ABSTRACT

Title of Dissertation:	DYNAMIC ORIGIN-DESTINATION DEMAND ESTIMATION AND PREDICTION FOR OFF-LINE AND ON-LINE DYNAMIC TRAFFIC ASSIGNMENT OPERATION
	Xuesong Zhou, Ph.D., 2004
Dissertation Directed By:	Professor Hani S. Mahmassani, Department of Civil and Environmental Engineering

Time-dependent Origin-Destination (OD) demand information is a fundamental input for Dynamic Traffic Assignment (DTA) models to describe and predict time-varying traffic network flow patterns, as well as to generate anticipatory and coordinated control and information supply strategies for intelligent traffic network management. This dissertation addresses a series of critical and challenging issues in estimating and predicting dynamic OD demand for off-line and on-line DTA operation in a large-scale traffic network with various information sources.

Based on an iterative bi-level estimation framework, this dissertation aims to enhance the quality of OD demand estimates by combining available historical static demand information and time-varying traffic measurements into a multi-objective optimization framework that minimizes the overall sum of squared deviations. The multi-day link traffic counts are also utilized to estimate the variation in traffic demand over multiple days. To circumvent the difficulties of obtaining sampling rates in a demand population, this research proposes a novel OD demand estimation

formulation to effectively exploit OD demand distribution information provided by emerging Automatic Vehicle Identification (AVI) sensor data, and presents several robust formulations to accommodate possible deviations from idealized conditions in the demand estimation process.

A structural real-time OD demand estimation and prediction model and a polynomial trend filter are developed to systematically model regular demand pattern information, structural deviations and random fluctuations, so as to provide reliable prediction and capture the structural changes in time-varying demand. Based on a Kalman filtering framework, an optimal adaptive updating procedure is further presented to use the real-time demand estimates to obtain *a priori* estimates of the mean and variance of regular demand patterns. To maintain a representation of the network states which consistent with that of the real-world traffic system in a real-time operational environment, this research proposes a dynamic OD demand optimal adjustment model and efficient sub-optimal feedback controllers to regulate the demand input for the real-time DTA simulator while reducing the adjustment magnitude.

DYNAMIC ORIGIN-DESTINATION DEMAND ESTIMATION AND PREDICTION FOR OFF-LINE AND ON-LINE DYNAMIC TRAFFIC ASSIGNMENT OPERATION

By

Xuesong Zhou

Thesis or 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
Ph.D.
2004

Advisory Committee:
Professor Hani S. Mahmassani, Chair
Professor Arjang A. Assad
Professor Michael O. Ball
Professor Gang-Len Chang
Professor Ali Haghani

Dedication

To Ming

Acknowledgements

I am indebted to a great number of people who generously offered advice, inspiration and friendship throughout my time at the University of Maryland and the University of Texas at Austin.

I wish to express my sincerest thanks to my advisor, mentor and role model, Dr. Hani S. Mahmassani, for his encouragement, patience, and guidance throughout the course of my Ph.D. study. My career has been strongly influenced by his dedication to excellent research work and his passion for the field of operations research and transportation science.

Special thanks go to Professor Gang-Len Chang, who provided input on my research as well as truly valued advice. I would like to thank Professor Ali Haghani for his continued support that started the first day I arrived at the University of Maryland. I would like to thank Professors Michael O. Ball, and Arjang A. Assad, members of my dissertation committee, for their participation and constructive comments.

I have been benefited from the constructive interactions with Dr. Henry Lieu in the DTA research project. I also would like to thank Rebecca A. Weaver-Gill for providing immeasurable assistance throughout my graduate program.

Very special thanks are to my colleagues Xiao Qin, Robert H. Mahfoud and Hayssam Sbayti for their contribution to the development and calibration of offline and real-time DTA systems. I am also grateful to other research team members and friends Dr. Khaled Abdelghany, Dr. Ahmed Abdelghany, Dr. Hossein Tavana, Dr. Pamela Murray-Tuitte, in particular, Dr. Yi-Chang Chiu, Nathan N. Huyhn, who were

very patient in providing the necessary guidance and advice in my early Ph.D. Study at Austin. I also owe many thanks to Chung-Cheng Jason Lu, Sevgi Erdogan, Jing Dong and Xiang Fei for their help on my research at Maryland.

I would like to express my deep appreciation to my beloved wife, Ming Zhong for her unconditional support throughout my studies. Finally, I must acknowledge and thank my parents for their long time encouragement and support.

Table of Contents

List of Tables

List of Figures

1. Introduction

1.1 Motivation

Population growth and economic development lead to increasing demand for travel and pose mobility challenges on capacity-limited transportation networks, especially in metropolitan areas. As the population of travelers and vehicle ownership rates continue to increase, traffic congestion is expected to remain as a top public concern for urban areas. It has been estimated (Lomax et al. 2002), each person in the 75 largest U.S. urban areas wastes, on average, 26 hours a year in congestion delays. According to the same source, the annual cost of congestion (based on wasted time and fuel) was an average of $520 per person in 2000. Such statistics underscore the need for effective solutions to slow the growth of congestion. By integrating advances in telecommunication, computation and information technologies, Intelligent Transportation Systems (ITS) aims to improve transportation system efficiency and traveler convenience.

Accurate OD trip desire estimates are required by the many traffic planning applications to evaluate network flow conditions that result from the travel decisions of individual travelers. Moreover, on-line applications of intelligent traffic network management call for the reliable forecasts of dynamic demand and resulting network flow states so that proactive, coordinated traffic information and route guidance instructions can be generated to network travelers for their pre-trip planning and en-route diversion. However, estimating complex traffic demand is difficult in its own right, as traffic demand can vary significantly by time of day and day of week over

different locations and evolve dramatically due to the feedback of implemented strategies. The inability of providing high quality OD demand estimates becomes a critical bottleneck in the evaluation and implementation of various promising traffic information and management scenarios, and consequently limits the potential for ITS deployments to alleviate traffic congestion and enhance mobility in urban networks.

In general, OD trip desire information can be obtained from direct interview surveys or estimated from real-time traffic surveillance data. Household, destination and roadside interview surveys, typically used in the transportation planning analysis, provide valuable samples about the detailed travel activities of each tripmaker, such as the origin and destination, the mode used and the travel time. Populating OD demand patterns from survey samples, however, is a very costly and time-consuming process. Besides, the above traditional survey methods cannot provide up-to-date dynamic demand inputs required by on-line ATMS and ATIS applications.

Thanks to the advances of surveillance, telecommunication and information technologies, the continuing deployment of Intelligent Transportation Systems offers more reliable and less costly channels to measure the complex transportation system dynamics. To name a few, these emerging technologies include vehicle identification though video image processing and Radio Frequency (RF) transponders, vehicle tracking through GPS receivers. These real-time traffic measurements provide a data rich environment to enhance modeling capabilities to capture the underling travel decision processes.

Substantial research over the last two decades has been devoted to the Dynamic Traffic Assignment (DTA) problem, which seeks to distribute given time-

dependent origin-destination (OD) trips over time and space in a transportation network according to specific user behavior and system assumptions. From the OD demand estimation perspective, these DTA models offer more realistic and tractable tools that describe the inter-relation between OD demand flows and network flow patterns. The combination of dynamic traffic assignment models and dynamic OD demand estimation models structures an integrated traffic state estimation framework that can produce consistent and realistic assignment results for general traffic networks.

To date, the potential benefits of utilizing new types of traffic measures to enhance travel demand modeling capability have not been adequately exploited, and the theoretical and algorithmic aspects on the dynamic OD demand estimation and prediction problem are still relatively undeveloped. As shown below, many fundamental issues need to be addressed to fulfill the methodological capabilities required by the off-line and on-line operation of traffic state estimation and prediction systems. These challenging questions place a greater need for flexible and systematic modeling methodologies and efficient solution algorithms.

1. How to effectively extract information from historical static demand data and from emerging point-to-point vehicle identification data?
2. How to recognize structural demand changes unfolding in complex real-world environments in addition to providing reliable OD demand forecasting under regular conditions?
3. How to accumulate the knowledge on day-to-day demand evolution from real-time estimation results?

4. How to prevent the propagation of demand estimate errors in a continually running real-time DTA system and to ensure its consistency with the real-world traffic system?

1.2 Objectives

This dissertation addresses a series of problems pertaining to the provision of accurate and reliable dynamic OD demand information for DTA planning and operational applications, namely demand estimation, prediction, day-to-day information updating as well as on-line consistency checking and updating, in the context of large-scale real-world networks with various information sources. The four principle objectives are listed as follows.

1. Estimate dynamic OD demand trips based on multiple data sources, including historical static OD demand information, multiple-day link counts, as well as vehicle identification data from point-to-point traffic detectors.
2. Formulate and develop a real-time OD demand estimation and prediction model that provides robust forecasting for both regular and irregular OD demand patterns.
3. Develop a day-to-day updating framework to adaptively and adequately capture travel demand evolution based on real-time demand estimates.
4. Construct an efficient and effective real-time demand adjustment model to maintain the consistency between the on-line DTA simulator and the actual traffic system.

The first objective is primarily intended to estimate traffic demand dynamics in the context of long-term planning and medium-term operations, where the major concern is to improve the accuracy and reliability of OD demand estimates through

effective exploitation of multiple sources of information. Typically, the OD demand estimation problem in planning applications deals with a large-scale urban network that may have thousands of links and hundreds of traffic analysis zones (TAZ). Real-time traffic surveillance data from traditional loop detectors, on the other hand, may only be available on a small subset of links in the study network, which might lead to serious identification problems in inferring a unique dynamic OD demand table for all the OD pairs. Therefore, an important and challenging problem in applying the DTA methodology to current planning practice is how to maximize the utilization of various real-time traffic measurements in conjunction with other available information sources in order to reduce the uncertainty of OD demand inputs. Historical static demand data obtained from direct household surveys or estimated from socio-economic, geographic and transportation supply characteristics, contain valuable information about the origin destination distribution of traffic demand. In addition, most traffic surveillance devices, such as loop detectors, once installed, can record multiple days of traffic data continuously at minimal additional cost. Increasing deployment of road pricing and traffic information projects also dramatically improves the public usage of RF-based vehicle tags and the availability of vehicle identification data. All of these data sources carry useful information and hold the promise of improving the identifiability of the dynamic OD demand problem.

The second, third and fourth objectives are critical in the successful deployment of the real-time traffic network state estimation and prediction systems, where dynamic OD demand information serves as an essential input for dynamic

traffic assignment. In order to provide accurate and robust demand forecasting under different traffic demand patterns, there is a great need to characterize regular demand pattern information, structural deviations and random fluctuations in traffic demand dynamics. As regular OD trip desires can be viewed as a repeated process, historical demand estimates, obtained from household interview surveys and off-line estimation results, serve as an informative source for long-range demand prediction under normal conditions. On the other hand, it is necessary to recognize the existence of structural deviations of real-time OD demand from the average pattern, which might be caused by severe weather conditions, special events, as well as the responses of travelers to information and/or other system management measures. With increasing availability, traveler information is expected to play a more active role in gradually changing day-to-day trip-making decisions and temporally re-distributing OD trip desires. In addition, random fluctuations would account for the effect of other unobserved factors and the inherent stochastic nature of daily time-varying demand. Clearly, a structural model with meaningful components can provide robust demand estimate under both regular and irregular conditions.

The third objective is to handle a common issue in the early deployment of real-time OD estimation and prediction, where only unreliable historical demand data with significant uncertainty is available, often consisting of out-of-date survey data and limited surveillance data. In this case, as the prior estimate cannot adequately describe the average conditions, the real-time estimate becomes more informative in the sense that it captures the prevailing demand pattern and encapsulates up-to-date demand information. Hence, it is advantageous to utilize new real-time estimation

results to update the *a priori* estimate of the regular pattern and adequately capture the day-to-day demand variations.

The internal traffic network representation of a real time DTA system forms a basis for evaluating the effectiveness of different management decisions. The fourth objective is related to the demand consistency checking and updating module, which aims to effectively control and reduce the propagation of demand estimation and prediction error in the DTA simulator. Without correcting OD demand estimation errors in the DTA simulator, the inconsistency in OD flows would accumulate in the DTA simulator and further propagate into the internal representation of path and link flows, making the network state prediction become highly unreliable. Therefore, an efficient feedback controller is critically required in real-time DTA systems to correct the demand prediction errors in the DTA simulator.

In brief, the overall objective of this dissertation is to provide comprehensive solutions to enhance the dynamic OD demand and prediction capabilities, and to fulfill various functional requirements of off-line and on-line DTA operation. Additionally, this dissertation seeks to apply the proposed models and solution algorithms in real-world networks in order to obtain insights on estimation and computational performance under alternative modeling formulations and implantation strategies, as well as different degrees of data availability.

1.3 Overview of Proposed Methods

To enhance the quality of OD demand estimates, this dissertation first combines available historical static demand information and time-varying traffic measurements from point detectors and point-to-point detectors into a flexible multi-

objective optimization framework that minimizes the overall sum of squared errors. The multi-day link traffic counts are also utilized to estimate the variation in traffic demand over multiple days.

To circumvent the difficulties of obtaining sampling rates in a demand population, this research proposes a novel OD demand estimation formulation to effectively exploit split fraction information provided by emerging Automatic Vehicle Identification (AVI) sensor data, and presents several robust formulations to accommodate possible departure from idealized conditions in the demand estimation process.

A structural real-time OD demand estimation and prediction model in conjunction with a polynomial trend filter is then developed and implemented to systematically model regular demand pattern information, structural deviations and random fluctuations so as to provide reliable prediction under stable condition and capture the structural changes in time-varying demand.

An optimal adaptive updating procedure is further presented to use the real-time demand estimates to capture the day-to-day variations of regular demand patterns. To maintain a representation of the network states consistent with that of the real-world traffic system, this research proposes a dynamic OD demand optimal adjustment model and efficient sub-optimal feedback controllers to regulate the demand input for the real-time DTA simulator while reducing the adjustment magnitude.

In the following, the specific problems and proposed models are addressed sequentially.

Given a traffic network with a historical demand matrix, link counts and point-to-point tagged vehicle counts, the off-line dynamic OD demand estimation problem is to find a consistent dynamic OD demand matrix that minimizes (1) the deviations between estimated traffic states and traffic measurements and (2) the deviations between estimated demand and target demand.

To capture the inter-relation between OD demand and traffic network conditions, off-line OD demand estimation models in this dissertation are based on an iterative bi-level estimation framework (Tavana, 2001). The upper-level problem is a constrained ordinary least squares problem, which is to estimate the dynamic OD demand using flow proportion matrices that map dynamic OD demand to path or link flows. Those mapping matrices are in turn generated from the dynamic traffic network assignment problem at the lower level, which is solved by a DTA simulator, namely DYNASMART-P (Mahmassani et al. 2000).

The upper-level estimation problem is to minimize the overall sum of three types of deviations. These deviation terms include

1. deviations between simulated link flows and observed link flows,
2. deviations between simulated link-to-link split fractions and observed link-to-link split fractions,
3. deviations between estimated dynamic OD demand and historical static OD demand.

For the first type of deviations, link flow proportion matrices are used to map dynamic OD flows with link flows. In addition, the link-to-link split fractions are expressed as the ratio of the sub-path flows that pass two AVI readers over the link

flows, where link-to-link flow proportion matrices are used to link dynamic OD flows to the sub-path flows. Three types of deviations are combined in a multi-objective optimization framework, and an interactive approach is used to determine an appropriate weighting scheme to fuse different information sources and find the best compromise solution.

Given historical demand information and real-world traffic measurements from various surveillance devices (e.g. occupancy and volume observations from loop detectors on specific links), the dynamic OD demand estimation and prediction problem seeks to estimate time-dependent OD trip demand patterns at the current stage, and predict demand volumes over the near and medium terms in a general network. In this research, actual dynamic OD demand is decomposed to three meaningful components in a structural state space model, that is,

true demand = regular pattern + structural deviations + random fluctuations.

In particular, the regular pattern can be estimated from archived data and updated with new demand information, while the structural deviation is modeled as a time-varying process with smooth trend. Random fluctuations can be modeled as an auto-regressive moving average component.

In order to capture possible structural changes in time-varying OD demand patterns, a polynomial trend filter is designed to estimate and predict demand deviations recursively from the *a priori* estimate of regular demand patterns. In the corresponding Kalman filtering framework, a transition equation is used to describe system evolution dynamics, while a measurement equation is to describe the mapping between observations and system variables.

At each estimation stage, the recursive prediction-correction framework updates estimation according to

new estimate = previous prediction + gain factor × (measurement – previous prediction)

where gain factor is determined optimally by taking into account mean and covariance of estimates.

Consequently, the prediction result is given by propagating mean and covariance matrices of OD demand according to the system transition equation.

Given new real-time estimates obtained every day, the day-to-day updating procedure seeks to update the *a priori* estimates of the mean and variance of the regular demand pattern. A Kalman filtering framework is applied in this research to construct this updating model to capture the day-to-day demand evolutions. This model can be naturally integrated into a real-time DTA system, providing an effective and efficient approach to utilize the real-time traffic data continuously in operational settings.

Given predicted OD demand and the deviations between actual system and simulated states, the dynamic OD demand consistency checking is to find OD demand adjustment that minimizes (1) the deviation between the adjusted link density and real-world density and (2) demand adjustment magnitude. Similar to the OD demand estimation formulation, link proportion matrices are used to map OD demand adjustment to changes of simulated link density. Several efficient solution algorithms and implementation strategies are proposed to design a robust demand consistency checking module in a real-time computation environment.

1.4 Organization of the Dissertation

The dissertation includes seven chapters. Chapter 2 provides a comprehensive review and discussion on several topics relevant to the dynamic OD demand estimation and prediction problem for DTA applications. In chapter 3, a bi-level off-line OD demand estimation formulation is presented to incorporate static demand and multi-day achieved link counts. Chapter 4 proposes several OD population demand estimation models to extract useful information from partially observed vehicle identification data. Chapter 5 develops a Kalman filtering model to systematically represent regular demand pattern, structural deviation and random fluctuation into real-time demand estimation and prediction, and then proposes an optimal day-to-day demand updating model. Chapter 6 presents predictive and reactive feedback control models for OD demand consistency checking and correction. In Chapter 7, a series of experiments are performed to demonstrate the capabilities of proposed models and provide insight on the system performance.

2. Literature Review

This chapter reviews topics relevant to dynamic OD demand estimation and prediction for off-line and on-line dynamic traffic assignment operations. After a short introduction to traditional traffic detection techniques, Section 2.1 highlights both potentials and challenges for applying emerging vehicle identification and location technologies in the context of traffic state estimation and prediction. Section 2.2 overviews major traffic state estimation and prediction approaches based on statistical models and macroscopic traffic flow models. Section 2.3 reviews the dynamic traffic assignment approach as a traffic state estimation and prediction tool, with a special focus on simulation-based DTA models. The principle functional requirements of OD demand estimation & prediction and consistency checking & updating components in on-line DTA systems are also presented. Section 2.4 reviews both static and dynamic OD demand estimation problems with different information availabilities, and it highlights the connection between dynamic OD demand estimation and dynamic traffic assignment. Various dynamic demand process models and corresponding trade-offs under regular and irregular conditions are discussed in Section 2.5, followed by a review of OD demand consistency checking and updating formulations and solution algorithms in real-time DTA systems in Section 2.6.

2.1 Traffic Surveillance Technologies

There are a number of surveillance techniques available for the traffic monitoring and management purposes, each with ability to collect and process real-time traffic data in specific types, including point, point-to-point and path

measurements. This review focuses on the essential operating characteristics of different traffic sensor types and their comparative advantages and disadvantages in traffic state estimation and prediction applications.

By counting traffic passing through a specific location during a period of time, a wide range of vehicle detection devices provide various point measurements such as lane occupancy, traffic volume, vehicle headway, as well as time-mean speed. As the earliest vehicle detection device, pneumatic tubes have been applied in traffic engineering practice since the 1930s, and they are still commonly in use as temporary counting devices. In the 1960s, many intrusive sensors such as inductive loops, magnetometers, and piezoelectric cables were introduced for automatic vehicle detection and classification. Among them, inductive loops have become the predominant vehicle detection device in the United States, due to their associated low unit equipment cost and relatively high performance. On the other hand, intrusive type detectors have to be directly installed on the pavement surface, causing considerable traffic disruption and high risk for maintenance workers during the installation and repair activities. High failure rates and significant downtimes are two other major issues in operating inductive loop detectors. For example, Bikowitz and Ross (1982) indicated that approximately 25 percent of inductive loop detectors in New York State were not functioning properly at any given time.

To overcome the disadvantages of inductive loop detectors, many roadside and overhead sensors are developed, including passive acoustic, passive infrared, microwave radar detectors. These non-intrusive devices are able to provide traffic measures without stopping traffic in installation and maintenance, but their

application is also limited by relatively high cost and high false detection rates under certain traffic or weather conditions.

Many vehicle identification devices have been developed to track the identities of vehicles through mounted transponder tags or license plate numbers when vehicles pass multiple but non-contiguous reader stations. A raw tag read typically records a vehicle ID number, the related time stamp and passing site location. If two readers at different locations sequentially identify the same probe vehicle, then the corresponding data reads can be fused to calculate the journey travel time and the counts of identified vehicles between instrumented points.

In conventional license plate surveys, part of a registration number (e.g. only last three digits) might be recorded in order to reduce manual data collection effort and avoid high recording errors when recording the complete registration number. Several statistical methods (e.g. Makowski and Sinha, 1976; Maher, 1985; Watling and Maher, 1992) have been presented to reduce “spurious matches”, which indicates that different vehicles observed at two points share the same partial registration number.

Automatic license plate matching techniques have entered the traffic surveillance field since 1970s, and many statistical and heuristic methods have been proposed to reduce reading errors and to provide reliable data association (Turner et al., 1998). Due to the difficulties in recognizing dirty and obscure characters, license-plate based vehicle identification techniques typically lead to relatively low identification rates.

