data. Transportation Research Record, 905, 52-60, 1983.

11. Doyle, J.C., B.A. Francis, and A.R. Tannenbaum. Feedback Control Theory. Macmillan Coll Div, 1992

12. Garrison, William L., Jerry D Ward, Tomorrow's Transportation: Changing Cities, Economies, and Lives, Boston Artech House, Inc., 2000.

13. Gazis, D. C., R. Herman, and R. W. Rothery. Nonlinear follow the leader models of traffic flow, Operations Research, 9, 545-567, 1961.

14. Gazis, D.C. and C.H. Knapp. On-Line Estimation of Traffic Densities from Time-Series of Flow and Speed Data. Transportation Science Vol. 5, No. 3, 283-301, 1971.

15. Greenberg, H. An Analysis of Traffic Flow. Operations Research, Vol 7, 78-85, 1959.

16. Greenshields, B.D. A Study in Highway Capacity. Highway Research Board, Proceedings, Vol. 14, 458, 1935.

17. Hauer, E., V.F. Hurdle. Discussion on FREFLO. Transportation Research Record, 722, 75-76, 1979.

18. Haugh, L.D. nd G.E.P. Box. Identification of Dynamic Regression (Distributed lag) Models Connecting Two Time Series. Journal of the American Statistical Association, 72, 121-130, 1977.

19. Helbing, D., *Verkehrsdynamik*, Springer, Berlin, 1997.

20. Helbing, D. Traffic and related self-driven many-particle systems. Reviews of Modern Physics 73, 1067-1141, 2001.

21. Herman, R., E. W. Montroll, R. B. Potts and R. W. Rothery. Traffic Dynamics: Analysis of Stability in Car Following. Operations Research, E. 17, pp. 86-106, 1959.

22. Herman, R. and R. B. Potts. Single Lane Traffic Theory and Experiment. Proceedings Symposium on Theory of Traffic Flow. Ed. R. Herman, Elsevier Publications Co., pp. 120-146, 1959.

23. Herman, R. Technology, Human Interaction, and Complexity: Reflections on Vehicular Traffic Science. Operations Research archive, Volume 40, Issue 2, March–April, pp. 199-212, INFORMS, Linthicum, Maryland, USA 1992.

24. Hoogendoorn, S.P. A Macroscopic Model for Multiple User-Class Traffic Flow. Proceedings of the 3rd TRAIL PhD. Congress, TRAIL Research School, Part I, 1997.

25. Hoogendoorn S. P. Multiclass Continuum Modelling of Multilane Traffic Flow. TRAIL Thesis Series, Delft University Press, The Netherlands, 1999.

26. Hoogendoorn, S.P. and P.H.L. Bovy. Multiclass Macroscopic Traffic Flow Modeling: A Multilane Generalisation Using Gas-Kinetic Theory. Proceedings of the 14th International Symposium of Transportation and Traffic Theory, 27-50, 1999.

27. http://www.dynamictrafficassignment.org, website of Office of Operation Research Development and Technology, Washington, D.C.

28. http://www.its.dot.gov, website of Joint Program Office of the Intelligent Transportation System program, U.S. Dept of Transportation, Washington, D.C.

29. http://www.itsa.org, website of ITS America Organization, Washington, D.C.

30. Huynh, N., H. S. Mahmassani, and H. Tavana. Adaptive Speed Estimation Using Transfer Function Models for Real-Time Dynamic Traffic Assignment Operation. The 81th Transportation Research Board Meeting Preprint CD-ROM, Washington, D.C., 2002.

31. Kalman R. E. A New Approach to Linear Filtering and Prediction Problems. Journal of Basic Engineering 82D (I), 35-45, 1960.

32. Kuhn, T. Black-Body Theory and the Quantum Discontinuity, 1894-1912. The University of Chicago Press, 1978.

33. Kühne, R.D. Macroscopic freeway model for dense traffic – stop-start waves and incident detection. Proceedings of the Ninth International Symposium of Transportation and Traffic Theory, 21-42, ed. by I. Volmuller & R. Hamerslag, 1984.

34. Kühne, R.D., R. Beckscchulte. Non-linearity Stochastic of Unstable Traffic Flow. Proceedings of the Twelfth International Symposium of Transportation and Traffic Theory, 21-42, ed. by C.F. Daganzo, 1993.

35. Kühne, R.D, and P. Michalopoulos. Traffic Flow Monograph: Chapter 5. FHWA, U.S. Department of Transportation, 1998, pp 21-25.

36. Leboeuf, J.N., Tajima, T. and Dawson, J.M. "A Magnetohydrodynamic Particle Code for Fluid Simulation of Plasmas", Journal of Comparative Physics 3 1, 3, 379-408, 1979.