Many feature-based vision and pattern recognition algorithms (e.g. Evans, 1993; Shuldiner et al., 1996; Coifman et al., 1998) have been presented to track individual vehicle trajectories using camera surveillance data. By means of vehicle signature matching techniques (Coifman, 1998), several coupled point detectors can also be used to approximate point-to-point travel measures such as link segment travel time information. However, the required high coverage densities for vehicle signature matching techniques dramatically reduce the economic feasibility of their application in a large-scale network.

Radio Frequency Identification (RFID) technologies first appeared in AVI applications during the 1980s and has become a mature traffic surveillance technology that produces various traffic measures with high accuracy and reliability. Currently, many RFID-based AVI systems are widely deployed in road pricing, parking lot management, as well as real-time travel time information provision. For instance, around 51 AVI sites were installed and approximately 48,000 tags had been distributed to users at San Antonio by 2001, corresponding to a 5% market penetration rate, while Houston's TranStar fully relies on AVI data to provide travel time information currently (Haas et al. 2001). It should be also noted that, utilizing AVI data for traffic OD volume estimation, especially in the early deployment stage, can be constrained by low market penetration rates of AVI tags. A simulation-based study conducted by Van Aerde et al. (1993) shows that low market penetration rates directly result in small data samples in statistical reference and high variances in travel time and OD flow estimates.

With advances in Geographic Information System (GIS) and telecommunication, many automatic vehicle location (AVL) technologies, such as Global Positioning System (GPS), electronic distance measuring instruments (DMI's) and cellular telephone tracking, provide new possibilities for traffic monitoring to semi-continuously obtain detailed passing time and location information along individual vehicle trajectories. As pointed out by Tavana et al. (1999), the popular use of cellular phones can dramatically increase the quality and quantity of traffic data, as a source of probe vehicle information as well as a source of live human reports. However, privacy concerns and expensive one-time installation costs are two important disadvantages influencing the AVL deployment progress.

Table 2-1 Comparison of Traffic Surveillance Devices.

Type	Categories	Data Quality	Measurements	Costs and Concerns
Loop Detector	Point Detection	High accuracy Relatively low reliability	Occupancy, volume and point speed	Low installation cost, high maintenance cost
AVI	Identification	Accuracy depends on market penetration level of tagged vehicles	Point-to-point flow information for tagged vehicles such as travel time and volume	High installation cost, public privacy concerns
GPS	Location	Accuracy depends on market penetration level of probe vehicles	Semi-continuous path trajectory for individual equipped vehicles	High installation cost, public privacy concerns

Table 2.1 summarizes the advantages and disadvantages of three representative traffic surveillance devices, namely loop detector, AVI and GPS. The primary advantage of vehicle identification and location devices is that they provide a

data rich environment for operating ATIS and ATMS. Specifically, point-to-point traffic measurements, even under low penetration levels, offer a super set of traveler's activities for traffic state estimation and prediction. Coupled with additional sources such as link counts, it can greatly improve the observability of the dynamic traffic OD demand estimation problem.

2.2 Traffic State Estimation and Prediction Models

Essentially, any application of real-time traffic measurements for supporting ATIS and ATMS functionalities involves estimation and/or prediction of traffic states. Depending on underlying traffic process assumptions, the existing traffic state estimation and prediction models can be classified into three major approaches.

1. Approach purely based on statistical methods, focusing on traffic flow/ travel time forecasting on a freeway segment or an arterial street.

2. Approach based on macroscopic traffic flow models, focusing on traffic flow estimation on successive segments of a freeway corridor.

3. Approach based on dynamic traffic assignment models, focusing on wide-area estimation and prediction of origin-destination demand, route choice probabilities, as well as resulting traffic network flow patterns.

Although the focus of this dissertation is mainly on the DTA based methods, an introduction to the first two approaches is useful for obtaining insights into many modeling issues in OD demand estimation and prediction. The pure statistical approach is mainly used to estimate and predict local traffic conditions for route guidance, adaptive ramp metering and signal control. Several early traffic signal control systems, such as UTCS-2 (FHWA, 1973), directly used historical normal

pattern to make traffic prediction for the future control horizon (5-15 min). The application of time series methods in traffic prediction started from Ahmed and Cook's work (1979), which predicts traffic occupancies in the context of incident detection. Focusing on Autoregressive Integrated Moving Average (ARIMA) models, they pointed out that first order differencing is adequate to remove the non-stationarity of original time series, and a complex model is not necessarily more accurate than a simple moving average model. The follow-up study by Ahmed (1989) further showed that model ARIMA(0,1,3) outperforms the double exponential and the moving average models, as well as the adaptive-exponential model in predicting traffic volume and density.

In order to combine both the historical data and real-time observations into the estimation and prediction process, Okutani and Stephanedes (1984) applied the Kalman filtering theory in dynamic traffic flow prediction, where the deviations from the historical average are considered as state variables that are assumed to follow an Autoregressive process.

The spectral analysis method proposed by Nicholson and Swann (1974) is able to provide reliable prediction under stable traffic conditions, but fails to capture rapid traffic flow changes. Davis and Nihan (1991) used nonparametric statistical methods (e.g. pattern recognition algorithms) to forecast freeway traffic congestion. Recognizing that most of the statistical based models only deal with local traffic conditions, Whittaker et al. (1997) attempted to apply a multivariate Kalman filtering model to predict freeway traffic flows on several links, but major difficulties arise in modeling the interaction of traffic flow process between links. Specifically,

autocorrelation, partial autocorrelation, cross-correlation characteristics, and order of parameters (i.e. time lags) between link flows cannot be determined in advance, especially for congested traffic. In general, these purely inductive methods have difficulties in providing complete information for a partially observed system, as they lack the underlying descriptive models that relate unobserved states to traffic measurements.

Based on macroscopic traffic flow models, the second approach aims to estimate and predict traffic flow, density and queue lengths on each link segment of the corridor under consideration, given a freeway corridor with partially loop detector coverage. Early studies started by Gazis and Szeto (1972) for traffic state estimation on short road sections. More realistic and detailed models (e.g., Cremer and Papageorgiou, 1981) were proposed and applied in the early 1980s. Most of the studies in this category adopted the Kalman filtering technique to estimate traffic state variables on each section under the assumption that on-ramp flows, off-ramp flows, and entering flows from the upstream section are known. As traffic flow models are integrated into the transition equation to describe the traffic flow dynamics evolution, the resulting nonlinearity in those models leads to extended Kalman filtering formulations and solution algorithms that involve complex linearization.

The DTA-based approach detailed in the following section represents a more comprehensive solution methodology to estimate and predict the time-varying traffic flow pattern over a general network.

2.3 Dynamic Traffic Assignment Models

Wardrop (1952) proposed the user equilibrium and system optimal principles of route choice behavior in his seminal paper, and Beckman et al. (1956) formulated the static user equilibrium traffic assignment problem as an equivalent convex mathematical programming problem. Since their influential contributions, the development of the static network assignment formulations, algorithms and applications have made remarkable progress. The books by Sheffi (1985) and Patriksson (1994) provide the most comprehensive coverage on the static traffic assignment problem and its variants.

Recognizing the limitations of static traffic assignment models in describing the dynamics of network flow propagation and dynamic travel behavior in response to real-time information, dynamic traffic assignment models have attracted active research and development attention, starting from Merchant and Nemhauser's pioneer work (1978a, b) that presented a mathematical programming formulation for a one-destination system optimal DTA problem. Peeta and Ziliaskopoulos (2001) provided a state-of-art review and detailed discussions of formulation approaches, model objectives, underlying assumptions, solution methodologies, traffic flow modeling strategies as well as operational requirements and capability.

The existing DTA methodologies are classified into two major groups: analytic approach and simulation based approach, and the former line further includes three types of formulations: mathematical programming, optimal control and variational control. The analytic approach has the potential on deriving theoretical insights. However, well-behaved mathematical formulations are currently unavailable

and principal difficulties occur in how to ensure the FIFO property and how to preclude the holding of traffic in SO assignment.

In order to avoid mathematical intractability, simulation-based dynamic traffic assignment is intended to capture dynamic tripmaker decisions and complex traffic processes in the practical deployment for realistic networks. For instance, a simulation based DTA system, DYNASMART (Mahmassani et al., 1994), can allow (1) a richer representation of traveler behavior decisions, (2) an explicit description of traffic processes and their time-varying properties, and (3) a more complete representation of the network elements, including signalization and other operational controls.

The simulation based DTA methodology has been further extended into the design of real-time DTA systems in order to provide real-time traffic state estimation and prediction for ITS network applications. In DYNASMART-X developed by Mahmassani et al. (1998), several components and modules are integrated to perform the following DTA functional capabilities.

1. Estimate the current traffic states in the network.
2. Provide future network traffic states for a pre-defined horizon in response to various control and information dissemination strategies.
3. Estimate dynamic OD demand in the current stage
4. Forecast dynamic OD demand in future stages
5. Maintain the consistency between the internal representation of the real-time DTA model and the real world system

FIGURE 2.1 depicts a high-level view of the DYNASMART-X system structure and interrelationships among the components and modules. Through a Common Object Request Broker Architecture (CORBA), the above components are integrated into the real-time DYNASMART-X system.

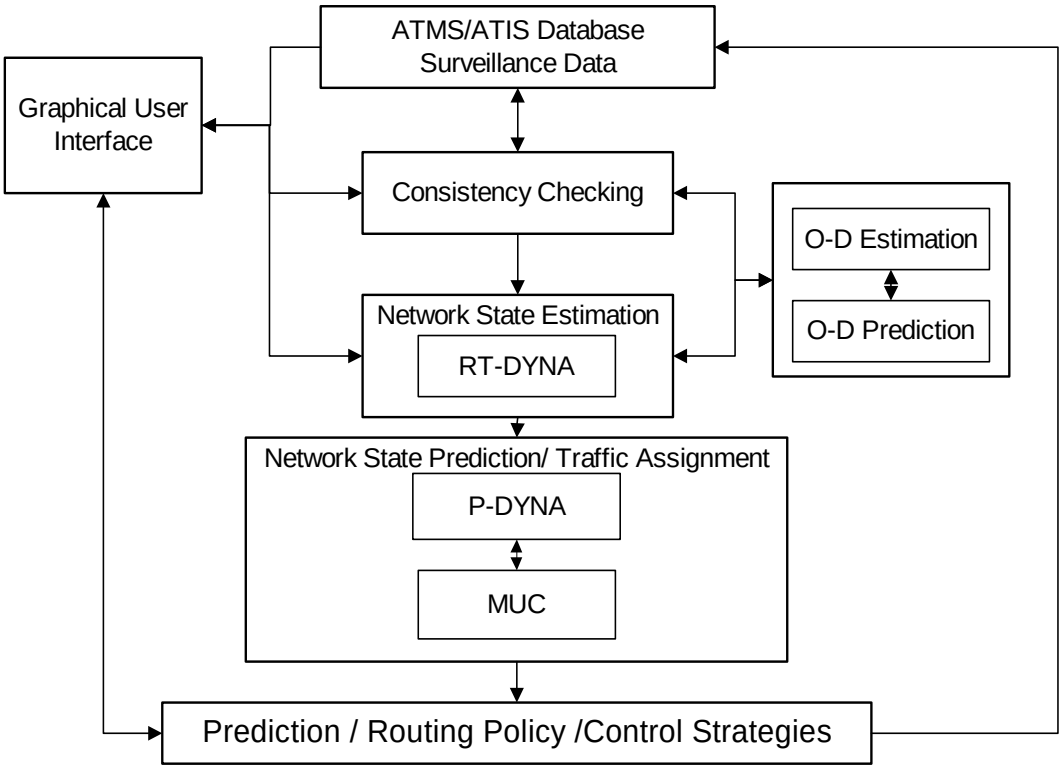

FIGURE 2-1 DYNASMART-X Functional Diagram.

Peeta and Mahmassani (1995) introduced a rolling-horizon solution procedure for solving the real-time DTA problem. This operational method has been detailed and implemented by Mahmassani et al. (1998) and Mahfoud (2002) using an asynchronous multi-horizon architecture, where multiple modules (tasks) execute in different processors and communicate with each other asynchronously. Different modules can run with different execution cycles, and each cycle length is determined

by the corresponding operational needs (e.g. providing descriptive information versus normative information) and resource constraints (e.g. input/output dependency).

The asynchronous multi-horizon architecture and various components in DYNASMART-X are described in detail as the following. Given currently available OD demand information, the network state estimation (RT-DYNA) simulates network flow patterns periodically (every *assignment interval*). The network state prediction module is executed less frequently (every *roll period*), in which P-DYNA receives the current network condition in RT-DYNA as its initial condition and projects the network state for a period in the future (*stage length*). The Multiple User Class (MUC) assignment algorithm is incorporated into the network state prediction component to provide new routing information and control strategies (Peeta and Mahmassani, 1995). Based on the predicted OD demand, every *roll period* a new network state is forecasted, which overwrites the one obtained from the previous roll.

Interfacing with traffic detectors and probes on the transportation network, the OD Estimation (ODE) module is activated every *OD observation interval*. The complete OD demand prediction-correction cycle operates every *OD estimation period*, which might consists of multiple *OD observation intervals*. The length of *OD estimation period* is constrained by the sampling rate of the traffic surveillance system and the computation time for estimating the related OD demand. In addition, an *OD estimation period* should be long enough to adequately model the correlated measurement noise among consecutive *OD observation intervals*. As ODE requires the link proportions generated from P-DYNA, its *estimation period* should be short

enough to accommodate the computation time required by P-DYNA in predicting the network state for the entire *stage length*.

Based on the *a posteriori* estimate of OD demand from ODE, the OD Prediction (ODP) module produces the time-dependent OD desires in the network to be used in the simulation-assignment procedure for relatively long time duration (*prediction horizon*). When OD estimation completes its execution, OD prediction should starts immediately to make new demand information available. On the other hand, ODP can be also periodically launched to ensure that predicted OD demand always readily available for RT-DYNA and P-DYNA.

Utilizing up-to-date surveillance data from different traffic sensor sources, the consistency checking and updating module aims to minimize the deviations between the real-world measurements and the simulated states. In particular, a Short-Term Consistency Checking (STCC) model is designed to regulate traffic flow propagation in the DTA simulator, while a Long-Term Consistency Checking (LTCC) model adjusts demand input that is fed into the DTA simulator. These two components run periodically and their frequencies can vary depending on computation requirements and adjustment accuracy.

Figure 2.2 depicts one execution sample in the asynchronous multi-horizon implementation of DYNASMART-X. The cycles for the modules are defined as follows. The state estimation module and short-term consistency checking are run every 30 seconds. The state prediction (P-DYNA) is run every 5 minutes with a stage length of 20 minutes. An OD observation interval is 5 minutes. Every 10 minutes, the OD estimation procedure starts estimation of the current stage and then prepares the

estimated demand state for the OD prediction module, which in turn predicts demand for the next 30 minutes. Long-term consistency checking corrects the predicted OD demand based on real-time traffic surveillance data every 5 minutes.

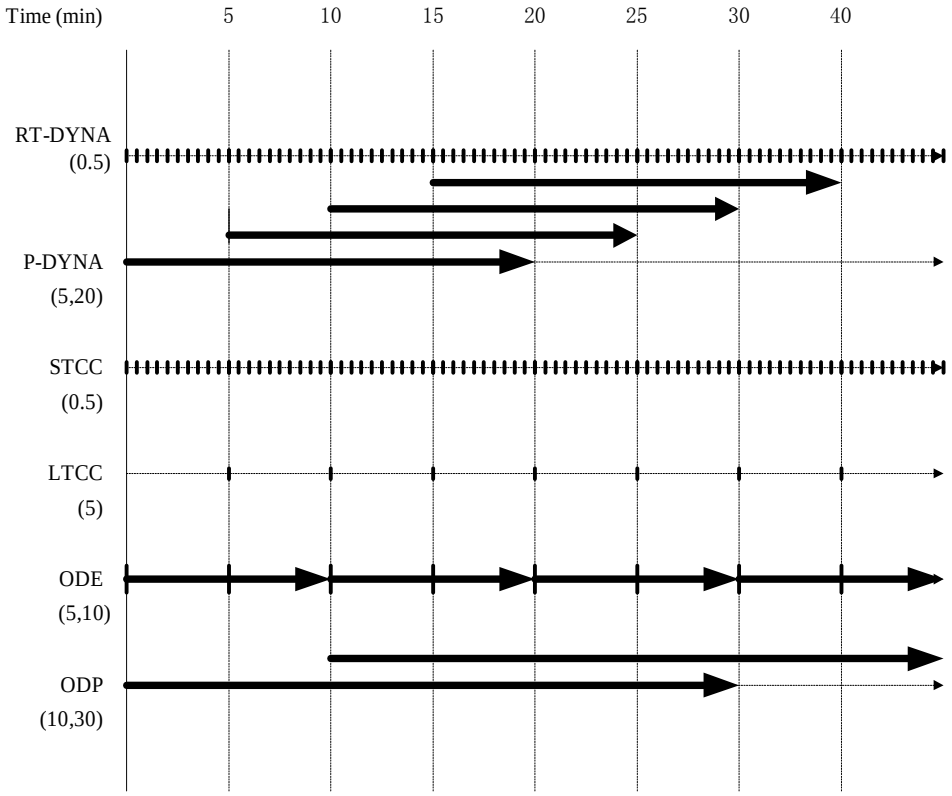

FIGURE 2-2 Asynchronous Multi-Horizon Implementation of DYNASMART-X

2.4 OD Demand Estimation Models

The general OD demand estimation problem is to find an estimate of OD demand matrix by effectively utilizing traffic flow observations and other available information. The review focuses on different estimation and different assumptions for the OD demand estimation problem under different information sources.

Existing OD demand estimation models belong to two major categories: static models or dynamic models. Assuming constant trip desires over the estimation horizon, static OD demand estimation models estimate a static OD demand table

based on daily or hourly average traffic counts. To realistically represent traffic formation and congestion on the traffic network, dynamic models utilize time-varying traffic flow observations to estimate traffic demand that varies over time.

2.4.1. Static OD Demand Estimation

Early studies (e.g. Van Zuylen and Willumsen, 1980) assumed that averaged link counts follow a Poisson distribution, and used the entropy minimizing principle to construct the OD estimation problem that minimizes the log likelihood function (2.1) subject to link volume observation constraint (2.2), where C is the vector of link counts, D is the OD demand matrix and A(•) is the traffic assignment function providing a mapping matrix between OD flows and link flows. The assignment mapping matrix can be obtained by combining route choice proportions and the link path incidence matrix.

$$\text{Ln } L = \text{Sum over } [C \ln \text{A}(D) - \text{A}(D)] + \text{constant} \qquad (2.1)$$

$$\text{Subject to } C = \text{A}(D) \qquad (2.2)$$

By assuming a multivariate Normal distribution for traffic counts, Maher (1983) and Cascetta (1984) proposed a Bayesian estimator and a generalized least squares (GLS) estimator, respectively. If the error terms are Normal variables with zero mean and variance-covariance matrix W, the corresponding log likelihood function is

$$\text{Ln } L = -\frac{1}{2}[C - \text{A}(D)]^{\text{T}} W^{-1}[C - \text{A}(D)] + \text{constant.} \qquad (2.3)$$

In a real traffic network, the number of links with observations is more likely to be less than the number of unknown OD pairs, and static OD demand estimation

that purely relies on averaged link counts might lead to an underdetermined system. As a result, additional information must be supplemented to find a unique OD demand estimate. *A priori* information on trip demand, which typically comes from either sample surveys or outdated estimates, has been widely used as an important supplementary data source. A Bayesian approach (Maher, 1983) and a GLS estimator (Cascetta 1984) can be used to combine traffic counts and target demand, leading to an optimization problem as the following form:

$$\text{Min } [C - \text{A}(D)]^{\text{T}} W^{-1} [C - \text{A}(D)] + [D - D']^{\text{T}} Z^{-1} [\text{D} - \text{D}'] \qquad (2.4)$$

Subject to $D \geq 0$

where W and Z denote dispersion matrices and D' is the target demand.

By assuming that sample N follows a multinomial distribution with known sampling fractions α, Spiess (1987) proposed an OD demand estimation model to incorporate trip sample counts, collected from household or origin surveys. The resulting log likelihood is

$$\text{Ln } L = \text{Sum over } [N \ln (\alpha D)] + \text{constant.} \qquad (2.5)$$

Cascetta and Nguyen (1988) gave an excellent review for estimating static OD demand matrix using traffic counts.

To capture congestion effects in traffic networks, many researchers attempted to integrate equilibrium assignment into the static OD demand estimation process. Nguyen (1977) and LeBlanc and Farhangian (1982) incorporated link count observations into a variable demand UE assignment program as equality side-constraints so that the estimated link flows can reproduce observed link counts. Along this direction, Bell (1993) used link flow observations to construct inequality side

constraints for a Logit based stochastic assignment model, and proposed Lagrangian solution algorithms. Fisk (1989) combined the entropy maximum model with an UE assignment program to construct a bi-level mathematical programming problem. Florian and Chen (1992) and Yang (1992) further presented a more flexible bi-level framework to estimate consistent OD demand, where the upper level is a GLS-based OD estimation model and the lower level is an UE assignment program.

2.4.2. Dynamic OD Demand Estimation

Substantial research efforts have been devoted to the dynamic demand estimation problem over the past 20 years. An important research direction is to improve the representation capability of the estimation model to adequately describe dynamic traffic and behavioral processes in networks. Accordingly, existing models can be grouped into two classes (Chang and Tao, 1999): DTA based vs. non-DTA based, depending on whether a DTA component is incorporated into the estimation process.

Early research methods are proposed to estimate time-dependent OD flows on individual components such as a single intersection or a freeway corridor. Cremer and Keller (1981, 1984 and 1987) developed four methods for identification of dynamic origin-destination flows, including a least squares estimator, a constrained optimization method, a simple recursive estimation formula and a Kalman filtering method. All these four methods estimate dynamic OD split fractions based on the entry and exit flow measurements, under the simplifying assumption of constant link travel time.

Bell (1983) introduced additional travel time dispersion parameters to relax the constant travel time assumption, and proposed a constrained weighted least squares formulation to estimate dynamic OD flows. In order to establish a more realistic measurement equation, Chang and Wu (1994) further used nonlinear macroscopic traffic relations to compute segment density, speed and travel time. Extending the concepts and solution methodologies of the static OD estimation problem, Cascetta et al. (1993) proposed a generalized least squares (GLS) estimator for dynamic OD demand in a general network. A simplified assignment model was used in this study, in particular, path choice fractions are first calculated from a route choice model and then the resulting path flows are propagated to link flows based on link travel times. Along this line, Ashok (1995) introduced stochastic assignment matrices into the demand estimation process to take into account measurement errors in travel times and inaccuracies in route choice models.

Growing interest in the application of simulation-based dynamic traffic assignment (DTA) models has been accompanied by research into the estimation of dynamic OD trip desires. Tavana (2001) proposed a bi-level generalized least squares optimization model and an iterative solution framework to estimate dynamic OD demand while seeking to maintain internal consistency between the upper-level demand estimation problem and the lower-level dynamic traffic assignment problem. In addition, a nonlinear least squares formulation was proposed to explicitly consider dependence of link flow proportions on demand flows, where derivatives of link flow proportions with respect to OD demand are obtained from a DTA simulation program, namely DYNASMART (Mahmassani et al., 1994). Tavana (2001) and

Peeta and Ziliaskopoulos (2001) offered extensive literature reviews of the dynamic OD demand estimation problem and its inherent connection to the dynamic traffic assignment problem.

If the number of independent link flow observations is less than the number of cells in the unknown dynamic OD demand matrix, other available information sources, such as *a priori* estimates of OD demand matrices, must be combined to ensure identifiability of the corresponding OD demand problem. Cascetta (1993) presented a bi-objective formulation that minimizes both (1) deviations between estimated link flow and observed traffic counts, and (2) deviations between estimated dynamic OD demand and historical or assumed *dynamic* OD demand. It should be noted that this formulation does not provide an explicit model to utilize historical *static* demand tables, which are typically available in conventional planning applications. By using prior information on OD demand flows, Tavana (2001) proposed several Bayesian inference formulations to update the OD demand estimates obtained from link flow observations. The use of the flow counts across screen-lines and cordon-lines in dynamic OD demand estimation was introduced by Chang and Wu (1994) and Chang and Tao (1996), in order to extract and supply more information from existing traffic surveillance and survey data.

Vehicle identification data represent another important and emerging data source for estimating dynamic OD demand. Two classes of demand estimation problems using vehicle identification data should be distinguished: the estimation of tagged vehicle demand and the estimation of population demand. If AVI data are collected based on the license plate numbers, where all vehicles have unique vehicle

identification code, then two types of problems are equivalent. In a transponder based AVI system, only a subset of vehicles are equipped with transponder tags, so the second problem needs to be explicitly considered to estimate the population trip desires.

Several studies are devoted to the first class of problem. Based on transponder tag data collected from a freeway corridor in Houston, Dixon and Rilett (2002) applied Cascetta's framework (1993) to calculate link flow proportions based on observed travel time from AVI counts, and they presented both off-line GLS and on-line Kalman filtering models for estimating tagged OD demand. Due to budget constraints, AVI readers in most applications are installed at limited critical locations, such as highway/freeway corridors and on-ramps and off-ramps, indicating that point-to-point counts rather than complete origin-destination counts are more likely to be available for OD estimation in a general network. Antoniou et al. (2004) introduced path-flow proportion matrices that relate OD demand flows to sub-path tag counts, and extended Ashok's framework (1996) to estimate and predict tagged vehicular OD demand flows.

Several models have been developed to the estimation of population demand using AVI counts. Recognizing the low identification rate associated with license plate based AVI data, Van der Zijpp (1997) proposed a constrained optimization formulation to jointly estimate unknown OD demand flows and identification rates. Along the same line, Asakura et al. (2000) provided an off-line least squares model to simultaneously determine OD demand and location-dependent identification rates, and further investigated day-to-day fluctuations in estimated OD demands. Dixon

(2000) proposed a three-stage procedure to estimate population OD demand from transponder based AVI data: (1) estimate the tagged OD demand matrix from AVI data, (2) estimate market penetration rates using AVI data and link counts, and then (3) scale the estimated tagged vehicle demand to the total population demand using estimated market penetration rates. In brief, the above models have to estimate either market penetration rates or identification rates so as to relate the AVI samples to population demand using a multiplicative function structure. The estimation of market penetration rates or identification rates, however, is a difficult problem in its own right, as these two types of rates are essentially time-dependent and location-dependent random variables. Moreover, the estimation error of market penetration rates and identification rates can dramatically impact the reliability of the final population demand estimate.

2.5 Dynamic OD Demand Prediction Models

Existing OD demand prediction models can be categorized according to the underlying assumptions in representing dynamic demand processes. Given historical mean (i.e. regular pattern) D_k^r, and real-time estimate $D_{k-p|k}$ at stage k where $p \geq 0$, the short term traffic state forecasting problem is to predict the future value of $D_{k+h|k}$ at stage k, where prediction horizon $h>0$.

Assuming that the deviations of flows from historical averages come from a stationary time series, Okutani and Stephanedes (1984) applied an auto-regressive model shown as Equation 2.6 to forecast time-varying traffic flows.

$$D_{k+1|k} = D_{k+1}^r + \sum_{p=0}^{q} \phi_{k-p} (D_{k-p|k} - D_{k-p|k}^r) + w_k \qquad (2.6)$$

where ϕ_{k-p} are predetermined autoregression coefficients, w_k is system evolution noise and q is the maximum time lag.

Along the same line, Ashok and Ben-Akiva (1993) formulated the deviations of OD demand from historical averages as AR processes, and further developed a Kalman filter for real-time OD demand estimation and prediction, in which a 4^{th}–order AR model is adopted based on calibration results from several data sets. Ashok and Ben-Akiva (2000) extended the above framework to construct an alternative demand estimation and prediction formulation that models origin trips and destination fractions separately. Clearly, an autoregressive model is suitable to describe a stationary random process with constant mean and variance. However, if the prevailing OD demand is structurally different form the regular demand pattern, demand deviations will not satisfy the fundamental stationarity assumption for AR processes, and such non-stationarity could seriously degrade the overall prediction performance. Moreover, an AR type model with high-order terms requires extensive off-line calibration effort for the autocorrelation coefficients, and the corresponding augmented state space dramatically increases the on-line computational burden, especially for large-scale network applications.

Alternatively, without requiring prior demand information, a simple random walk model can be built relatively easier for short-term demand prediction, corresponding to an AR(1) model with autocorrelation coefficient of 1. Cremer and Keller (1981, 1987) and Chang and Wu (1994) applied the random walk model

$$D_{k+1|k} = D_{k|k} + w_k \tag{2.7}$$

to predict dynamic OD flow split parameters, by directly extending the latest estimates as the future forecasts.

In order to describe the nonlinear trend characteristics in dynamic OD demand, Mahmassani et al. (1998) and Kang (1999) proposed a polynomial trend filter to estimate time-dependent OD flows on a general network, while historical data were used to calibrate polynomial function parameters in demand evolution processes.

$$D_{k+\zeta|k} = \sum_{p=0}^{m} \beta_p \zeta^m = \beta_0 + \beta_1 \zeta + \beta_2 \zeta^2 + \cdots \beta_p \zeta^p + \cdots + \beta_m \zeta \qquad (2.8)$$

where m is the order term, β_p is the polynomial function parameter.

Majority of real-time Demand estimation studies focus on how to generate reliable and accurate demand estimates, with little attention paid to the procedure of utilizing newly generated real-time demand estimates. Ashok (1996) suggested several heuristic approaches to update the mean estimate of historical demand with recent real-time estimates. To the author's knowledge, no model has been proposed to utilize both the first-order and second-order statistics (i.e. mean and variance) of real-time estimates to update the *a priori* estimate of the regular demand pattern. There is a need to develop a satisfactory updating model that can effectively and systematically extract the historical demand information for capture day-to-day demand evolution and provide up-to-date regular demand reference for on-line estimation and prediction purposes.

2.6 Dynamic OD Demand Consistency Checking and Updating Models

Network state consistency over time is one of major concerns in a real-time DTA system. Doan, Ziliaskopoulos and Mahmassani (1999), Kang (1999) and Peeta and Ziliaskopoulos (2001) enumerated possible error sources that can cause the potential divergence of the predicted system state from the actual traffic conditions unfolding on-line. These error sources can be categorized into the following groups: (i) incorrect prediction of the dynamic OD demand, (ii) incorrect route choice predictions, (iii) incorrect traffic flow modeling, (iv) incorrect assumptions on driver behavior and/or response to information provided, (v) unpredicted incidents, incorrect assumptions on system related parameters, (vi) noise and/or sparsity in measurements, as well as (vii) failure of the ATIS system components.

Obviously, the demand estimation/prediction errors can sequentially propagate to path and link flow representations along the assignment and simulation processes. From the perspective of OD demand estimation and prediction, the role of DTA simulators is to provide a process model that links demand states to traffic measurements. The discrepancy between the simulated states and the real-world system indicates mismatch in the descriptive DTA process model, directly affecting reliability of the mapping matrices (e.g. link flow proportions that relate OD demand flows to link flows) in the measurement equation. Without correcting these errors, the demand error in the DTA simulator can be accumulated and amplified over time in the recursive estimation and prediction structure. This operational issue in the deployment of a real-time DTA system has motivated the development of OD demand consistency checking and updating models.

Kang (1999) designed a diagnostic architecture for checking system consistency in a real-time DTA system. Based on a proportional-integral-derivative (PID) feedback control framework, Mahmassani et al. (1998) and Kang (1999) further proposed a real-time demand consistency updating module, which heuristically adjusts the demand level according to the discrepancy between simulated and observed link density.

Peeta and Bulusu (1999) proposed a mathematical programming approach for ensuring on-line consistency, which seeks to minimize deviations between real-time traffic measurements and predicted network states. The OD demand adjustment is calculated by iteratively solving a deterministic DTA problem and a least squares problem within a stage-based rolling horizon framework. The resulting rank deficient least squares problem is solved by a generalized singular value decomposition strategy, leading to intensive computational requirements.

The existing models and algorithms discussed here for maintaining OD demand consistency have not provided an entirely satisfactory approach that meet the following critical requirements: (1) an effective optimization formulation that lead to a unique adjustment solution, (2) efficient and robust algorithms applicable to large-scale realistic networks.

2.7 Summary

This chapter has reviewed several topics relevant to the dynamic OD demand estimation and prediction problem for off-line and on-line DTA applications. A wide range of traffic surveillance data sources available to the OD demand estimation and prediction process are first discussed, followed by the review of three traffic state

estimation and prediction approaches using real-time traffic data. The modeling and solution methodologies for dynamic traffic assignment are further detailed, with a special focus on its real-time system structure and implementation strategies. The remaining sections sequentially review the static OD demand problem, dynamic OD demand estimation problem, as well as dynamic OD demand prediction problem. In particular, this review discusses the connection between OD demand estimation and traffic assignment, and various demand process models for OD demand prediction. Section 2.7 overviews OD demand consistency checking and updating models, which serve as the essential supporting component that corrects OD demand estimation/prediction errors in on-line DTA systems.

3. Dynamic Origin-Destination Demand Estimation Using Static Historical OD Demand and Multi-Day Link Counts

3.1 Introduction

Based on an iterative bi-level estimation framework, this chapter aims to enhance the quality of OD demand estimates by efficiently combining available historical static demand information and time-varying traffic point measurements from multiple days into a flexible multi-objective optimization framework that minimizes the overall sum of squared errors. Section 3.2 first proposes a dynamic OD estimation model to extract information from a historical static OD demand matrix. Section 3.2 also discusses several possible strategies for ensuring identifiability of the estimation problem in DTA planning applications, and an interactive approach is also presented to determine the appropriate weighting scheme and find the best compromise solution. The one-day demand estimation formulation is extended to an estimation model in Section 3.3 for inferring day-to-day demand variations by using multi-day link counts, followed by a description of the related hypothesis testing procedure. The effectiveness of the proposed methods is illustrated in Section 3.4 using available real-world measurements in a real network.

3.2 Model Framework using Static Historical OD Demand Information

The model presented in this chapter is an extension of the iterative bi-level OD demand estimation framework proposed by Tavana and Mahmassani (2001) and Tavana (2001). Specifically, the upper-level problem is a constrained ordinary least squares problem, which is to estimate the dynamic OD demand based on given link

flow proportions. The link flow proportions are in turn generated from the dynamic traffic network loading problem at the lower level, which is solved by a DTA simulation program, namely DYNASMART-P (Mahmassani et al., 2000).

The following notation is used to represent all the variables in the demand estimation formulation. This section is only concerned with demand estimation using one-day link counts, so the subscript of day *m* is dropped for simplicity.

l = subscript for links with traffic flow measurements, l=1,…,L.

L = number of links in the network that have flow measurements.

τ= subscript for departure time intervals, τ =1, 2, …, $T_{\mathrm{d.}}$

t= subscript for observation time interval, i.e. sampling time interval t =1,…, $T_{\mathrm{o.}}$

T = number of aggregated departure time intervals in the estimation period.

i = subscript for origin zone, i =1,…, I.

I = number of origin zones in the network.

j = subscript for destination zone, j =1,…, J.

J = number of destination zones in the network.

m = subscript for day of week.

M = number of days of week, m =1,…, M. Here, M = 5, representing Monday through Friday.

$c_{(l,t),m}$ = measured traffic volume on link l, during observation interval t, on day m.

C_m= vector of measured flows on the links, consisting of element $c_{(l,t),m}$.

$d_{(i,j,\tau),m}$ = demand volume with destination in zone j, originating their trip at zone i during aggregated departure interval τ on day m.

D_m= vector of OD demand flows, consisting of elements $d_{(i,j,\tau),m}$ on day m.

$p_{(l,t),(i,j,\tau),m}$ = link flow proportions, that is the proportion of demand flow $d_{(i,j,\tau),m}$ that flows onto link l during observation interval t.

P_m= matrix of link flow proportions, consisting of element $p_{(l,t),(i,j,\tau),m}$.

$\varepsilon_{(l,t),m}$ = the combined error terms in estimation of traffic flow on link l during observation interval t on day m.

E_m= vector of combined error terms, consisting of elements $\varepsilon_{(l,t),m}$ for link flow.

$g_{(i,j)}$ = target demand, which is the total traffic demand during period of interest for each origin-destination pair (i, j).

G = target demand vector, which is a vector of total traffic demand during period of interest, consisting of elements $g_{(i,j)}$.

$\eta_{(i,j),m}$ = the combined error terms in estimation of total traffic demand during period of interest from zone i to zone j, on day m.

A = mapping matrix between time-dependent demand and total demand.

$\prod_m$ = vector of combined error terms, consisting of elements $\eta_{(i,j),m}$ for total traffic demand during period of interest.

Two objectives are considered in this formulation. The first one is to minimize the deviation between observed link flows and estimated link flows, as shown in Equation (3.1a) or (3.1b). The second objective is to minimize the deviation between the target demand and estimated demand. Suppose that the target demand is a historical static demand table for the entire study horizon, so the second objective function can be explicitly written as the difference between the static demand and the sum of dynamic demand over the study period, as shown in Equation (3.2a) or (3.2b).

$$C = P \cdot D + E \tag{3.1a}$$

$$\text{or} \quad c_{(l,t)} = \sum_{i,j,\tau} p_{(l,t),(i,j,\tau)} d_{(i,j,\tau)} + \varepsilon_{(l,t)} \tag{3.1b}$$

$$G = A \cdot D + \Pi \tag{3.2a}$$

$$\text{or} \quad g_{(i,j)} = \sum_{\tau} d_{(i,j,\tau)} + \eta_{(i,j)} \tag{3.2b}$$

From a multi-objective programming standpoint, the above bi-objective programming problem can be transformed into a single-objective problem by either a weighting formulation or a ε-constraint formulation. The former leads to a relatively simple quadratic programming problem, which coincides with an ordinary linear regression model, while the latter introduces hard nonlinear constraints if the deviation is represented by the squared error. The weighted formulation is adopted to combine the two sets of deviations, with respective weights *w* and (1-*w*) for the first and second objectives. The weights *w* and (1-*w*) could be interpreted as the decision maker's relative preference or importance belief for the different objectives; they could also be considered as the dispersion scales for the first and second error terms

in the ordinary least-squares estimation procedure. In general, if the provided target demand is not reliable, i.e. the error term $\eta_{(i,j)}$ has a high variance, a small value of *w* is used; and vice versa. The resulting bi-level dynamic OD estimation problem with a single day of link-level observations is presented in Equations (3.3) and (3.4), which is to minimize the combined deviations, subject to the dynamic traffic assignment constraint and non-negativity constraints for demand variables.

$$\min Z = \left\{ (1-w) \sum_{l,t} [\sum_{i,j,\tau} p_{(l,t),(i,j,\tau)} \cdot d_{(i,j,\tau)} - c_{(l,t)}]^2 + w \sum_{i,j} [\sum_{\tau} d_{(i,j,\tau)} - g_{(i,j)}]^2 \right\} \quad (3.3)$$

$$\text{s.t.} \quad p_{(l,t),(i,j,\tau)} = \text{ass}\mathit{ignment}\left\lfloor d_{(i,j,\tau)} \right\rfloor \quad \text{from DTA, } \forall l,t,i,j,\tau \quad (3.4)$$

$$d_{(i,j,\tau)} \geq 0, \ \forall i,j,\tau$$

where *w* is a positive weight.

If a time-dependent demand matrix is available a priori, the above formulation can be written as Equation (3.5), where $g_{(i,j)}$ is extended to $g_{(i,j\tau,)}$ for each departure time interval:

$$\min Z = \left\{ (1-w) \sum_{l,t} [\sum_{i,j,\tau} \hat{p}_{(l,t),(i,j,\tau)} \cdot d_{(i,j,\tau)} - c_{(l,t)}]^2 + w \sum_{i,j,\tau} [d_{(i,j,\tau)} - g_{(i,j,\tau)}]^2 \right\} \quad (3.5)$$

A natural attempt would be to split the given static demand $g_{(i,j)}$ into equal portions of $g_{(i,j\tau,)} = \frac{g_{(i,j)}}{T}$ for each time interval and use Equation (3.4), which has a similar structure as the static OD estimation case. However, this scheme would implicitly impose a uniform temporal pattern on the target demand, thereby biasing the resulting estimation. More precisely, let us define the combined error term

$\eta_{(i,j,\tau)} = d_{(i,j,\tau)} - g_{(i,j,\tau)}$ for each departure time interval, then $\eta_{(i,j)} = \sum_{\tau} \eta_{(i,j,\tau)}$. The bias for the latter formulation with respect to the previous formulation for each OD pair (i, j) is shown in Equation (3.6).

$$\sum_{\tau} \left[d_{(i,j,\tau)} - g_{(i,j,\tau)} \right]^2 - [\sum_{\tau} d_{(i,j,\tau)} - g_{(i,j)}]^2 = \sum_{\tau} \eta_{(i,j,\tau)}^{\ 2} - [\sum_{\tau} \eta_{(i,j,\tau)}]^2 = -2 \sum_{\tau 1} \sum_{\tau 1 \neq \tau 2} \eta_{(i,j,\tau 1)} \eta_{(i,j,\tau 2)} \quad (3.6)$$

The iterative solution algorithm for the proposed bi-level programming problem is briefly described as follows.

Step 1: (Initialization) $k = 0$. Start from an initial guess of the traffic demand matrix D_0, obtain link flow proportions P_0 from the DTA simulator.

Step 2: (Optimization) Substituting link flow proportions P_k, solve the dynamic OD estimation problem as Equation (3.3) to obtain demand D_k.

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

Step 4: (Evaluation) Calculate the deviation between simulated link flows and observed link counts, and calculate the deviation between estimated demand D_k and target demand.

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

In the following, two key questions are addressed using the above formulation in planning applications. One is how to assess the weight *w*, the other is how to deal with estimation with partial observation.

3.2.1 Retrieving Best Compromise Solution

It may be possible to obtain the least-squares estimate of the weight value through linear regression. However, in planning analysis, it is more desirable to incorporate the planners' knowledge and experience in the estimation process, reflecting different degrees of confidence in the different sources of information. Furthermore, planners might like to adjust their preferences progressively as they develop better understanding of the problem. For these reasons, an interactive approach is presented to determine the weight for the above bi-objective problem, consisting of the following two steps. A representative subset of non-dominated solutions is first generated by varying the weight, and then the decision maker can determine the weight that results in the best compromise solution based on the following three criteria, as commonly used in the multi-objective programming field.

1. Minimum combined deviation. This is equivalent to the objective function value in Equation (3.3).

2. Best trade-off. The trade-off measurement can be computed by $\frac{\partial Z_1}{\partial Z_2}$, where Z_1 and Z_2 are the first and second objectives. Given two different weights w^0 and w^1, the corresponding objective values are $Z(w^0) = (Z_1(w^0), Z_2(w^0))$ and $Z(w^1) = (Z_1(w^1), Z_2(w^1))$, then $\frac{\partial Z_1}{\partial Z_2}$ can be numerically approximated from

the ratio of change between the Z_1 and Z_2 as shown in Equation (3.7). Intuitively, a trade-off in the OD estimation problem means how much deviation from the target demand the decision maker would give up to decrease the deviation for link counts by one unit.

$$\frac{\partial Z_1}{\partial Z_2} = \frac{Z_1(w^0) - Z_1(w^1)}{Z_2(w^0) - Z_2(w^1)} \tag{3.7}$$

3. Minimum distance from the ideal point. Planners can define the goal f_1 and f_2 as the maximum possible deviation for the first and the second objectives, and then the goals for both objectives make up an ideal point $f^* = (f_1, f_2)$ or a utopia point. The best compromise solution is the one with minimum distance from the ideal point.