37. Ledoux, C. An Urban Traffic Flow Model Integrating Neural Networks. Transportation Research C: Vol. 5, No. 5, 287-300, 1997.

38. Leonard, D.R., P. Gower and N.B. Taylor. CONTRAM: Structure of the Model. Transport and Road Research Laboratory (TRRL) Research Report 178. Department of Transport, Crowthorne. 1989. See also www.contram.com.

39. Levin, M., and Y. Tsao. On forecasting Freeway Occupancies and Volumes. Transportation Research Record 773, 47-49, 1980.

40. Lieberman, E.B. et al. Variable cycle signal timing program: Volume 4 -- prediction algorithms, software hardware requirements, and logical flow diagrams. FHWA, USDOT, NTIS: PB 241720, May 1974.

41. Lighthill, M. H. and G. B. Whitham. On Kinematic Waves-II. A Theory of Traffic Flow on Long Crowded Roads. Proceedings, Royal Society (London), A229, No. 1178, pp. 317-345, 1955.

42. Liu, L., and D.M. Hanssens. Identification of Multiple-input Transfer Function models. Communication in Statistics – Theory and Methodology, 11(3), 297-314, 1982.

43. Liu, G., A.S. Lyrintzis and P.B. Michalopoulos. Improved High-Order Model for Freeway Traffic Flow. Transportation Research Record, 1644: 37-46, 1998.

44. Ljung, G.M. and G.E.P. Box. The Likelihood Function of Stationary Autoregressive-Moving Average Models. Biometrika, 66, 265-270, 1979.

45. Lu J. Prediction of Traffic Flow by an Adaptive Prediction System. Transportation Research Record 1287, 54-61, 1990.

46. Mahmassani, H.S., A.F. Abdelghany N. Huynh, X. Zhou, Y-C. Chiu, and K.F. Abdelghany. DYNASMART-P (version 0.926) User's Guide. Technical Report STO67-85-PIII, Center for Transportation Research, University of Texas at Austin. 2001.

47. Mahmassani, H.S., T.Y. Hu and S. Peeta. Microsimulation-based procedures for dynamic network traffic assignment", Proceedings of the 22nd European Transport Forum, H2, 53-64, PTRC, London, 1994.

48. Mahmassani, H.S. Dynamic Network Traffic Assignment and Simulation Methodology for Advanced Systems Management Applications. Networks and Spatial Economics, Vol. 1, pp. 267-292, 2001.

49. Massachusetts Institute of Technology ITS Program. DynaMIT Evaluation Final Reports Draft: Development of a Deployable Real-Time Dynamic Traffic Assignment System. Volpe National Transportation Systems Center, 2003.

50. May, A.D. Traffic Flow Fundamentals. Prentice-Hall, Inc, NJ, 1990.

51. Michalopoulos P. G., and D. E. Beskos. Improved Continuum Models of Freeway Flow. Proceedings of the Ninth International Symposium on Transportation and Traffic Theory, 89-111, 1984.

52. Michalopoulos P. G., P. Yi, and A. S. Lyrintzis. Development of an Improved High-order Continuum Traffic Flow Model. Transportation Research Record, 1365, 125-132, 1992.

53. Munjal, P. and J. Pahl. An Analysis of the Boltzmann-type Statistical Models for Multi-lane Traffic Flow. Transportation Research, Vol. 3, 151-163, 1969.

169

54. Nagel, K., and M. Schreckenberg. A cellular automaton model for freeway traffic. Journal of Physics, I France 2, 2221-2229, 1992.

55. Nagel, K., and M. Paczuski. Emergent traffic jams. Physics Review, E 51, 2909-2918, 1995.

56. Nelson, P. A Kinetic Model of Vehicular Traffic and its Associated Bimodal Equilibrium Solutions. Transport Theory and statistical Physics, Vol. 24, Nos. 1-3, 383-409, 1995.

57. Newell, G. F. Nonlinear effects in the dynamics of car following. Operations Research, 9, 209-229, 1961.

58. Okutani, I., and Y.J. Stephanedes. Dynamic Prediction of Traffic Volume through Kalman Filtering Theory. Transportation Research., 18B(1), 1–11, 1984.

59. Oswald, R.K., W.T. Scherer, B.L. Smith. Traffic Flow Forecasting Using Approximate Nearest Neighbor Nonparametric Regression. Final project of ITS Center project: Traffic forecasting: non-parametric regressions, December 2000.

60. Pankratz, A. Forecasting with Dynamic Regression Model. Wiley-Interscience Publication, 1991.

61. Papageorgiou, M., Posch B. and Schmidt G. Comparison of macroscopic models for control of freeway traffic. Transportation Research, 17B, 107-116, 1983.