3.2.2 Utilizing Limited Real-Time Data

Given a subset of links with real-time link flow observations, a fundamental question is how to identify the demand dynamics using limited information, in particular, how to obtain a unique solution for the above ordinary least-square formulation. This requires that the number of decision variables (OD demand flows) be less than the number of constraints (the number of link observations plus the number of OD pairs in the static demand matrix), as shown in Inequality (3.8).

$$L \times H + I \times J \geq I \times J \times T \tag{3.8}$$

If the given link observations cannot satisfy Inequality (3.8), then the OD estimation turns out to be an under-determined problem, which can have numerous multiple solutions. To ensure the identifiability of the dynamic OD estimation problem, the following four possible approaches can be used.

The first simple remedy is to increase the length of departure time intervals so as to reduce the number of decision variables, but this aggregation scheme will undermine the capability of modeling OD demand dynamics. The second method is to shorten the length of observation time intervals to increase the number of observation. However, a short observation time interval would increase the possibility of linear correlation in the link flow matrix P, which makes the estimation result unstable. In fact, to obtain a unique solution, one still needs to verify the rank condition, that is, the sum of rank for matrices G and C is greater than the number of variables. Since the coefficient vectors in matrix G correspond to independent OD demand, the rank of matrix A is always $I \times J$. The link flow proportion vector can be expressed as Equation (3.9).

$$p_{(l,t),(i,j,\tau)} = \sum_{k \in K(i,j)} \alpha_{(l,t),(i,j,\tau)} q_{(k,\tau)} \tag{3.9}$$

where $\alpha_{(l,t),(i,j,\tau)}$ is the time-dependent link-path incidence indicator, $q_{(k,\tau)}$ is the path flow choice probability of selecting path k at the departing interval t, and $K(i,j)$ is the set of paths between origin i and destination j.

Clearly, the path flow choice probability is determined by traffic assignment, and the link-path incidence is governed by the traffic flow propagation process. For instance, consider two consecutive short intervals t_1 and t_2; it is highly possible that the path flow choice probability and the time-dependent link-path incidence are unchanged, and the corresponding two link flow proportion vectors at t_1 and t_2 are the same. This implies that one cannot arbitrarily shrink the observation interval to

increase the number of observations, and the redundant information does not increase the chance of making the problem identifiable.

For a traffic network with partial observations, not all OD demand will pass through those links that have flow measurements. In other words, only those OD demand flows that have impact on the measured flows can be inferred from the observed flows. Based on the link flow proportions generated from the network loading (simulation-assignment) result, one can denote OD pairs (i, j) with $p_{(l,t),(i,j,\tau),m} > 0$ as relevant OD pairs, and those with $p_{(l,t),(i,j,\tau),m} = 0$ as irrelevant OD pairs. Consequently, only relevant OD pairs need to enter the OD estimation problem. However, this procedure is still an ad hoc technique that highly relies on the quality of simulated link flow proportions, and it is still possible to rule out actual relevant OD pairs. Therefore, one should try to estimate the full OD matrix table as completely as possible in planning practice.

The fourth approach is to apply the polynomial transformation (Kang, 1999) shown in Equation (3.10). In particular, if the degree of the polynomial model N is less than the number of departure intervals T, the number of total decision variables can be reduced. An advantage of this method is to use fewer decision variables to represent the dynamics of demand, especially the trend information. However, it should be noted that a low-order polynomial model may not always capture the full randomness of demand, and a high order model might lead to wild oscillations even if it provides better goodness of fit.

$$d_{(i,j,\tau)} = \sum_{n=0}^{N} b_{(i,j)}^{n} \tau^{n} \tag{3.10}$$

where n is the order term, b is the parameter to be estimated.

3.3 Multi-day (Weekdays) OD Demand Estimation

3.3.1 Model Specification

The formulation of the OD demand estimation problem with single-day link observations is extended to a multi-day context. Considering five weekdays, a more extensive model can be expressed as Equation (3.11).

This formulation is analogous to a multiple linear regression model that has the standard form as Equation (3.12). From the multiple linear regression point of view, Y are dependent variables, (X_1 X_2 X_3 X_4 X_5) are independent variables, (D_1 D_2 D_3 D_4 D_5)^T are coefficients to be estimated and Ψ are error terms.

$$\begin{pmatrix} (1-w)C_1 \\ (1-w)C_2 \\ (1-w)C_3 \\ (1-w)C_4 \\ (1-w)C_5 \\ wG \\ wG \\ wG \\ wG \\ wG \end{pmatrix} = \begin{bmatrix} (1-w)P_1 & 0 & 0 & 0 & 0 \\ 0 & (1-w)P_2 & 0 & 0 & 0 \\ 0 & 0 & (1-w)P_3 & 0 & 0 \\ 0 & 0 & 0 & (1-w)P_4 & 0 \\ 0 & 0 & 0 & 0 & (1-w)P_5 \\ wA & 0 & 0 & 0 & 0 \\ 0 & wA & 0 & 0 & 0 \\ 0 & 0 & wA & 0 & 0 \\ 0 & 0 & 0 & wA & 0 \\ 0 & 0 & 0 & 0 & wA \end{bmatrix} \cdot \begin{pmatrix} D_1 \\ D_2 \\ D_3 \\ D_4 \\ D_5 \end{pmatrix} + \begin{pmatrix} (1-w)E_1 \\ (1-w)E_2 \\ (1-w)E_3 \\ (1-w)E_4 \\ (1-w)E_5 \\ w\Pi_1 \\ w\Pi_2 \\ w\Pi_3 \\ w\Pi_4 \\ w\Pi_5 \end{pmatrix} \quad (3.11)$$

$$Y = \begin{pmatrix} X_1 & X_2 & X_3 & X_4 & X_5 \end{pmatrix} \cdot \begin{pmatrix} D_1 \\ D_2 \\ D_3 \\ D_4 \\ D_5 \end{pmatrix} + \Psi \quad (3.12)$$

Let us consider three possible assumptions about D_m that lead to different forms for Equation (3.12).

The first is that the OD demand matrices D_m on different days are different. Accordingly, link flow proportions, i.e. X_m in Equation (3.12), would also have different values.

The second situation corresponds to identical D_m on different days. Thus, link flow proportions over different days would be generated identically from the DTA simulation, and Equation (3.12) can be collapsed to the simple form shown in Equation (3.13).

$$Y = \begin{pmatrix} X_0 & X_0 & X_0 & X_0 & X_0 \end{pmatrix} \cdot \begin{pmatrix} D_0 \\ D_0 \\ D_0 \\ D_0 \\ D_0 \end{pmatrix} + \Psi \tag{3.13}$$

However, this assumption D_m is likely to be too stringent for the real-world traffic demand. In order to recognize the inherent stochasticity of traffic demand, the third assumption views the multi-day OD demand as the outcome of a common underlying random process with mean D_0 and variance ε_D, that is, $D_m = D_0 + \varepsilon_D$. In this way, Equation (3.12) can be simplified to Equation (3.14).

$$Y = \begin{pmatrix} X_1 & X_2 & X_3 & X_4 & X_5 \end{pmatrix} \cdot \begin{pmatrix} D_0 \\ D_0 \\ D_0 \\ D_0 \\ D_0 \end{pmatrix} + \Psi' \tag{3.14}$$

where Ψ' is the combined error for Ψ and ε_D.

The above multi-day OD estimation problem can still be solved by the bi-level ordinary least-squares method used for single-day estimation described previously. Its mathematical formulation consists of objective function (3.15) and constraint (3.16). Note that, the objective is to minimize *multi-day* discrepancies, and the link flow proportions are obtained from the DTA simulator individually for different demand matrices.

obj.

$$\min Z = \sum_m \left\{ (1-w) \sum_{l,t} [\sum_{i,j,\tau} \hat{p}_{(l,t),(i,j,\tau),m} \cdot d_{(i,j,\tau),m} - c_{(l,t),m}]^2 + w \sum_{i,j} [\sum_{\tau} d_{(i,j,\tau),m} - g_{(i,j)}]^2 \right\} \quad (3.15)$$

s.t. $\quad d_{(i,j,\tau),m} \geq 0 \,, \forall i, j, \tau, m$

$$\hat{p}_{(l,t),(i,j,\tau),m} = \text{assignment} \quad (d_{(i,j,\tau),m}) \quad \text{from DTA, } \forall l, h, i, j, \tau, m \qquad (3.16)$$

3.3.2 Analysis of Day-to-Day Variability

Hypotheses for D_m

In order to identify day-to-day variability of OD demand, two potential models are assumed. The null hypothesis (H_0) for Dm is that the means of multi-day demand are identical, corresponding to a reduced model. The alternative hypothesis (H_1) for Dm is that the means of multi-day demand patterns are different, corresponding to a full model.

$H_0 : D_{m1} = D_{m2}, \quad \text{for } m_1 \neq m_2.$

$H_1 : D_{m1} \neq D_{m2}, \quad \text{for } m_1 \neq m_2.$

Standard F-test

Statistical testing is performed to compare means across multiple days. The F-statistic tests the null hypothesis that the multiple means of OD demands across all days are equal. If the computed F-statistic value is greater than the corresponding critical value (for the desired significance level), then the null hypothesis can be rejected, and the multi-day mean OD demands may be considered significantly different from one another.

$$F = \frac{\left(SSE_{reduced} - SSE_{full}\right)/\left(k-g\right)}{SSE_{full}/\left[n-\left(k+1\right)\right]} \tag{3.17}$$

where n is the number of observations; k is the number of restrictions causing change of the full model to the reduced model; g+1 is the number of coefficients in the reduced model.

SSE is the sum of square errors, which is calculated according to (3.18),

$$\begin{aligned} SSE &= \sum_m \left\{ (1-w)\sum_{l,t}[\sum_{i,j,\tau} \hat{p}_{(l,t),(i,j,\tau),m} \cdot \hat{d}_{(i,j,\tau),m} - c_{(l,t),m}]^2 + w\sum_{i,j}[\sum_{\tau} \hat{d}_{(i,j,\tau),m} - g_{(i,j)}]^2 \right\} \\ &= \sum_m SSE_m \end{aligned} \tag{3.18}$$

It is worth noting that the variables (X_1 X_2 X_3 X_4 X_5) in this study would have different values in the full model and the reduced model due to the inherent dependency of link flow proportions and OD traffic demand matrices.

Full model calibration

Based on the structure of the model (3.12), the procedure for full model calibration is equivalent to performing individual model calibration for each day.

After obtaining five estimated OD demand matrices individually, i.e. ($\hat{D}_1$, $\hat{D}_2$, $\hat{D}_3$, $\hat{D}_4$, $\hat{D}_5$), one just needs to substitute them in Equation (3.18) to compute SSE_{full}.

Reduced model calibration

The following procedure is adopted to calibrate the reduced model:

Step 1: Compute average the $\overline{D}$ and variance σ_D^2 of the estimated OD demand ($\hat{D}_1$, $\hat{D}_2$, $\hat{D}_3$, $\hat{D}_4$, $\hat{D}_5$) obtained from the full model calibration.

Step 2: Randomly generate OD demand on five days ($\dot{D}_1$, $\dot{D}_2$, $\dot{D}_3$, $\dot{D}_4$, $\dot{D}_5$) based on the average $\overline{D}$ and the variance σ_D^2.

Step 3: Obtain five link flow proportions matrices ($\hat{P}_1$, $\hat{P}_2$, $\hat{P}_3$, $\hat{P}_4$, $\hat{P}_5$) through DTA simulation, given the randomly generated ($\dot{D}_1$, $\dot{D}_2$, $\dot{D}_3$, $\dot{D}_4$, $\dot{D}_5$).

Step 4: Substitute ($\hat{P}_1$, $\hat{P}_2$, $\hat{P}_3$, $\hat{P}_4$, $\hat{P}_5$) into formulation (3.18) to estimate a common demand matrix D_0 for all days. Then, the estimated D_0 is taken as a new average value used for randomization in Step 2. If D_0 is stable, stop and the estimated D_0 in this step is the optimal one. Otherwise, go back to Step 2 and repeat the process.

Finally, $SSE_{reduced}$ can be computed based on the optimal OD demand D_0.

3.4 Numerical Experiments

3.4.1 Network Configuration and Traffic Measurements

In this section, the proposed off-line OD estimation models are tested using a simplified road network of the Irvine, CA test bed, which consists of two interstate freeways (I-5, I-405), a state highway 133, as well as other main arterials.

As shown in Figure 3-1, the simplified network includes 16 OD zones, 31 nodes and 80 directed links (32 freeway and 48 arterial). Traffic counts are measured on 16 links; at 30-second interval on 10 freeway links, and at 5-minute interval on 6 arterial links. In addition, a static planning OD demand table is given and used as the target demand. The time of interest in the following experiments is the morning peak period (6:30 am – 8:30 am) on four weekdays (Tuesday ~ Friday). It would have been ideal to investigate the OD demand variability over five weekdays, but the data for Monday could not be used in the estimation due to poor data quality (caused by sensor malfunction). Simulation is performed for three hours (6:00 am – 9:00 am). In order to gain more reliable estimation results, the starting period from 6:00 to 6:30 is used as a warm-up period. Moreover, the ending period from 8:30 to 9:00 is not considered in the statistics, since a large number of vehicles departing after 8:30 would not have finished their trips, resulting in incomplete link-flow proportions.

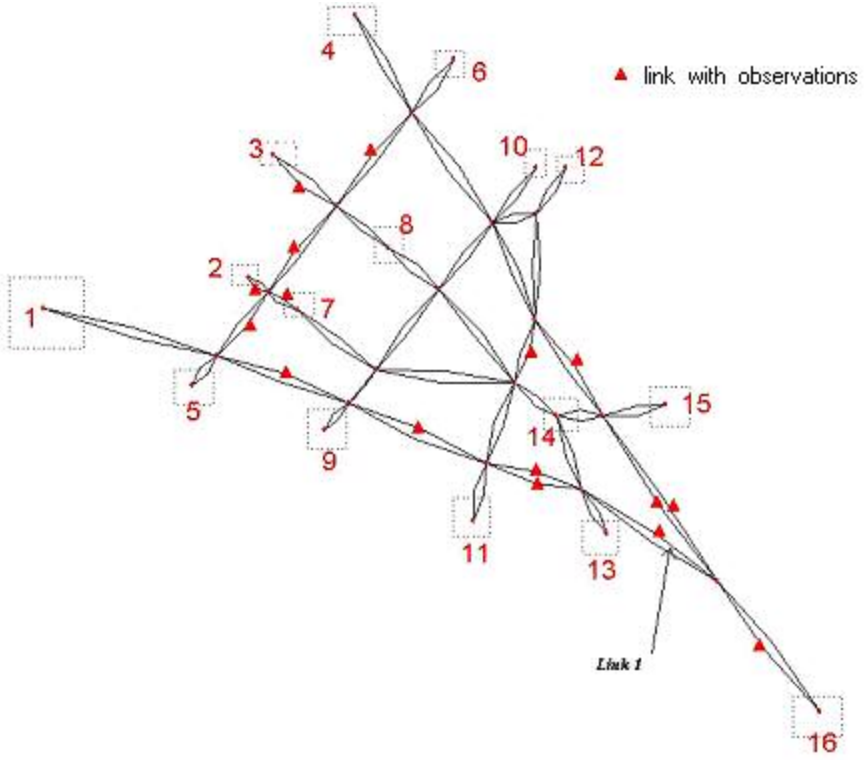

FIGURE 3-1 Irvine simplified network.

3.4.2 Weighting Scheme in the Upper Level Optimization

Before conducting the OD estimation, let us first want to determine the appropriate weighting value in the weighted objective function. Using the data for Tuesday, ten different values of *w* between 0 and 1 are used to generate a representative set of non-dominated solutions, as plotted in Figure 3-2.

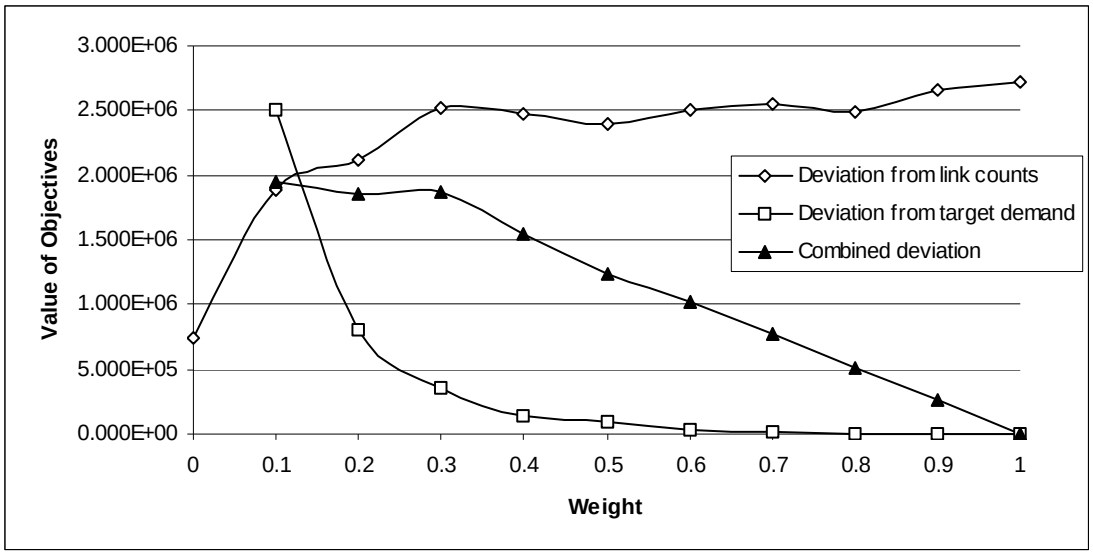

FIGURE 3-2 A representative set of non-dominated solutions.

For $w = 0$, i.e. OD estimation without target demand, the resulting deviation from the target demand is 2.42×10^8, which falls outside the range of the plot. As expected, greater weight on the target demand can result in smaller deviation from the target demand but larger deviation from the link flows. Interestingly, the rate of change for the second objective is much more dramatic than the first objective, and the total weighted deviation is governed by the deviation from the target demand. In addition, $w = 1$ minimizes the deviation from the target demand, actually yielding an overall deviation of zero, since the target demand can always be a feasible solution. In contrast, for $w = 0$ (i.e. only the deviations of link flows are considered), the solution does not fit all the link counts perfectly.

The three criteria discussed in Section 3.2.1 for determining a best compromise solution are examined here to determine the appropriate weight. First, the minimum combined deviation condition is not suitable in our case, since $w = 1$ always provides the best result but does not consider the link flows. It is easy to check

that w = 0.9 corresponds to the "best" trade-off, but the absolute deviation for the link counts is still very high. The ideal point is illustrated in Figure 3-3, where the goals for the first and second objectives are set to 2.00×10^6 and 0, respectively. The plot reveals that the solutions corresponding to weights of 0.2 and 0.5 are very close to the ideal point. This study will use w = 0.5 since this provides better trade-offs.

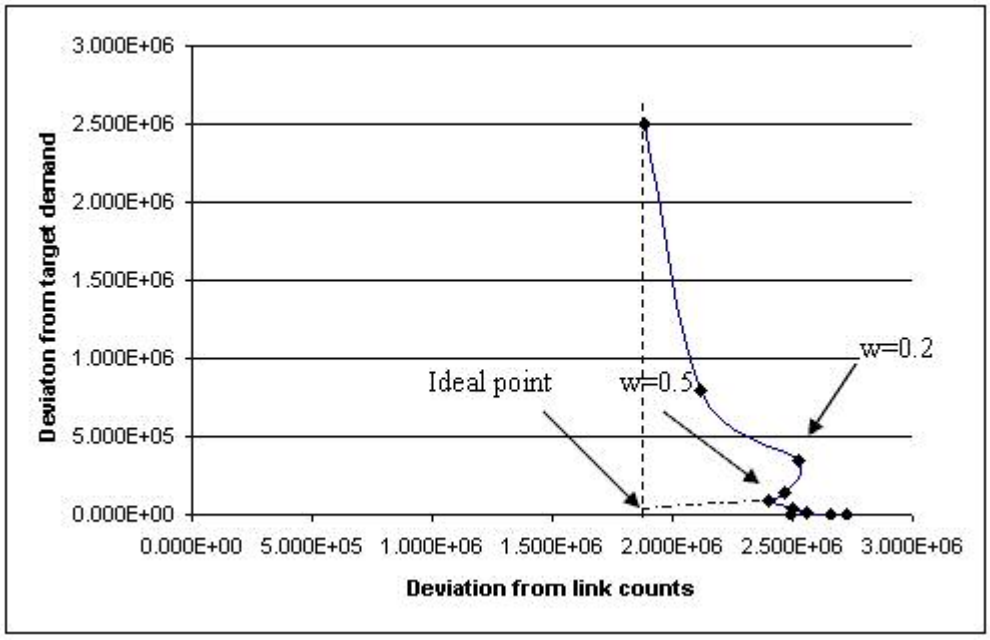

FIGURE 3-3 A representative set of non-dominated solutions.

3.4.3 Day-to-day OD Demand Patterns

The estimated demand patterns for three OD pairs are shown in Figures 3-4, 3-5 and 3-6. In particular, the selected OD pairs (12,1), (16, 1) and (16, 4) are representative OD pairs with the highest trip demand. In Figure 3-4, the estimated dynamic demand on different days is consistent with the magnitude of the historical static demand. As expected, all the three OD pairs show significant within-day dynamics. In particular, OD pair (12, 1) corresponds to a slow increasing trend, and

OD pairs (16, 1) and (16, 4) have similar peaking patterns: the peak occurs around 7:15 am, and the height of the peak is about 20~30% above the minimal demand level during the two hours. Note that the latter two OD pairs start from the same origin zone 16, so they are more likely to have common departure time patterns. Moreover, the dynamic demand pattern can be verified by the observed link flows as shown in Figure 3-7. The reason of selecting link 1 (in Figure 3-1) is that it carries the demand flow for OD pair (16, 1). It is easy to see the time of flow peak is relatively later than the time of demand peak, which is around 7:15am to 7:45am, due to the traffic flow propagation. As can be seen from the plot, the simulated flows on link 1 match the observed flows quite well.

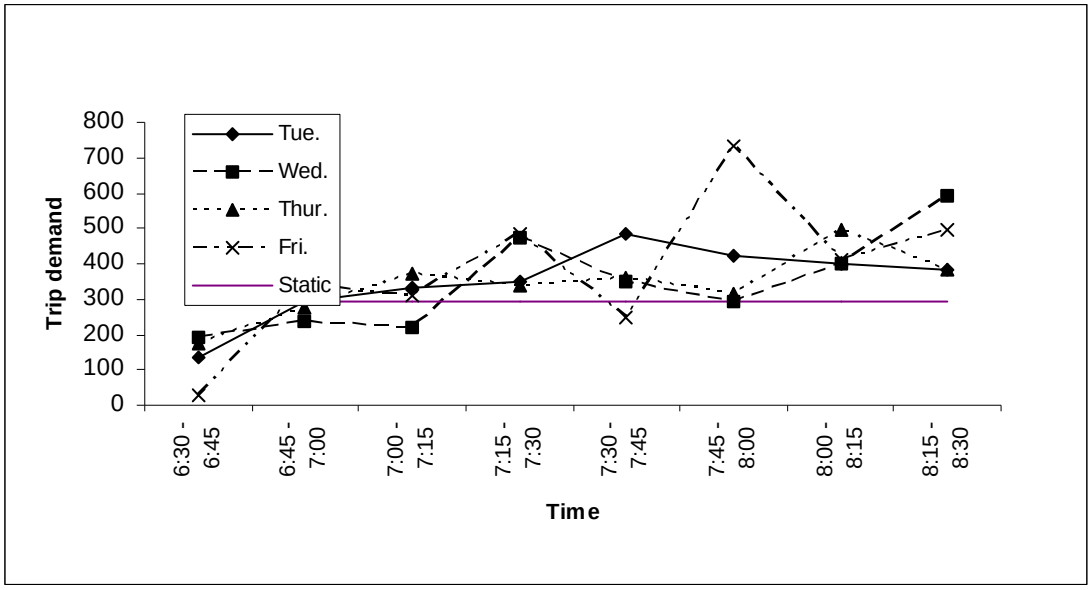

FIGURE 3-4 Estimated trip demand patterns for OD pair (12, 1).

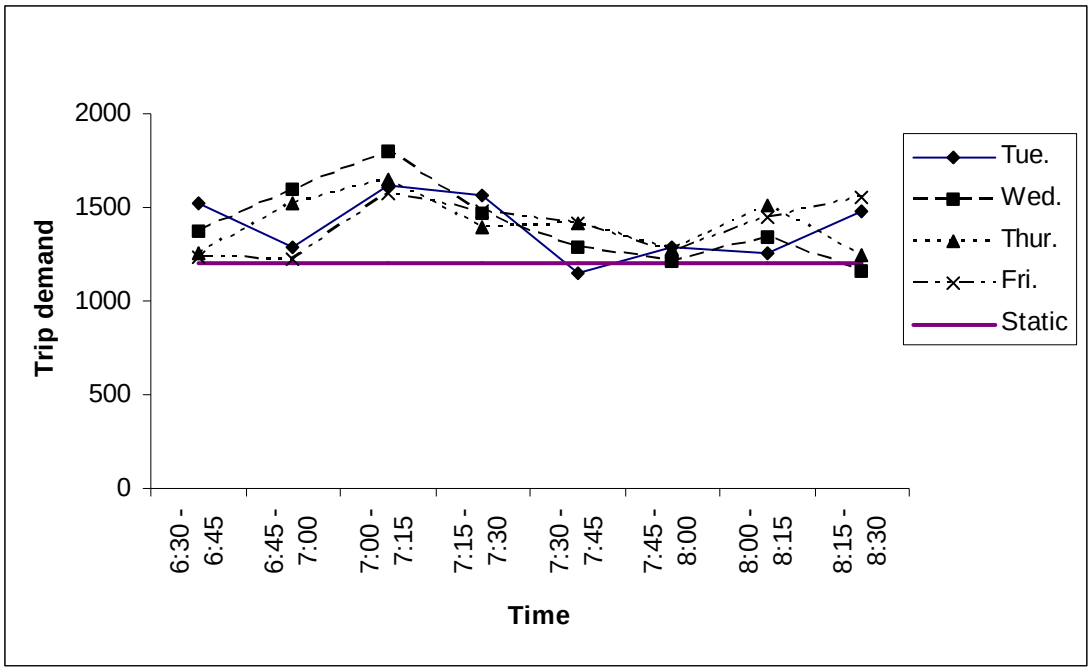

FIGURE 3-5 Estimated trip demand patterns for OD pair (16, 1).

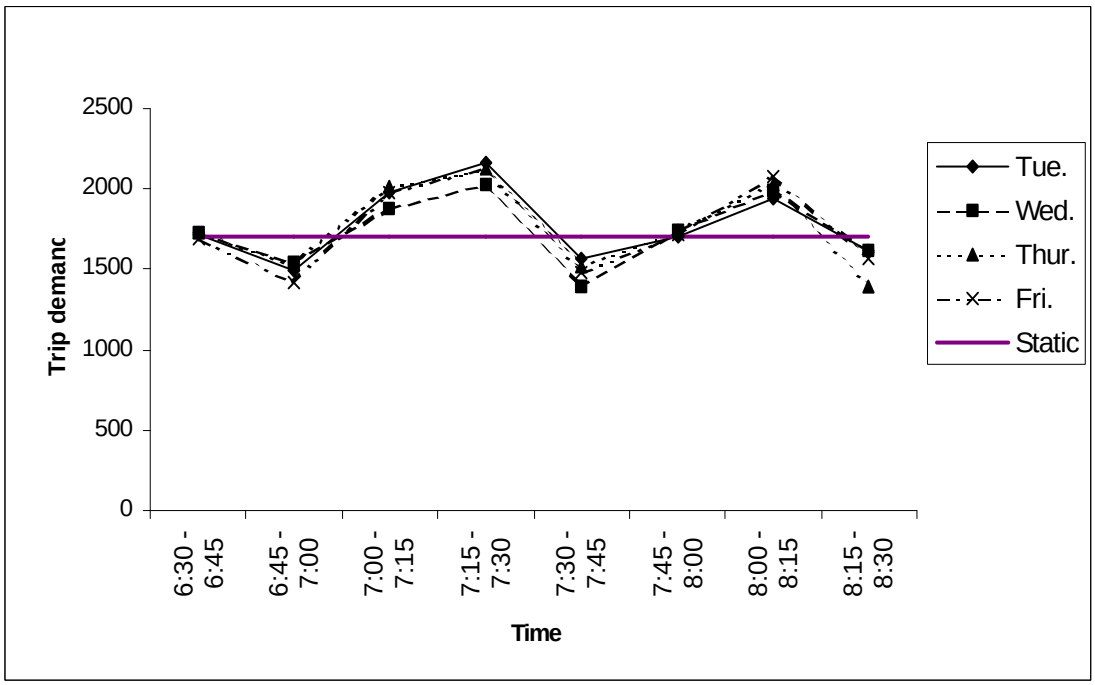

FIGURE 3-6 Estimated trip demand patterns for OD pair (16, 4).

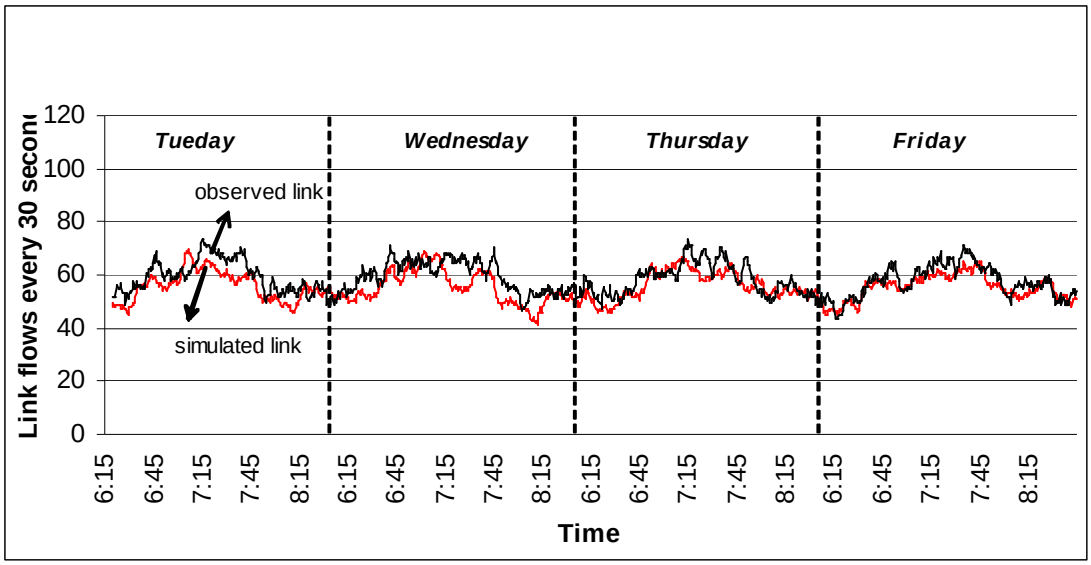

FIGURE 3-7 Observed vs. simulated link flows for Link 1.

Based on the estimated results, all three OD pairs exhibit different patterns of day-to-day variation. In general, OD pair (12, 1) exhibits greater variation in terms of the time of peak, OD pair (16, 1) shows greater variation in terms of the height of peak, while OD pair (16, 4) exhibits much more stable patterns. This information is useful to help the transportation manager alleviate congestion by redistributing demand flow spatially and temporally on the network.

3.4.4 Hypothesis Testing for the Mean of the Demand

The sum of squared errors SSE for estimation (that is, the total objective value) is required for computation of the F-statistic. Calibration of the full model and the reduced model provide the value of SSE without additional effort. The results are $SSE_{full} = 4.90 \times 10^6$ for the full model and $SSE_{reduced} = 5.55 \times 10^6$ for the reduced model. In addition, during 2 hours (or 120 minutes) and for 16×16–16 = 240 OD pairs, the number of observations n = (120×2×10+120/5×6+240)×4 = 11136. The

number of restrictions from the full model to the reduced model is k = 120/15×240×3 = 5760, and the number of coefficients in the reduced model is $g+1$ = 120/15×240 = 1920. The F-statistic is

$$F = \frac{\left(SSE_{reduced} - SSE_{full}\right)/\left(k-g\right)}{SSE_{full}/\left[n-\left(k+1\right)\right]} = 0.186.$$

By selecting the significance level α as 0.05, the critical value $F_{\alpha,k-g,n-(k+1)}$ =1.10, so the null hypothesis cannot be rejected, indicating that the mean OD demand pattern for the network is essentially identical across multiple days.

Since link-flow proportions are generated by simulation instead of known a priori, the F-test statistics for the model specification in this example should be viewed as an approximate indicator. Moreover, demand flows associated with different OD pairs might not be independent of each other, so the conclusion from the statistical testing should again be interpreted with considerable caution.

3.5 Summary

Time-dependent OD demand matrices are a critical input to dynamic traffic assignment methodology in real-time operational and planning applications. This chapter introduces and highlights the potential of using multiple sources of information to estimate the dynamic OD demand for planning applications. The particular sources available here include the historical static information and ITS real-time link-level information. First, a bi-level iterative dynamic OD estimation model is extended to combine both (1) deviation between estimated link flows and real-time link counts and (2) deviation between estimated time-dependent demand and given

historical static demand. The two objectives are combined using a weighted function, where the weighting value is determined by an interactive approach to obtain the best compromise solution. In particular, the trade-offs among several methods that are designed to use limited real-time information to infer the demand dynamics are discussed. The model was extended to use the multiple days of link counts to estimate the variations in traffic demand over multiple days. Indicated by the estimation results based on the Irvine network, the ideal point approach is suitable for determining the weights in the multi-objective framework for estimating OD demand with various information sources, and major OD pairs exhibit different patterns of day-to-day variation.

4. Off-line Dynamic OD Demand Estimation Using Automatic Vehicle Identification Data

4.1 Introduction

Previous chapter addressed the dynamic OD demand estimation problem using multi-day link counts and historical static OD demand data, and formulated a statistical testing procedure to investigate day-to-day demand variations. Together with commonly used point sensors, AVI systems provide valuable point-to-point traffic data for identifying the complex OD demand system state. To estimate population OD demand, this study is intended to use extract OD demand distribution information from partially observed AVI counts without estimating market penetration rates. Specifically, population OD split fractions are sampled from point-to-point AVI counts, as opposed to being treated as unknown variables in the early dynamic OD estimation models (Cremer and Keller, 1981, 1984, 1987). Section 4.2 first illustrates the use of point-to-point split fractions in dynamic population OD demand estimation with AVI data, and addresses the resulting modeling and formulation issues. Based on two idealized assumptions, a nonlinear ordinary least-squares model is presented in Section 4.3 to extract OD distribution information from AVI counts. Section 4.4 proposes two OD demand estimation formulations in the presence of possible identification and representativeness errors. Sections 4.5 and 4.6 extend the iterative solution algorithm from Chapter 3 and discuss the identification condition and properties of the resulting estimation problem. Section 4.8 uses synthetic AVI traffic counts to systematically investigate the relative value of AVI

information under different market penetration rates, identification rates and detector location schemes.

4.2 Problem Statement and Illustrative Examples

The following notation is used to represent all the variables in the dynamic OD demand estimation formulation.

L = set of links in the network.

L_{lc} = set of links with link count observations.

L_{vi} = set of links with vehicle identification observations.

l = subscript for link with traffic measurements.

i = subscript for origin zone, $i \in I$.

j = subscript for destination zone, $j \in J$.

τ= subscript for departure time intervals, $\tau = 1, 2, \ldots, T_d$.

t = subscript for observation time interval, i.e. sampling time interval, t=1, 2, …, T_o.

tg = superscript for tag-equipped vehicles.

id = superscript for identified vehicles.

k = superscript for iteration counter.

$c_{(l,t)}$ = number of vehicles on link l during observation interval t .

C = vector of measured flows on the links, consisting of element $c_{(l,t)}$.

$c_{(l,s,t)}$ = number of vehicles observed on link *s*, traveling from link *l* during observation interval *t*.

$d_{(i,j,\tau)}$ = demand volume with destination in zone j, originating their trip at zone i during departure interval τ.

D = dynamic OD demand matrix, consisting of elements $d_{(i,j,\tau)}$.

$c_{(i,j,\tau)}$= number of vehicles observed in destination zone *j*, originating their trip at zone i during departure interval τ.

$b_{(i,j,\tau)}$ = origin-to-destination split fraction, i.e. proportion of traffic departing from origin *i* during departure time interval *τ*, heading towards destination *j*.

$b_{(l,s,t)}$ = link-to-link split fraction, i.e. proportion of traffic passing link *l* during observation time interval t, heading towards link *s* (link *s* is not necessary to be a downstream link of link *l*) .

$p_{(l,t),(i,j,\tau)}$ = link flow proportions, i.e. proportion of vehicular demand flows from origin *i* to destination j, starting their trips during departure interval τ, contributing to the flow on link *l* during observation interval *t*.

$\hat{p}_{(l,t),(i,j,\tau)}$ = estimated link flow proportions based on a DTA program.

$p_{(l,s,t),(i,j,\tau)}$ = link-to-link flow proportions, i.e. proportion of vehicular flows from origin *i* to destination j, starting their trips during departure interval τ, contributing to the link-to-link flow from link *l* (during observation intervals *t*) to link *s*.

$\hat{p}_{(l,s,t),(i,j,\tau)}$ = estimated point-to-point-flow proportions based on a DTA program.

$\hat{P}$ =estimated flow proportion matrix that includes elements $\hat{p}_{(l,t),(i,j,\tau)}$ and $\hat{p}_{(l,s,t),(i,j,\tau)}$.

$\eta_{(l,s,t)}$ = sampling error term in estimation of link-to-link split fraction $b_{(l,s,t)}$.

$\zeta_{(l,s,t)}$= combined error term in estimation of link-to-link split fraction $b_{(l,s,t)}$.

$\varepsilon_{(l,t)}$ = combined error term in estimation of traffic flow on link l during observation interval t.

$g_{(i,j)}$ = target demand, which is total traffic demand during period of interest for origin-destination pair (i, j) .

G = historical OD demand matrix, consisting of elements $g_{(i,j)}$.

Consider a traffic network consisting of multiple origins $i \in I$ and destinations $j \in J$, as well as a set of nodes connected by a set of directed links. The analysis period of interest, is discretized into departure time intervals $\tau = 1, 2, \ldots, T_d$. Link counts $c_{(l,t)}$ are available on link $l \in L_{lc}$ during observation interval $t = 1, 2, \ldots, T_o$. AVI reader stations are located on link $l \in I_{vi}$, and vehicle identification data include point counts $c^{id}_{(l,t)} \; \forall \; l \in L_{\text{vi}}, t = 1, \ldots, T_o$ and point-to-point counts $c^{id}_{(l,s,t)} \; \forall \; l,s \in L_{\text{vi}}, t = 1, \ldots, T_o$. The sampling time intervals for AVI counts and traffic link counts are assumed to be the same for notation simplicity. As AVI reader stations are installed on link segments in a network, the "point-to-point counts" will be equivalently referred to as "link-to-link counts" in order to maintain continuity with "link counts" from point sensors. Given link counts, vehicle identification counts and prior information on OD trips, the dynamic OD demand estimation problem seeks to find time-dependent OD trip

desires so as to minimize deviations between the observed traffic flows and the traffic flows resulting from a DTA program, and deviations between estimated OD demand flows and the historical demand matrix.

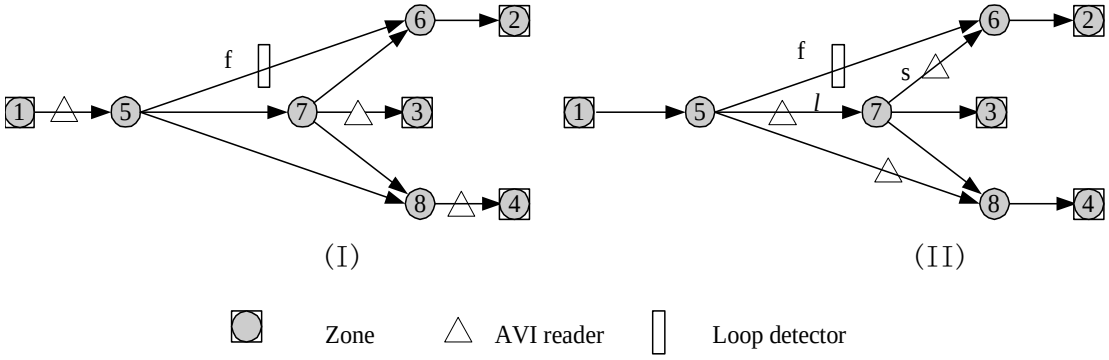

FIGURE 4-1 Example of utilizing AVI point-to-point counts

In order to circumvent difficulties in estimating market penetration rates of AVI tags, this study utilizes probe vehicle data to extract spatial distribution information of trip-makers in traffic networks. This approach can be illustrated using a small network shown in Figure 4-1. The problem is to estimate population OD demand flows $d_{(1,2)}$, $d_{(1,3)}$ and $d_{(1,4)}$ from available AVI and loop detector counts. For simplicity, subscripts τ and t are dropped in the following discussion. In part (I) of Figure 4-1, an AVI reader located on link (1,5) records tagged vehicle flows departing from zone 1, that is, $\sum_j d^{tg}_{(1,j)} = d^{tg}_{(1,2)} + d^{tg}_{(1,3)} + d^{tg}_{(1,4)}$. Origin-destination AVI counts $d^{tg}_{(1,3)}$ and $d^{tg}_{(1,4)}$ are also observed. If tagged vehicles are representative of the total population, then AVI counts can be used to estimate the OD split fractions for the population demand traveling from zone 1, leading to the following measurement equations:

$$\frac{d^{tg}_{(1,3)}}{\sum_j d^{tg}_{(1,j)}} = \frac{d_{(1,3)}}{\sum_j d_{(1,j)}} + \eta_{(1,3)} \quad (4.1)$$

$$\frac{d^{tg}_{(1,4)}}{\sum_j d^{tg}_{(1,j)}} = \frac{d_{(1,4)}}{\sum_j d_{(1,j)}} + \eta_{(1,4)} \quad (4.2)$$

where $\eta_{(1,3)}$, $\eta_{(1,4)}$ are sampling errors.

The loop detector on link (5,6) captures partial OD flow $d_{(1,2)}$, so demand $d_{(1,2)}$ can be related to link count $c_{(f)}$ using link flow proportion $p_{(f),(1,2)}$, where f denotes link (5,6).

$$c_{(f)} = p_{(f),(1,2)} d_{(1,2)} + \varepsilon_{(f)} \quad (4.3)$$

Combining the above three measurement equations, one can create a system of nonlinear equations to estimate unknown population OD demand flows $d_{(1,2)}$, $d_{(1,3)}$ and $d_{(1,4)}$, without knowing the market penetration rate of vehicle tags.