62. Papageorgiou, M., Blosseville, J. M. and Haj-Salem, H. Macroscopic modelling of traffic flow on the Boulevard Périphérique in Paris. Transportation Research B, 23, 29-47, 1989.

63. Papageorgiou M. Some Remarks on Macroscopic Traffic Flow Modeling. Transportation Research A, Vol. 32, No. 5, 323-329, 1998.

64. Paveri-Fontana, S. L. On Boltzmann-like Treatments for Traffic Flow: A Critical Review of the Basic Model and an Alternative Proposal for Dilute Traffic Analyis Transportation Research, Vol 9, 225-235, 1975.

65. Payne, H. J. Models of Freeway Traffic and Control. Mathematical Models of Public Systems, edited by G. A. Bekey (Simulation Council, La Jolla, CA), Vol. 1, 51-61, 1971.

66. Payne, H. J. FREFLO: A Macroscopic Simulation Model of Freeway Traffic. Transportation Research Record 722, 68-77, 1979.

67. Pipes, L. A. An operational analysis of traffic dynamics, Journal of Applied Physics, 24, 274-281, 1953.

68. Priestley, M.B. Fitting Relationships between Time Series. Bulletin of the International Statistical Institute, 34, 295-324, 1971.

69. Prigogine, I. A Boltzmann-like Approach to the Statistical Theory of Traffic Flow. Theory of Traffic Flow, ed. By R. Herman, 1961.

70. Prigogine, I. and R. Herman. Kinetic Theory of Vehicular Traffic. Elsevier, 1971.

71. Qin, X. and H.S. Mahmassani. Adaptive Calibration of Dynamic Speed-Density Relations for Online Network Traffic Estimation and Prediction Applications. Transportation Research Record 1876, 82-89, 2004.

72. Richards, P. I. Shock Waves on the Highway. Operation Research, 4 (1), pp. 42-51, 1956.

73. Smith, B.L., B.M. Williams, and R.K. Oswald. Comparison of Parametric and Nonparametric Models for Traffic Flow Forecasting. Transportation Research Part C: Vol.10, 303–321, 2002.

74. Stephanedes, Y.J., P.G. Michalopoulos and R.A. Plum. Improved Estimation of Traffic Flow for Real-Time Control. Transportation Research Record, 795, 28-39, 1981.

75. Tavana, H., and H. S. Mahmassani. Estimation and Application of Dynamic Speed-Density Relations Using Transfer Function Models. The 79th Transportation Research Board Meeting Preprint CD-ROM, Washington, D.C., 2000.

76. The National ITS Architecture 5.0, 2003

77. Treiber, M., A. Hennecke, and D. Helbing. Congested traffic states in empirical observations and microscopic simulations. Physics Review, E62, 1805-1824, 2000.

78. University of Maryland DTA Group. DYNASMART-P Evaluation for Operational Planning Applications. Maryland Transportation Initiative, 2003.

79. University of Maryland DTA program. DYNASMART-X Evaluation for Real-Time TMC Application. Maryland Transportation Initiative, 2004.

80. University of Maryland DTA program. DYNASMART-X User's Guide. Maryland Transportation Initiative, 2004.

81. Urban Traffic Control System and Bus Priority System Traffic Adaptive Network Signal Timing Program: Software Description. FHWA, USDOT, Aug, 1973.

82. Vythoulkas, P.C. Alternative Approaches to Short Term Forecasting for Use in Driver Information Systems. Proceedings of the Twelfth International Symposium of Transportation and Traffic Theory, 485-506, ed. by C.F. Daganzo, 1993.

83. Whitham, G. B. Linear and Nonlinear Waves. Wiley & Sons, NY, 68-95, 1974.

84. Williams B.M., P. Durvasula and D. Brown. Urban Freeway Traffic Flow Prediction: Application of Seasonal ARIMA and Exponential Smoothing Models. Transportation Research Record 1644, 132-141, 1998.

85. Williams, B.M. Multivariate Vehicular Traffic Flow Prediction: Evaluation of ARIMAX modeling. Transportation Research Record 1776, 194-200, 2001.

86. Yin, H., S.C. Wong, J. Xu, and C.K. Wong. Urban Traffic Flow Prediction Using a Fuzzy-neural approach. Transportation Research C: Vol. 10, 85-98, 2002.

87. Zhang, H. M. A Theory of Nonequilibrium Traffic Flow. Transportation Research -B, Vol. 32, No. 7, 485-498, 1998.

88. Zhang, H. M. Analyses of the Stability and Wave Properties of a New Continuum Traffic Theory. Transportation Research -B Vol.33, 399-415, 1999.