Part (II) of Figure 4-1 shows a more general case, where direct origin destination tagged vehicle counts are unavailable. Denote links (5,7) and (7,6) as links l and s, respectively. Link-to-link counts $c_{(l,s)}$ include partial OD flows $d_{(1,2)}$, $d_{(1,3)}$, $d_{(1,4)}$, while link counts $c_{(l)}$ record partial OD flows $d_{(1,2)}$. The resulting split fraction for link pair (l,s) is

$$\frac{c^{id}_{(l,s)}}{c^{id}_{(l)}} = \frac{\sum_{i,j} p_{(l,s)(i,j)} d_{(i,j)}}{\sum_{i,j} p_{(l)(i,j)} d_{(i,j)}} + \eta_{(l,s)} = \frac{p_{(l,s)(1,2)} d_{(1,2)}}{p_{(l)(1,2)} d_{(1,2)} + p_{(l)(1,3)} d_{(1,3)} + p_{(l)(1,4)} d_{(1,4)}} + \eta_{(l,s)} . \quad (4.4)$$

In this case, the information on OD demand distributions can be partially revealed from link-to-link split fractions, i.e. the ratio of the link-to-link AVI counts and link AVI counts. Because link proportions are determined by the route choice behavior

and traffic flow propagation, the above link-to-link split formulation introduces additional complexity and uncertainty in demand estimation, compared to the formulation using OD split factions. A simulation-based DTA program, namely DYNASMART-P, be used to estimate the link flow and link-to-link flow proportions.

4.3 Nonlinear Least Squares Formulation

This section formulates a dynamic OD estimation model based on the following two assumptions.

1. AVI readers can correctly identify every tagged vehicle, i.e. 100% identification rates.
2. Tagged vehicles are a representative set of the entire population.

Under the first condition, $c^{id}_{(l,s,t)} = c^{tg}_{(l,s,t)}$; $c^{id}_{(l,,t)} = c^{tg}_{(l,t)}$. Under the second condition, the tagged vehicles probabilistically represent the entire population, and the split fractions of tagged vehicles can be used as sample estimates for the population split fractions. To construct a rigorous statistical inference model, the following discussion examines the properties of a random AVI sample for estimating split fractions. Recognizing that there are $c^{id}_{(l,s,t)}$ tagged vehicles choosing link s out of $c^{id}_{(l,t)}$ vehicles observed on link l at time t, link-to-link identified vehicle count $c^{id}_{(l,s,t)}$ essentially follows a binomial distribution with sample size $c^{id}_{(l,t)}$ and success probability $\frac{c_{(l,s,t)}}{c_{(l,t)}}$, i.e. $c^{id}_{(l,s,t)} \sim$ Binomial $[c^{id}_{(l,t)}, \frac{c_{(l,s,t)}}{c_{(l,t)}}]$. Since

$$\frac{c_{(l,s,t)}}{c_{(l,t)}} = \frac{\sum_{i,j,\tau} p_{(l,s,t)(i,j,\tau)} d_{(i,j,\tau)}}{\sum_{i,j,\tau} p_{(l,t)(i,j,\tau)} d_{(i,j,\tau)}}, \tag{4.5}$$

the sample proportion $\frac{c^{id}_{(l,s,t)}}{c^{id}_{(l,t)}}$ is an unbiased estimator of slit fractions for the population proportion. That is,

$$\frac{c^{id}_{(l,s,t)}}{c^{id}_{(l,t)}} = \frac{c_{(l,s,t)}}{c_{(l,t)}} + \eta_{(l,s,t)} = \frac{\sum_{i,j,\tau} p_{(l,s,t)(i,j,\tau)} d_{(i,j,\tau)}}{\sum_{i,j,\tau} p_{(l,t)(i,j,\tau)} d_{(i,j,\tau)}} + \eta_{(l,s,t)}, \quad (4.6)$$

and the corresponding mean and variance of sampling error $\eta_{(l,s,t)}$ are $E(\eta_{(l,s,t)}) = 0$ and

$$\mathrm{Var}(\eta_{(l,s,t)}) = \frac{b_{(l,s,t)} \cdot (1 - b_{(l,s,t)})}{c^{id}_{(l,t)}}, \quad (4.7)$$

where $b_{(l,s,t)} = \frac{c_{(l,s,t)}}{c_{(l,t)}}$. Equation (4.7) indicates that the variance of sampling error $\eta_{(l,s,t)}$ decreases as the size of the sample $c^{id}_{(l,t)}$ increases.

Let links *s* and *s′* denote two distinct links reachable from link *l*. If links *s* and *s′* are two independent choice alternatives for vehicles traveling on link *l* at time *t*, link-to-link AVI counts $c^{id}_{(l,s,t)}$ and $c^{id}_{(l,s',t)}$ follow a multinomial distribution, leading to the covariance of sampling errors as

$$Cov(\eta_{(l,s,t)}, \eta_{(l,s',t)}) = \frac{b_{(l,s,t)} \cdot b_{(l,s',t)}}{c^{id}_{(l,t)}}. \quad (4.8)$$

If certain vehicles in link-to-link flows $c_{(l,s,t)}$ and also appear in flows $c_{(l,s',t)}$, links *s* and s′ cannot be viewed as independent choice alternatives for vehicles traveling on link *l* at time *t*. In this case, link counts $c_{(l,s,t)}$ and $c_{(l,s',t)}$ can be partitioned into three mutually exclusive categories $\omega \in \Omega$:

$\omega = 1$: Only choose *s*; $\omega = 2$: only choose *s′*; $\omega = 3$: choose both *s* and *s′*.

The resulting covariance can be expressed in terms of variance and covariance between disjoint sets.

$$Cov(\eta_{(l,s,t)},\eta_{(l,s',t)}) = \sum_{\omega\in\Omega}\sum_{\omega'\in\Omega} Cov(\eta_{(l,\omega,t)},\eta_{(l,\omega',t)}) = \frac{1}{c^{id}{}_{(l,t)}} \sum_{\omega\in\Omega}\sum_{\omega'\in\Omega} b_{(l,\omega,t)} \cdot b_{(l,\omega',t)} \quad (4.9)$$

where $b_{(l,\omega,t)}$ is the proportion of link flow from link l at time t heading towards category $\omega\in\Omega$.

In the following two cases, the complicated nonlinear equation (4.6) can be simplified. First, if link volume $c_{(l,t)}$ is observed from a loop detector for link $l \in I_{vi}$, a linear measurement equation can be derived as:

$$\frac{c^{id}{}_{(l,s,t)}}{c^{id}{}_{(l,t)}} = \frac{c_{(l,s,t)}}{c_{(l,t)}} + \eta_{(l,s,t)} = \frac{\sum_{i,j,\tau} p_{(l,s,t)(i,j,\tau)} d_{(i,j,\tau)}}{c_{(l,t)}} + \eta_{(l,s,t)}. \quad (4.10)$$

Second, if AVI readers cover the entry links of origin i and the exit links of destination j, then AVI origin-destination counts are directly used to infer destination distributions:

$$\frac{c^{id}{}_{(i,j,\tau)}}{c^{id}{}_{(i,\tau)}} = \frac{d_{(i,j,t)}}{\sum_{j} d_{(i,j,t)}} + \eta_{(i,j,\tau)}. \quad (4.11)$$

Substituting the estimates of flow proportion matrices from a DTA problem into Equation (6) yields a complete measurement equation for link-to-link split fractions

$$\frac{c^{id}{}_{(l,s,t)}}{c^{id}{}_{(l,t)}} = \frac{\sum_{i,j,\tau} \hat{p}_{(l,s,t),(i,j,\tau)} d_{(i,j,\tau)}}{\sum_{i,j,\tau} \hat{p}_{(l,t)(i,j,\tau)} d_{(i,j,\tau)}} + \zeta_{(l,s,t)}, \quad (4.12)$$

where $\zeta_{(l,s,t)}$ refers to the combined error in estimation of link-to-link split fraction $b_{(l,s,t)}$. The combined error term $\zeta_{(l,s,t)}$ includes the following error sources.

1. Model assumption errors related to perfect representativeness and 100% identification rates.
2. Sensor errors (i.e. identification errors) related to link-to-link AVI count $c^{id}_{(l,s,t)}$ and link count $c^{id}_{(l,t)}$.
3. Sampling errors $\eta_{(l,s,t)}$.
4. Aggregation errors related to time-varying OD demand flows.
5. Estimation errors related to link flow and link-to-link flow proportions from the DTA program, which can be further caused by inconsistency in dynamic traffic assignment assumptions on the route choice behavior, traffic flow propagation, as well as input data errors related to traffic control and information strategies.

Because the split fractions only carry information on OD demand distributions, it is necessary to combine other information sources that describe OD population demand volumes in order to estimate a complete OD matrix. The observed traffic volume on link l during time interval t can be related to the OD demand flows using the link flow proportions, leading to the following measurement equation

$$c_{(l,t)} = \sum_{i,j,\tau} \hat{p}_{(l,t)(i,j,\tau)} \cdot d_{(i,j,\tau)} + \varepsilon_{(l,t)} . \quad (4.13)$$

If a static OD demand matrix is available from existing survey data or other planning applications, the difference between the static demand and the sum of dynamic demand over the study period can be expressed as

$$g_{(i,j)} = \sum_{\tau} d_{(i,j,\tau)} + \xi_{(i,j)} . \quad (4.14)$$

In addition, OD demand flows should satisfy non-negativity constraints

$$d_{(i,j,\tau)} \geq 0 \qquad \forall i, j, \tau \,. \tag{4.15}$$

If the split fraction is small and the number of observations in the AVI point sample is large, the binomial probabilities can be approximated by a Poisson distribution. If the AVI point sample size is sufficiently large with a moderate value of the split fraction, error terms $\zeta_{(l,s,t)}$ can be assumed to follow a normal distribution with zero mean according to the central limit theorem. If assuming the error terms $\zeta_{(l,s,t)}$, $\varepsilon_{(l,t)}$ and $\xi_{(i,j)}$ are independent, one can construct an Ordinary Least Squares (OLS) estimator to fuse data from different information sources. The corresponding bi-level dynamic OD estimation problem is to minimize the combined deviations with respect to link counts, historical static demand and AVI split fractions, subject to the dynamic traffic assignment constraint and non-negativity constraints for demand variables.

$$\min Z = [Z_1(D,C) + Z_2(D,G) + Z_3(D,C^{id})] \tag{4.16}$$

where

$$Z_1(D,C) = w_1 \sum_{l \in L_{lc}, t} [\sum_{i,j,\tau} \hat{p}_{(l,t),(i,j,\tau)} \cdot d_{(i,j,\tau)} - c_{(l,t)}]^2 \tag{4.17}$$

$$Z_2(D,G) = w_2 \sum_{i,j} [\sum_{\tau} d_{(i,j,\tau)} - g_{(i,j)}]^2 \tag{4.18}$$

$$Z_3(D,C^{id}) = w_3 \sum_{l \in L_{vi}, t} \sum_{s \in L_{vi}} [\frac{\sum_{i,j,\tau} \hat{p}_{(l,s,t),(i,j,\tau)} d_{(i,j,\tau)}}{\sum_{i,j,\tau} \hat{p}_{(l,t)(i,j,\tau)} d_{(i,j,\tau)}} - \frac{c^{id}{}_{(l,s,t)}}{c^{id}{}_{(l,t)}}]^2 \tag{4.19}$$

s.t. $\hat{P} = assignment\,[D]$ from DTA, (4.20)

$$d_{(i,j,\tau)} \geq 0 \;\forall\; i,j,\; \tau. \tag{4.21}$$

where w_1, w_2 and w_3 are positive weights for deviations with respect to link counts, historical static demand and observed split fractions.

With the full information on the covariance and autoregression terms, the objective function can be formulated as a Generalized Least Squares (GLS) formulation to obtain an efficient estimator.

4.4 Considering Representativeness Bias and Identification Errors

The two idealized assumptions in the above analysis, 100% identification rates and perfect representativeness of AVI samples, are difficult to meet in the real world. First, recognition rates vary significantly among different AVI technologies. For instance, Active tags, especially used for toll collection purposes, can provide a satisfactory >99% identification rate, but the identification rates for license plates and passive tags are relatively lower and have large variability. On the other hand, the AVI sample data might not a perfectly representative miniature of the population. For instance, tag users might experience less congestion on dedicated lanes in toll plaza, but they might need to pay one-time charge or monthly fee for transponder tags. In this case, tag users and non-tag users might belong to different socio-economic groups with heterogeneous preferences on the value of time. Essentially, the representativeness of AVI data should be verified on a case-by-case basis.

When these two assumptions are not attainable, there is a great need to establish a flexible estimation framework that can accommodate possible departure from the idealized conditions. A natural approach is to establish a joint estimation model as follows:

$$\frac{c^{id}{}_{(l,s,t)}}{c^{id}{}_{(l,t)}} = \frac{\sum_{i,j,\tau} p_{(l,s,t)(i,j,\tau)} d_{(i,j,\tau)}}{\sum_{i,j,\tau} p_{(l,t)(i,j,\tau)} d_{(i,j,\tau)}} \alpha_{(l,s)} + \zeta_{(l,s,t)}, \tag{4.22}$$

where a fixed effect parameter $\alpha_{(l,s)}$ is introduced to take into account possible systematic errors in estimating link-to-link split fractions.

The validity of the two idealized assumptions can be measured using the following statistical procedure. A full estimation model incorporates a fixed effect parameter for each possible link pair (*l*,*s*) explicitly, leading to a new deviation function between the unknown OD demand and the identified vehicle counts:

$$Z_3(D, C^{id}) = w_3 \sum_{l \in L_{vi}, t} \sum_{s \in L_{vi}} [\frac{\sum_{i,j,\tau} \hat{p}_{(l,s,t),(i,j,\tau)} d_{(i,j,\tau)}}{\sum_{i,j,\tau} \hat{p}_{(l,t)(i,j,\tau)} d_{(i,j,\tau)}} \alpha_{(l,s)} - \frac{c^{id}{}_{(l,s,t)}}{c^{id}{}_{(l,t)}}]^2 . \tag{4.23}$$

The null hypothesis (H_0) states that the mean of deviations is zero, corresponding to a reduced model with fixed effect parameters as 1. The alternative hypothesis (H_1) states that the mean of deviations is non-zero, corresponding to the full model. A standard F-statistic test can be applied in this context.

$$\boldsymbol{H_0}: \alpha_{(l,s)} = 1 \tag{4.24}$$

$$\boldsymbol{H_1}: \alpha_{(l,s)} \neq 1 \tag{4.25}$$

Essentially, introducing fixed effect parameters complicates the model structures. The existence of numerous error sources, such as assignment modeling errors and temporal fluctuation of demand flows, might lead to inconclusive estimates for $\alpha_{(l,s)}$ with large variance. In addition, if only average AVI counts for the entire planning horizon are available (e.g. in the context of static OD estimation), then fixed

effect parameters, which are designed to deal with time series data, are inappropriate to be included in the estimation model.

It is desirable to design a population OD demand estimator without involving fixed effect parameters. In many instances, the traffic planner knows the likely sign of fixed effect parameters but with little information about the exact magnitude of representativeness and identification errors. For instance, AVI samples based on license plates have perfect representativeness but low identification rates, meaning that $\alpha_{(l,s)} \leq 1$ for sure. In an AVI system where passive tags offered to the public at no charge, both representative errors and identification errors can coexist but the impact of identification errors is more likely to be dominating. In these two cases, the mean deviation in the original measurement equation (4.12) becomes

$$E\left(\frac{c^{id}_{(l,s,t)}}{c^{id}_{(l,t)}} - \frac{\sum_{i,j,\tau} p_{(l,s,t)(i,j,\tau)} d_{(i,j,\tau)}}{\sum_{i,j,\tau} p_{(l,t)(i,j,\tau)} d_{(i,j,\tau)}}\right) < E\left(\frac{c^{id}_{(l,s,t)}}{c^{id}_{(l,t)}} - \frac{\sum_{i,j,\tau} p_{(l,s,t)(i,j,\tau)} d_{(i,j,\tau)}}{\sum_{i,j,\tau} p_{(l,t)(i,j,\tau)} d_{(i,j,\tau)}} \alpha_{(l,s)}\right) = 0, \quad (4.26)$$

so the residuals are more likely to be negative, as illustrated in Figure 4-2. In this study, two-sided penalty terms are introduced to reflect the asymmetric distribution of residuals.

$$Z_3(D, C^{id}) = w^+_3 \sum_{l \in L_{vi}, t} \sum_{s \in L_{vi}} \left[\frac{\sum_{i,j,\tau} \hat{p}_{(l,s,t)(i,j,\tau)} d_{(i,j,\tau)}}{\sum_{i,j,\tau} \hat{p}_{(l,t)(i,j,\tau)} d_{(i,j,\tau)}} - \frac{c^{id}_{(l,s,t)}}{c^{id}_{(l,t)}} \right]^+ + w^-_3 \sum_{l \in L_{vi}, t} \sum_{s \in L_{vi}} \left[\frac{\sum_{i,j,\tau} \hat{p}_{(l,s,t)(i,j,\tau)} d_{(i,j,\tau)}}{\sum_{i,j,\tau} \hat{p}_{(l,t)(i,j,\tau)} d_{(i,j,\tau)}} - \frac{c^{id}_{(l,s,t)}}{c^{id}_{(l,t)}} \right]^- \quad (4.27)$$

where w^+_3 and w^-_3 are penalty terms for positive and negative deviations, respectively.

The above objective function can be viewed as an adaptation of the goal programming approach, which is used to assign priority factors for over- or under-

achievement of a goal. The target value (goal) in this study is given by the observed AVI point-to-point split fraction, and the overall aim is to simultaneously minimize the positive and negative deviations (i.e. overestimation and underestimation) from the specified goals. An interactive approach can be used to determine the value of two-sided weights, which can be in turn interpreted as the relative confidence and preference of a decision maker on the possible sign of representativeness and identification errors.

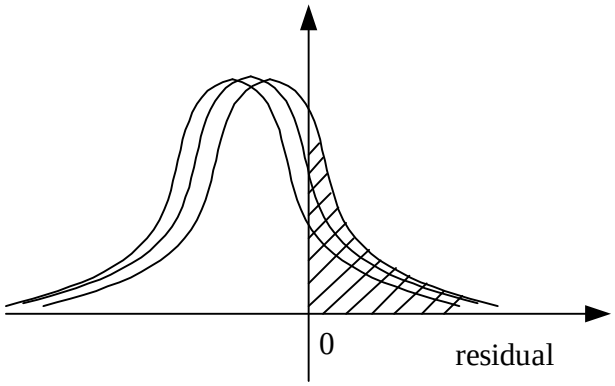

FIGURE 4-2 Positive and negative deviations in estimating split fractions.

Based on the imprecise knowledge of $E(\frac{c^{id}{}_{(l,s,t)}}{c^{id}{}_{(l,t)}} - \frac{\sum_{i,j,\tau} p_{(l,s,t)(i,j,\tau)} d_{(i,j,\tau)}}{\sum_{i,j,\tau} p_{(l,t)(i,j,\tau)} d_{(i,j,\tau)}}) < 0$, it is more desirable to penalize positive residuals than to penalize negative residuals, as shown in Figure 4-2. A simply weighting scheme is to use a one-sided penalty formulation by choosing $w^{+}_{3} > 0$ and $w^{-}_{3} = 0$. It is easy to verify that, the smaller the $\alpha_{(l,s)}$, the few chances for the deviation being positive and the less information obtained from the AVI counts in this one-sided formulation. In other words, the one-sided penalty formulation becomes weaker when fixed parameters are closer to 0. It

should be noted that, the value of w^+_3 still needs to be jointly determined by considering the relative confidence on link counts and historical demand information. For fixed values of w_1 and w_2, a larger w^+_3 means that the decision maker has more confidence on the quality of AVI data and vice versa. Compared to the least squares form, another advantage of the linear penalty function form is that it is much "smoother" to outlying observations in a sense that large deviations cannot substantially impair the estimation performance. In general, the first joint estimation model is a systematical statistical approach to verify the proposed hypothesis, but it leads to considerable modeling complexity and requires a large number of time-series observations. The one-sided penalty formation, on the other hand, can utilize the imprecise knowledge from the analyst and obviates the need for exactly estimating time-dependent and location-dependent representation identification rates and representativeness parameters.

4.5 Bi-Level Estimation Procedure

The above bi-level programming problem can be solved by an iterative solution algorithm.

Step 1: (Initialization) $k = 0$. Start from an initial guess of the traffic demand matrix $\hat{D}^0$, obtain flow proposition matrix $\hat{P}^k$ from the DTA simulator.

Step 2: (Optimization) Substituting flow proportion matrix $\hat{P}^k$, solve the upper-level estimation problem to obtain demand D^k.

Step 3: (Simulation) Using estimated demand $\hat{D}^k$, run the DTA simulator to generate new flow proportions $\hat{P}^{k+1}$.

Step 4: (Evaluation) Calculate the deviation between simulated link flows and observed link counts, the deviation between estimated demand $\hat{D}^{k+1}$ and target demand *G*, as well as the deviation between estimated link-to-link split fractions and observed link-to-link split fractions.

Step 5: (Convergence test) If the convergence criterion is satisfied (estimated demand is stable or no significant improvement in the overall sum of deviations), stop; otherwise $k = k + 1$ and go to Step 2.

The multi-objective optimization techniques presented in Chapter 3 can be applied here to determine the weights in the upper-level objective function. Standard nonlinear optimization algorithms, such as the projected gradient algorithm, can be applied to solve the proposed nonlinear least squares estimation problem. To evaluate gradient directions for the objective function shown in Equation (4.19), the derivative of objective function Z_3 with respect to demand $d_{(i,j,\tau)}$ is derived as

$$\frac{\partial Z_3}{\partial d_{(i,j,\tau)}} = 2w_3 \sum_{l \in L_{vi}, t} \sum_{s \in L_{vi}} \left[\frac{\sum_{i,j,\tau} \hat{p}_{(l,s,t)(i,j,\tau)} d_{(i,j,\tau)}}{\sum_{i,j,\tau} \hat{p}_{(l,t)(i,j,\tau)} d_{(i,j,\tau)}} - \frac{c^{id}{}_{(l,s,t)}}{c^{id}{}_{(l,t)}} \right] \times \frac{\hat{p}_{(l,s,t)(i,j,\tau)} - \hat{p}_{(l,t)(i,j,\tau)}}{\left[\sum_{i,j,\tau} \hat{p}_{(l,t)(i,j,\tau)} d_{(i,j,\tau)} \right]^2} \tag{4.28}$$

To further obtain a computationally feasible algorithm for the upper-level estimation problem, one can linearize the nonlinear function of split fraction $b_{(l,s,t)}$ based on a first-order Taylor series approximation around previous estimate of $\hat{D}^k$ at iteration *k*,

$$\frac{\sum_{i,j,\tau}\hat{p}_{(l,s,t),(i,j,\tau)}d_{(i,j,\tau)}}{\sum_{i,j,\tau}\hat{p}_{(l,t)(i,j,\tau)}d_{(i,j,\tau)}}=\frac{\sum_{i,j,\tau}\hat{p}_{(l,s,t),(i,j,\tau)}\hat{d}^{k}_{(i,j,\tau)}}{\sum_{i,j,\tau}\hat{p}_{(l,t)(i,j,\tau)}\hat{d}^{k}_{(i,j,\tau)}}+\sum_{i,j,\tau}\left[\frac{(\hat{p}_{(l,s,t)(i,j,\tau)}-\hat{p}_{(l,t)(i,j,\tau)})}{(\sum_{i,j,\tau}\hat{p}_{(l,t)(i,j,\tau)}\hat{d}^{k}_{(i,j,\tau)})^{2}}(d_{(i,j,\tau)}-\hat{d}^{k}_{(i,j,\tau)})\right]+\nu_{(l,s,t)} \tag{4.29}$$

where $\nu_{(l,s,t)}$ is the approximation error in the above Taylor series expansion for link-to-link split fraction $b_{(l,s,t)}$.

Substituting Equation (4.29) into measurement equation (4.12) for link-to-link split fractions yields the following transformed measurement equation in a more convenient form:

$$y^{k}_{(l,s,t)}=\sum_{i,j,\tau}h^{k}_{(l,s,t)}d_{(i,j,\tau)}+\varsigma_{(l,s,t)}, \tag{4.30}$$

where transformed measurement

$$y^{k}_{(l,s,t)}=\frac{c^{id}{}_{(l,s,t)}}{c^{id}{}_{(l,t)}}-\frac{\sum_{i,j,\tau}\hat{p}_{(l,s,t),(i,j,\tau)}\hat{d}^{k}_{(i,j,\tau)}}{\sum_{i,j,\tau}\hat{p}_{(l,t)(i,j,\tau)}\hat{d}^{k}_{(i,j,\tau)}}+\sum_{i,j,\tau}\left[\frac{(\hat{p}_{(l,s,t)(i,j,\tau)}-\hat{p}_{(l,t)(i,j,\tau)})\cdot\hat{d}^{k}_{(i,j,\tau)}}{(\sum_{i,j,\tau}\hat{p}_{(l,t)(i,j,\tau)}\hat{d}^{k}_{(i,j,\tau)})^{2}}\right] \tag{4.31}$$

transformed mapping proportion is

$$h^{k}_{(l,s,t)}=\frac{\hat{p}_{(l,s,t)(i,j,\tau)}-\hat{p}_{(l,t)(i,j,\tau)}}{(\sum_{i,j,\tau}\hat{p}_{(l,t)(i,j,\tau)}\hat{d}^{k}_{(i,j,\tau)})^{2}} \tag{4.32}$$

and new combined error term $\varsigma'_{(l,s,t)}=\nu_{(l,s,t)}+\zeta_{(l,s,t)}$.

The resulting deviation function for the AVI sample becomes

$$Z_3(D,\hat{D}^{k},C^{id})=w_3\sum_{l\in L_{vi},t}\sum_{s\in L_{vi}}\sum_{i,j,\tau}(y^{k}_{(l,s,t)}-h^{k}_{(l,s,t)}\cdot d_{(i,j,\tau)})^{2}, \tag{4.33}$$

and now objective function Z_3 in the upper-level is converted into a tractable linear least squares function.

4.6 Identification Conditions

The dynamic OD demand estimation problem has $|I|\times|J|\times T_d$ unknown demand variables. On the other hand, loop detectors can provide at most $|L_{lc}|\times T_o$ independent link volume observations. AVI data can provide at most $|L_{vi}|\times|L_{vi}|\times T_o$ link-to-link tagged vehicle counts, which dramatically alleviate the under-specification problem of OD estimation. With prior static information on dynamic OD demand, the number of independent observations should be greater than the number of unknown demand variables so as to identify a unique solution for the transformed linear least squares problem, leading to the necessary condition for uniquely estimating a dynamic OD demand matrix as

$$|L_{lc}|\times T_o + |L_{vi}|\times|L_{vi}|\times T_o + |I|\times|J| \geq |I|\times|J|\times T_d. \tag{4.34}$$

AVI data only provide OD demand distribution information, so OD demand volume information from loop counts and historical OD tables must be added to identify a unique solution. To quantify the minimum requirement for additional link counts and prior information, let us consider an extreme case, where AVI readers cover all the entry/exit links of each zone so that OD identified vehicle count $c^{id}_{(i,j,\tau)}$ are available for each OD pair.

$$\frac{c^{id}_{(i,j,\tau)}}{c^{id}_{(i,\tau)}} = \frac{d_{(i,j,\tau)}}{\sum_j d_{(i,j,\tau)}} + \eta_{(i,j,\tau)} \tag{4.35}$$

Because of summation equation $c^{id}_{(i,\tau)} = \sum_j c^{id}_{(i,j,\tau)}$, one OD split fraction observation is obviously redundant for origin zone *i*. Therefore, available AVI data provide at most $|I|\times(|J|-1)\times T_d$ independent observations in the proposed formulation, and the minimum information requirement for link counts and historical demand data is

$$|L_{lc}|\times T_o + |I|\times|J| \geq |I|\times T_d. \quad (4.36)$$

It should be noticed that, several factors significantly decrease the number of independent observations in an actual AVI data set. First, many OD flows might not be captured by any AVI readers due to the partial coverage of AVI detectors in a general network. This situation occurs when there is no feasible path from link *l* to link *s*, or the related feasible paths are not considered in the choice set. More precisely, if for OD pair (*i*,*j*)

$$p_{(l,s,t)(i,j,\tau)} = 0 \qquad \forall l, s \in L_{vi}\ , \quad (4.37)$$

then no information for OD pair (i,j) can be extracted from the AVI point-to-point sample data. In addition, as mentioned in the previous chapter, time-dependent link counts from two closely related detectors can be highly correlated. In the same way, time-dependent point-to-point AVI measures from two neighboring readers might produce few independent observations on OD demand distributions.

4.7 Numerical Experiments

4.7.1 Experiment Design

This section is intended to evaluate the performance of the proposed dynamic OD demand estimation models under different levels of market penetration rates and

identification rates. The experiments are conducted based on a simplified Irvine test bed network shown as Figure 3-1, which includes 16 OD zones, 31 nodes and 80 directed links (32 freeway and 48 arterial). The time of interest is the morning peak period (6:30 am – 8:30 am). Traffic link counts are measured on 16 links; at 30-second interval on 10 freeway links, and at 5-minute interval on 6 arterial links, but no real-world AVI traffic measurements are currently available in this data set. In order to capture a realistic OD demand pattern for the underlying real network, this study first uses actual link counts and a historical static demand table to estimate the OD traffic demand matrix, and then uses the estimated matrix as the "true" OD demand in the following experiments. The "true" OD demand is loaded onto the network using a DTA simulation program (i.e. DYNASMART-P) to generate both link counts and point-to-point counts as the "ground-truth" observations in the synthetic data set. The DTA simulator is also used to provide link flow proportions and link-to-link flow proportions to the OD estimation program. Note that, to ensure the internal consistency between link flow measurements and point-to-point flow measurements, this study uses simulated link counts as estimation input, instead of the actual link flow observations from the field data. In addition, stochastic disturbances following an independent normal distribution with zero mean is added into simulated link counts so as to emulate the effect of measurement errors, and the standard deviation of random errors is set to 10% of the corresponding simulated link volume. AVI readers are assumed to cover all the entry/exit links of each OD demand zone, indicating that OD AVI counts are available for each OD pair. In addition, both departure time interval in the dynamic OD demand matrix and the AVI observation

time interval are set to 5 minutes. For each experiment in this study, the initial demand is assumed to be 50% of the assumed actual values.

To quantify the accuracy of estimation results, the Root Mean Squared Error (RMSE) is selected as the performance measure:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{\tau=1}^{\Gamma}(\hat{d}_{(i,j,\tau,)} - d_{(i,j,\tau)})^2}{\Gamma \times \mathrm{I} \times \mathrm{J}}}$$

where $d_{(i,j,\tau)}$ = "true" demand volume for OD pair (i, j) during departure interval τ,

$\hat{d}_{(i,j,\tau)}$ = estimated demand volume for OD pair (i, j) during departure interval τ,

I = number of origin zones,

J = number of destination zones,

Γ = number of departure time intervals in the dynamic OD demand table.

4.7.2 Effect of Market Penetration Rates of AVI Tags

Identification rates in the first set of experiments are assumed as 100%. To test the nonlinear ordinary least squares model without fixed effect parameters, i.e. the objective function shown in Equation (4-19), the following two scenarios are used: (1) all OD pairs have the exactly same market penetration rate; (2) the market penetration rates are assumed to follow an independent uniform distribution with a range [$0.75\overline{\beta}$, $1.25\overline{\beta}$] among different origin-destination pairs at different departure times, where $\overline{\beta}$ is the mean market penetration rate for all OD pairs. Figure 4-3 shows the change of the solution quality in response to increasing market penetration

rates. Each data point in the figure represents the mean value of RMSE from five random replications. Overall, the estimation performance under the random market penetration scenario is relatively worse than that under constant market penetration. Because random variations in market penetration rates introduce almost equal possibilities of the small and large sample size with respect the mean sample size, the slightly higher estimation errors in the random penetration case suggest that the loss of estimation quality due to the small sample size is more significant than the gain due to the large sample size.

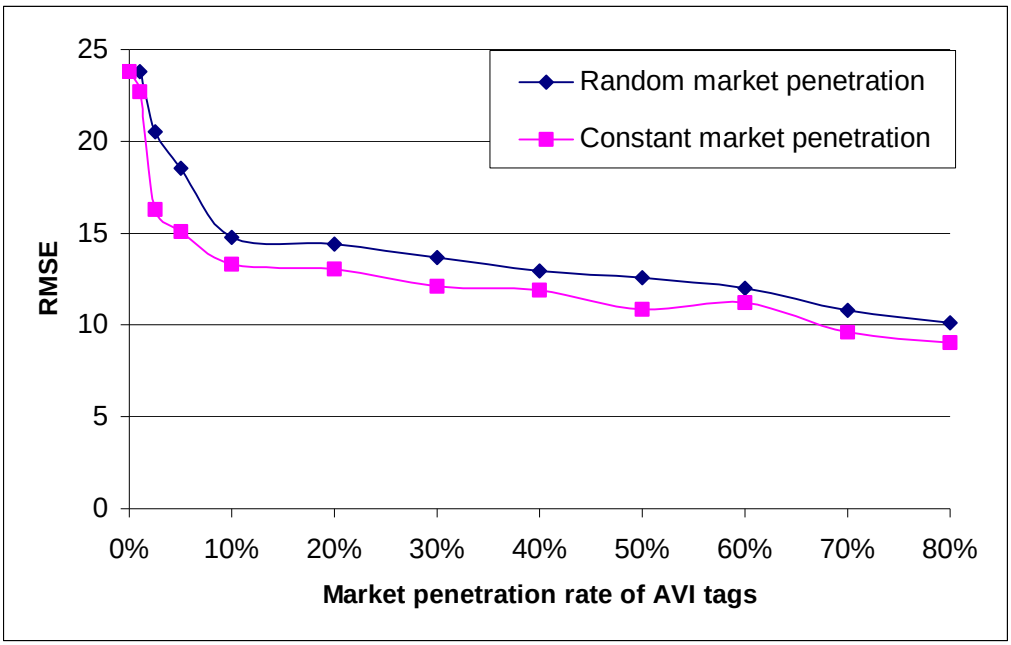

FIGURE 4-3 Dynamic OD estimation under different market penetration rates.

As shown in Figure 4-3, increasing market penetration rates gradually improves the quality of final OD demand estimates. When the market penetration rate is zero, OD demand is estimated only using dynamic link count data, corresponding to the do-nothing case. At a market penetration rate of 1%, the additional AVI information does not lead to significant error reductions compared to the do-nothing

case for both penetration scenarios. At such a low market penetration level, the number of tagged vehicles observed at each observation time interval is too small to provide reliable samples for point-to-point split fractions, leading to large sampling errors for the corresponding measurement equations. When the market penetration rate increases to 5%, the nonlinear OLS model produces nearly 20% error reductions even in the case of random tag penetration. The quality improvement in terms of RMSE further rises to around 40% at a 10 % market penetration rate, and extra error reductions become relatively smaller beyond this market penetration level. In brief, the experiment results clearly demonstrate that the proposed OD estimation model is able to effectively utilize AVI information in the presence of random market penetration rates. On the other hand, a certain level of market penetration of vehicle transponders is necessary to achieve meaningful quality improvement using additional AVI data.

4.7.3 Effect of Identification Rates

In the next set of experiments, the market penetration rates are assumed to follow an independent uniform distribution with a range [7.5%, 12.5%] and the recognition rates of AVI readers are assumed to follow a random uniform distribution with mean γ and a range [γ–0.1, γ+0.1].

Table 4-1 summarizes the estimation errors resulted from the joint estimation model with fixed effect parameters and the one-sided positive penalty formulation (i.e. $w^{+}_{3} > 0$ and $w^{-}_{3} = 0$), and the corresponding percentage improvement compared to the do-nothing case (without AVI information). Each data point in the table represents the mean value of estimation performance from five random replications.

Table 4-1 Performance of estimation models in the presence of identification errors.

Identification Rates	[0.8, 1.0]	[0.7,0.9]	[0.6,0.8]	[0.5,0.7]	[0.4,0.6]
Joint Estimation Model	21.5	22.0	22.8	23.73	23.65
% Improvement	9.4%	7.6%	4.2%	0.2%	0.5%
One-sided penalty model	18.12	19.17	20.84	22.22	23.02
% Improvement	23.7%	19.4%	12.4%	6.5%	3.2%

The experimental results show that, the joint estimation model only produces a marginal performance improvement by using AVI information. The difficulty in applying this model can be attributed to the increase in unknown variables and the resulting high nonlinearity in the model structure due to the inclusion of the fixed effect parameters. In contrast, even under a medium level of identification rates [0.7, 0.9], the one-sided linear penalty estimator is still able to reduce error by nearly 30 percent, revealing that this parsimonious structure is quite robust to the imperfect observations. As expected, the estimation errors from the one-sided model become larger when identification rates decrease, and the value of AVI data tends to be insignificant. In fact, as observed values of split fractions are significantly smaller from the true values (due to identification errors), the estimated OD demand is less likely to be penalized by the one-sided function, leading a very weak estimator to reduce the solution variance. Overall, the one-sided linear penalty formulation presents a tractable and intuitive approach for incorporating partially observed point-to-point sensor data with small identification errors.

4.7.4 AVID etector Coverage

Obviously, the complete AVI coverage scheme in the above experiments requires extensive detector installation and maintenance efforts for a general traffic network. It is desirable to maintain the estimation quality while minimizing the number of AVI detectors in the network. A group of experiments are first conducted to reveal the relation between the estimation quality and the percentage of covered OD demand flows by AVI readers. The other experimental settings are 100% identification rate and uniformly distributed market penetration rates between [7.5%, 12.5%].

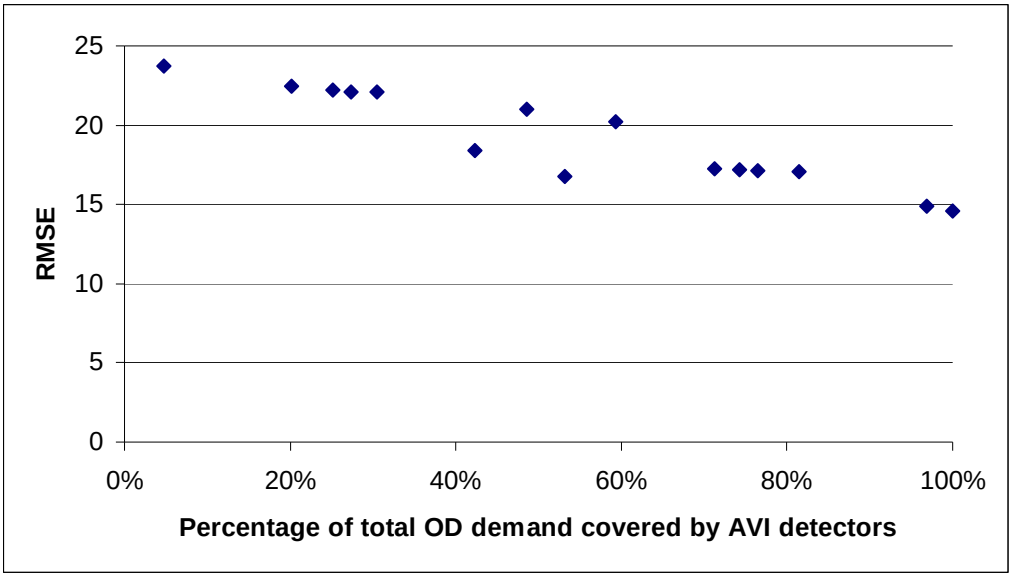

FIGURE 4-3 Relationships between estimation performance and AVI detector coverage

Total of 16 detector location schemes are randomly generated, and the resulting estimation errors are shown in Figure 4-3. Clearly, the estimation errors decrease with increasing AVI coverage of OD demand flows. The above strong correlation between these two attributes suggests a basis for optimizing AVI detector

locations for the estimation of population OD demand flows, that is, it is advantageous to cover as more OD demand flows as possible with limited number of AVI detectors.

The following experiments represent a preliminary attempt to optimize the AVI location for OD demand estimation purposes. The Irvine test bed network in the study has a clear triangular structure, where the OD demand flows among zones 1, 4 and 16 account for 36.4% of total OD trip desires. Thus, AVI detectors should be first located on the entry/exit links for these critical OD zones. Next, OD pair (16→12) is the uncovered dominant OD pair, which has the largest number of travelers beyond the internal OD demand flows between the zone subset {1, 4 and 16}, so zone 12 is added into the AVI coverage subset. In the same way, zones 13, 15, 5, 2 and 11 are sequentially added into the coverage plan, and the corresponding OD demand estimation errors at each step are reported in Table 4-2.

Table 4-2 Estimation performance under different AVI location schemes.

# of Zones Covered	3	4	5	6	7	8	9	16
Zones Covered	1,4,16	+12	+13	+15	+5	+2	+11	
Percentage of demand coverage	36.4%	47.3%	52.5%	61.4%	73.1%	76.2%	84.6%	100%
RMSE	20.896	19.119	18.41	16.812	16.647	16.061	15.378	14.775
Percentage improvement	12.1%	19.6%	22.6%	29.3%	30.0%	32.5%	35.3%	37.9%

Figure 4-4 further depicts the trade-off curve between the number of zones covered and the estimation error reduction. Compared to the do-nothing case, nearly 30% error reduction is obtained by locating AVI detectors to cover the six major

zones, which capture 61.4% of the total OD flows in the study network. For the remaining OD zones carrying less trip flows, the marginal improvements in fit are not as significant. Particularly, when 10 major OD demand zones are selected into the coverage plan, only 12.6% of total OD flows remain uncovered by the AVI detectors, leading to very small potential for further estimation error reduction. According to this preliminary study, locating AVI detectors on major OD demand zones with large traffic attraction/production can capture the essential OD distribution pattern in the network and dramatically improve the quality of OD estimates.

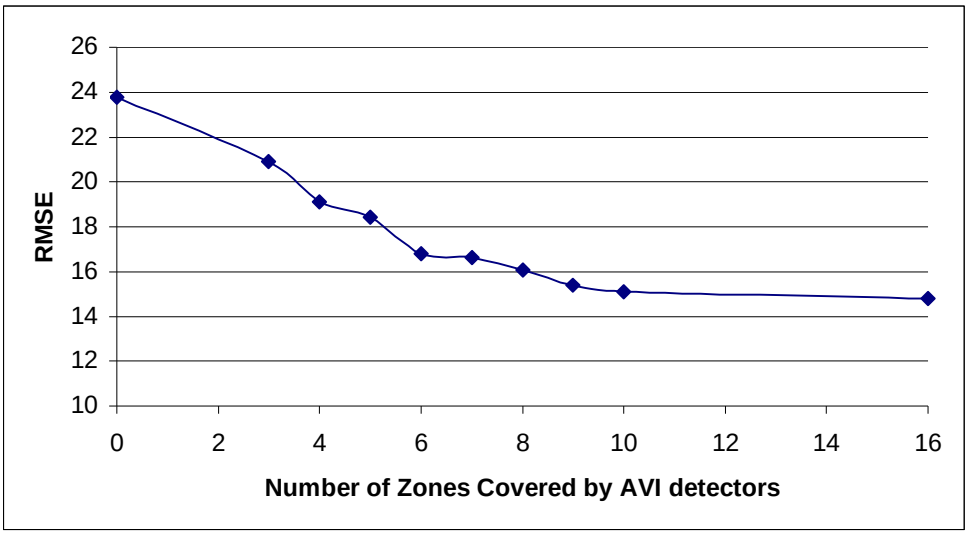

FIGURE 4-4 Trade-offs between estimation errors and number of zones covered by AVI detectors.

4.8 Summary

The emerging AVI technique provides valuable point-to-point flow observations for estimating dynamic OD trip desires. This chapter proposes a novel OD demand estimation approach to effectively exploit OD demand distribution information from AVI counts. A nonlinear ordinary least squares model is presented

to combine AVI counts with other available information sources into a multi-objective optimization framework. A joint estimation formulation and a two-sided linear penalty formation are further developed to take into account identification and representativeness errors. The resulting models are solved using an iterative bi-level estimation framework. Based a synthetic data set using the simplified Irvine test bed network, this study evaluates the performance of new estimation models and provides the following key findings. (1) Sufficient market penetration is required to obtain reliable information from AVI counts. (2) In the presence of identification errors, a parsimonious structure accounting for imprecise information can provide more robust estimates than a complex joint estimation model. (3) For the estimation of population OD demand flows, it is advantageous to locate AVI detectors on major OD demand zones with large traffic attraction/production to capture the essential OD distribution pattern in the network.

5. A Structural State Space Model for Real-Time OD Demand Estimation and Prediction in a Day-to-Day Updating Framework

5.1 Introduction

The off-line dynamic OD demand estimation problem has been presented and solved under different information sources in the previous chapter. This chapter considers the dynamic OD demand estimation and prediction problem for real-time dynamic traffic assignment operations. As discussed in Chapter 2, there is a great need for systematically integrating historical demand information and structural changes into a real-time demand process model, in order to provide accurate and robust demand prediction under regular and irregular conditions. In addition, it is also advantageous to progressively update the *a priori* estimate of the regular demand pattern using new real-time estimation results. Section 5.2 first overviews the recursive OD demand estimation-prediction mechanism and the demand data flow structure in the real-time DTA system. Section 5.3 introduces the core mathematical model, which formulates demand deviations from the *a priori* estimate of the regular pattern using a polynomial trend filter. This is followed by a detailed explanation of the process equation, measurement equation in the structural state space formulation. Additionally, the values of the historical demand information and polynomial trend components are analyzed. In Section 5.4, a Kalman filter based updating framework is further proposed to keep track of the up-to-date regular demand pattern using real-time demand estimates.

5.2 Recursive OD Demand Estimation-Prediction Mechanism

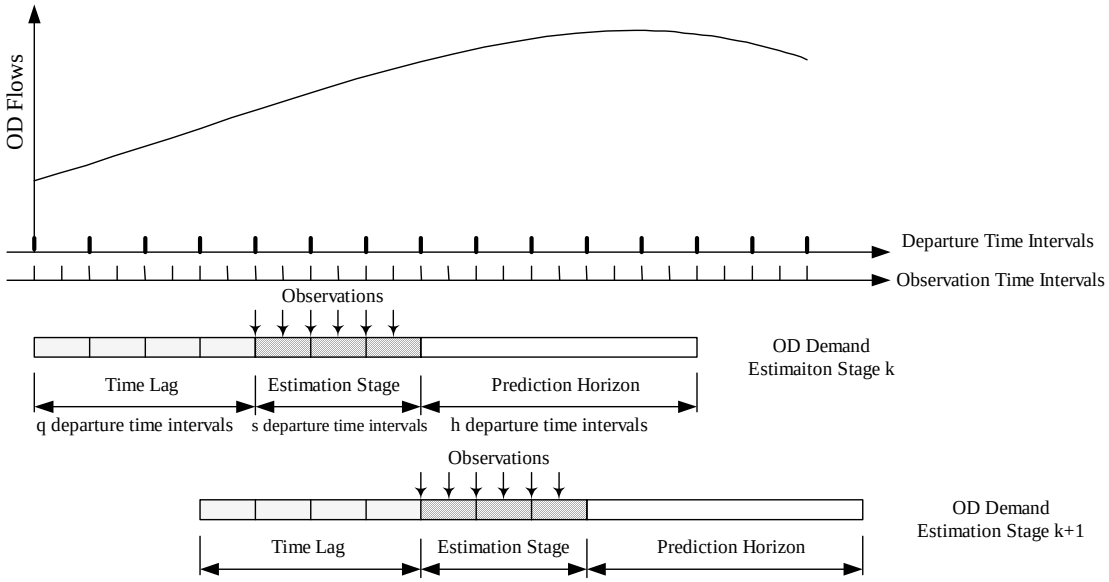

FIGURE 5-1 Illustration of recursive demand estimation-prediction mechanism.

As shown in Figure 5-1, the time is divided into consecutive demand departure time intervals and observation time intervals. The estimation stage (i.e. estimation period) represents the time duration for which an estimation-prediction cycle performs. Note that, an estimation stage can cover multiple departure time intervals and multiple observation time intervals. Starting at the end of an OD demand estimation period, an OD demand prediction horizon corresponds to the time length for which forecasted OD demand should be available for the simulation-assignment procedure. OD demand estimates are calculated using traffic measurements streaming in real-time, and predictions for a given planning horizon are prepared on the basis of the estimation results during the current estimation stage.

Because each traveler takes a certain time to complete his/her trip in a large city network, and the resulting travel time can be very long depending on trip length

and prevailing traffic conditions. The time lag in Figure 5-1 indicates that the traffic flow at the current demand estimation stage can include traffic demand flows departing from pervious estimation stage, leading to a thorny modeling issue in recursive on-line OD estimation-prediction. Failure to recognize the existence of lagged demand would attribute all current flows to demands departing during the current estimation stage, potentially causing serious bias in estimation results. One possible solution of handling lagged OD demand is to extend the dimension of the state variable vector so as to include all the lagged OD demand variables in the current estimation stage (Okutani and Stephanedes, 1984), but the resulting expanded state space could significantly increase the computational complexity. Alternatively, a polynomial trend model can offer a compact representation of lagged demands, which will be addressed in a later section.

Figure 5-2 depicts the demand data flow diagram in the real-time DTA system. As essential supporting components in the rolling horizon approach for solving a real-time DTA problem, real-time OD estimator and predictor are naturally integrated with other on-line DTA modules. Specifically, the DTA simulator is relied upon to generate link proportions for the OD estimation module at the current stage, and the OD demand predictor provides future OD demands for the assignment and simulation in the future stages. These real-time OD demand predicts are regulated by the OD demand consistency checking module before being loaded into the real-time network state estimator (RT-DYNA). Since the current state of RT-DYNA is forwarded into the network state predictor (P-DYNA) as the initial condition every roll period, the corrected OD demand by the consistency checking module is

indirectly updated in the network state prediction component. Finally, real-time OD demand estimates obtained on a new day are extracted to update the regular demand pattern database, which in turn provides the *a priori* estimate of the regular demand pattern for real-time estimation on next day.

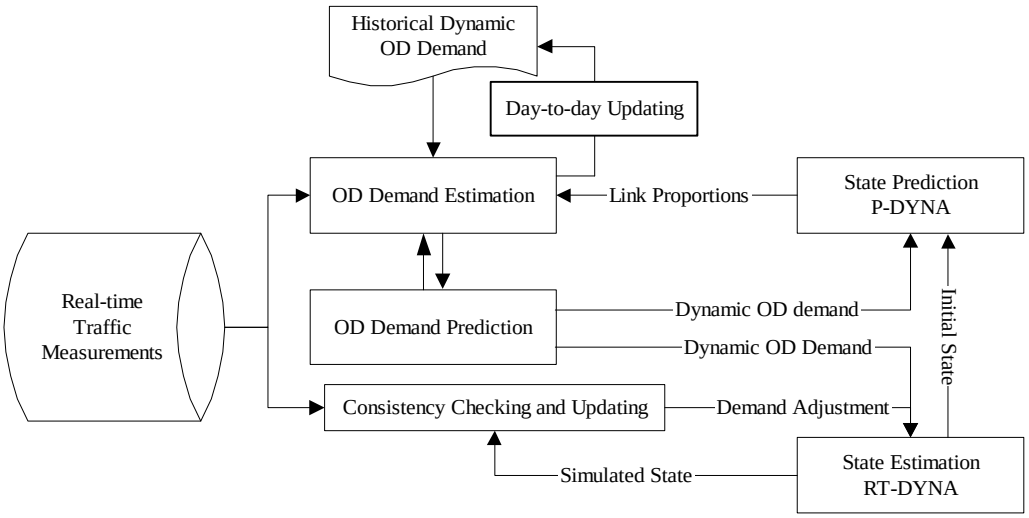

FIGURE 5-2 Demand data flow structure in a real-time DTA system.

The recursive demand estimation-prediction procedure is described in Algorithm 5.1. This scheme entails sequential execution of the OD estimator and predictor, in conjunction with real-time DTA simulators.

Algorithm 5.1. Recursive OD estimation and prediction

Step 1: Receive real-time traffic measurements from the surveillance system.

Step 2: Fetch link proportion data for the current estimation stage from the DTA simulator.

Step 3: (OD estimation) Estimate time-varying OD demand matrices involved in the current estimation stage using the Kalman filtering method.

Step 4: (OD prediction) Predict OD demand over next planning horizon to make latest demand information available for the network state predictor and estimator.

Step 5: Advance estimation period forward, and then go back to Step 1.

5.3 Kalman Filtering Model of the Dynamic OD Demand Estimation and Prediction Problem

5.3.1 Notation and Problem Definition

For convenient reference, the notation used in the real-time OD estimation and prediction model is first presented as follows.

l = subscript for link with traffic measurements, l=1,…,Nod.

(i, j) = subscript for origin-destination pair, (i, j) =1,…, N_{od}.

τ= subscript for aggregated departure time interval, τ =1, 2, …

t = subscript for observation time interval, i.e. sampling time interval, t=1, 2, …

k = subscript for estimation stage, k = 1, 2, 3, …

n = number of observation intervals per departure time interval.

s = number of departure time intervals per estimating period.

h = prediction horizon in numbers of departure time intervals.

q = maximum lag length in numbers of departure time intervals, i.e. the traffic flow at the current departure time interval τ might depart from intervals τ -1, τ -2, …, τ -q.

$c_{(i,t)}$ = number of vehicles measured on link l, during observation interval t .

$d_{(i,j,\tau)}$ = demand volume from origin-destination pair (i, j) during departure time interval τ.

$LP_{(l,t),(i,j,\tau)}$ = link proportions, that is the proportion of vehicles on link l at observation time t, coming from OD pair (i, j) at departure time τ, to the total demand of OD pair (i, j) at departure time τ.

$d^{r}_{(i,j,\tau)}$ = demand volume in regular demand pattern for OD pair (i, j) during departure time interval τ .

$\tilde{d}^{r}_{(i,j,\tau)}$ = *a priori* estimate of regular demand volume for OD pair (i, j) during departure time interval τ.

$\mu_{(i,j,\tau)}$ = structural deviation between *a priori* estimate $\tilde{d}^{r}_{(i,j,\tau)}$ and true demand $d^{r}_{(i,j,\tau)}$.

$\varepsilon_{(i,j,\tau)}$ = error term in approximating true demand for OD pair (i, j) with departure time τ.

$\mu'_{(i,j,\tau)}$, $\mu''_{(i,j,\tau)}$ = first and second derivatives of demand deviation $\mu_{(i,j,\tau)}$, respectively.

p = order index of a polynomial model.

$\mu^{(p)}_{(i,j,\tau)}$ = p^{th}-order derivative of demand deviation $\mu_{(i,j,\tau)}$.

λ = maximum order of a polynomial model.

$w^{(p)}_{(i,j,\tau)}$ = evolution noise for p^{th}-order derivative of demand deviation $\mu_{(i,j,\tau)}$.

$u_{(l,t)}$ = error term in estimation of link observation $c_{(l,t)}$.

$v_{(l,t)}$ = combined error term due to $u_{(l,t)}$ and $\varepsilon_{(i,j,\tau)}$ for link observation $c_{(l,t)}$.

$\hat{d}_{(i,j,\tau)}$ = estimated mean value of $d_{(i,j,\tau)}$.

$\hat{\mu}^{(p)}{}_{(i,j,\tau)}$ = estimated mean value of $\mu^{(p)}{}_{(i,j,\tau)}$.

Z_k= state variable vector at stage *k*.

Y_k= measurement vector at stage *k*.

H_k= measurement matrix, relating measurement Y_k to state Z_k.

w_k= process noise at stage *k*.

v_k= measurement noise at stage *k*.

$\hat{Z}_{k,k-1}$ = prediction of Z_k using observations up to stage *k*-1, i.e.

$E\left(Z_k \middle| Y_1, Y_2, \cdots, Y_{k-1}\right)$.

$\hat{Z}_{k,k}$ = estimation of Z_k using observations up to stage *k*, i.e. $E\left(Z_k \middle| Y_1, Y_2, \cdots, Y_k\right)$.

$P_{k,k-1}$= predicted state covariance matrix of Z_k at stage *k*-1,i.e. $Var\left(Z_k - \hat{Z}_{k,k-1}\right)$.

$P_{k,k}$= estimated state covariance matrix of Z_k at stage *k*, i.e. $Var\left(Z_k - \hat{Z}_{k,k}\right)$.

Consider a traffic network consisting of multiple origins $i \in I$ and destinations $j \in J$, as well as a set of nodes connected by a set of directed links. The time is discretized into departure time intervals τ =1, 2, … , and the time-dependent OD trip desires are expressed as the number of vehicle trips $d_{(i,j,\tau)}$, traveling from origin zone *i* to destination zone *j* during departure time interval τ, where $i \in I$, $j \in J$ and τ=1, 2, ….Link observations $c_{(l,t)}$ are available on link $l \in L$ during observation interval *t*=1, 2, …. The *a priori* estimate of regular demand volume for OD pair (*i*,*j*) during

departure time interval τ, denoted as $\tilde{d}^{r}_{(i,j,\tau)}$, is obtained from the off-line OD estimation results or day-to-day demand updating using previous real-time demand estimates. Given prior information on OD trips and link observations at the current stage, the dynamic OD demand estimation and prediction problem seeks to estimate time-dependent demand flows involved at the current estimation stage and to predict OD demand flows during the prediction horizon.

5.3.2 Transition Equation

As discussed previously, the true demand can be decomposed into three components, namely, regular patterns, structural deviations and random fluctuations. In reality, only the *a priori* estimate $\tilde{d}^{r}_{(i,j,\tau)}$ of the regular demand, reflecting prior survey data and surveillance information up to the previous day, is available before performing real-time estimation on the current day. For this reason, the true demand $d_{(i,j,\tau)}$ in the following study is modeled as a linear combination of the *a priori* estimate, structural deviation and random disturbance:

$$d_{(i,j,\tau)} = \tilde{d}^{r}_{(i,j,\tau)} + \mu_{(i,j,\tau)} + \varepsilon_{(i,j,\tau)}, \qquad (5.1)$$

where the random disturbance term is assumed to follow a Normal distribution with zero mean. Moreover, a polynomial trend model is introduced here to describe the structural deviations based on the following assumption.

Assumption 5.1 (Polynomial trend) Deviations at time $\tau + \zeta$ can be adequately represented *locally* by an $(\lambda+1)^{\text{th}}$-order polynomial function as Equation (5.2) near time τ for a small value of ζ, while derivatives of higher orders are assumed to be zero: $\mu^{(p)}_{(i,j,\tau)} = 0$ for $p>\lambda$.

$$\mu_{(i,j,\tau+\zeta)} = b_0 + b_1\zeta + b_2\zeta^2 + \cdots b_p\zeta^p + \cdots + b_\lambda\zeta^\lambda \tag{5.2}$$

By Taylor's theorem, the smooth function of $\mu_{(i,j,\tau+\zeta)}$ can be expanded about the point $\mu_{(i,j,\tau)}$ as

$$\mu_{(i,j,\tau+\zeta)} = \mu_{(i,j,\tau)} + \zeta\mu'_{(i,j,\tau)} + \frac{\zeta^2}{2!}\mu''_{(i,j,\tau)} + \cdots + \frac{\zeta^p}{(p)!}\mu^{(p)}{}_{(i,j,\tau)} + \cdots + \frac{\zeta^\lambda}{(\lambda)!}\mu^{(\lambda)}{}_{(i,j,\tau)} \tag{5.3}$$

A comparison of Equations (5.2) and (5.3) indicates that the polynomial coefficients in the original functional form can be obtained directly from

$$b_p = \frac{\mu^{(p)}{}_{(i,j,\tau)}}{p!}. \tag{5.4}$$

A more compact form for the p^{th}-order derivative of a polynomial is

$$\mu^{(p)}{}_{(i,j,\tau+\zeta)} = \sum_{p'=p}^{\lambda} \frac{\zeta^{(p'-p)}}{(p'-p)!}\mu^{(p')}{}_{(i,j,\tau)}. \tag{5.5}$$

For example, a third-order polynomial model in matrix form is

$$\begin{pmatrix} \mu_{(i,j,\tau+\zeta)} \\ \mu'_{(i,j,\tau+\zeta)} \\ \mu''_{(i,j,\tau+\zeta)} \end{pmatrix} = \begin{bmatrix} 1 & \zeta & \zeta^2/2! \\ & 1 & \zeta \\ & & 1 \end{bmatrix} \begin{pmatrix} \mu_{(i,j,\tau)} \\ \mu'_{(i,j,\tau)} \\ \mu''_{(i,j,\tau)} \end{pmatrix}. \tag{5.6}$$

The second assumption allows time-varying trends to evolve stochastically between time stages.

Assumption 5.2. (Evolution process) From stage *k* to stage *k*+1, the change of derivative $\mu^{(p)}{}_{(i,j,\tau)}$ can be described as

$$\mu^{(p)}{}_{(i,j,\tau+s)} = \sum_{p'=p}^{\lambda} \frac{s^{(p'-p)}}{(p'-p)!}\mu^{(p')}{}_{(i,j,\tau)} + w^{(p)}_{(i,j,\tau)} \tag{5.7}$$

where departure time index $\tau = ks$, and $w^{(p)}{}_{(i,j,\tau)} \sim N[0, W^{(p)}{}_{(i,j,\tau)}]$.

In other words, the change of demand derivations from departure time interval τ to τ+s is contributed by a local Taylor series expansion term plus an evolution noise term as a result of time shift s.

Incorporating the evolution noise term to Equation (5.6) yields a transition equation for a third-order polynomial trend model in the Kalman filtering formulation:

$$\begin{pmatrix} \mu_{(i,j,\tau+\zeta)} \\ \mu'_{(i,j,\tau+\zeta)} \\ \mu''_{(i,j,\tau+\zeta)} \end{pmatrix} = \begin{bmatrix} 1 & \zeta & \zeta^2/2! \\ & 1 & \zeta \\ & & 1 \end{bmatrix} \begin{pmatrix} \mu_{(i,j,\tau)} \\ \mu'_{(i,j,\tau)} \\ \mu''_{(i,j,\tau)} \end{pmatrix} + \begin{pmatrix} w_{(i,j,\tau)} \\ w'_{(i,j,\tau)} \\ w''_{(i,j,\tau)} \end{pmatrix}. \quad (5.8)$$

In the above system process equation, the state vector consists of the zeroth to 1th-order derivatives of demand structural deviations from the *a priori* estimate of regular demand pattern, and the transition matrix is independent of the current stage k and related departure time interval τ. The single OD-pair model can be easily extended to consider all the OD pairs in a network. Consider a third-order polynomial filter with departure time τ=ks at stage k, one can define the state vector as

$$Z_k = (\mu_{(1,1,\tau)}, \mu'_{(1,1,\tau)}, \mu''_{(1,1,\tau)}, \mu_{(1,2,\tau)}, \mu'_{(1,2,\tau)}, \mu''_{(1,2,\tau)}, \cdots, \cdots, \mu_{(N_{od},\tau)}, \mu'_{(N_{od},\tau)}, \mu''_{(N_{od},\tau)},)^T$$

and define the transition matrix as

$$A_k = Diag(A_k^{(1,1)}, A_k^{(1,2)}, \cdots, A_k^{(i,j)}, \cdots, A_k^{N_{od}})$$

$$\text{where } A_k^{(i,j)} = \begin{bmatrix} 1 & s & s^2/2! \\ & 1 & s \\ & & 1 \end{bmatrix} \quad (5.9)$$

for (i,j)=1, 2, …, N_{od}.

By assuming the evolution noise w_k as

$$w_k = (w_{(1,1,\tau)}, w'_{(1,1,\tau)}, w''_{(1,1,\tau)}, w_{(1,2,\tau)}, w'_{(1,2,\tau)}, w''_{(1,2,\tau)}, \cdots, w_{(N_{od},\tau)}, w'_{(N_{od},\tau)}, w''_{(N_{od},\tau)})^T$$

the complete transition equation in real-time OD estimation and prediction can be written as

$$Z_{k+1} = A_k Z_k + w_k . \tag{5.10}$$

where $w_k \sim N[0, W_k]$.

To obtain the future demand level with prediction horizon *h*, one needs to first predict the demand deviation at time $\tau + h$ based on estimated derivatives at the current stage, and then substitute the predicted demand deviation and the *a priori* estimate of the regular demand pattern into Equation (5.1). Thus,

$$E\left[d_{(i,j,\tau+h)} \middle| \mu_{(i,j,\tau)}\right] = \tilde{d}^{r}_{(i,j,\tau+h)} + \sum_{p=0}^{\lambda} \frac{h^{(p)}}{p!} E\left[\mu^{(p)}{}_{(i,j,\tau)}\right] \tag{5.11}$$

$$Var\left[d_{(i,j,\tau+h)} \middle| \mu_{(i,j,\tau)}\right] = \sum_{p=0}^{\lambda} \sum_{p'=0}^{\lambda} \left[\frac{h^{(p)}}{p!} \frac{h^{(p')}}{p'!} Cov\left(\mu^{(p)}{}_{(i,j,\tau)}, \mu^{(p')}{}_{(i,j,\tau)}\right) \right] \tag{5.12}$$

where $\tau = ks$.

Equation (5.11) indicates that prediction errors of the proposed model are dependent on the order of model λ and the length of prediction horizon *h*. Evidently, a complicated high-order polynomial model describes the dynamic demand process in greater detail than a low-order model, but it might lead to potential large prediction errors.

5.3.3 Value of Historical Demand Information

It is important to theoretically investigate possible benefits of incorporating the *a priori* estimate of the regular demand pattern in the proposed structural model. The following hypothetic analysis is based on two assumptions: (1) a high-order

polynomial model shown as Equation (5.13) adequately represents the true demand flows; (2) the true OD demand flows are directly measured without observation errors.

$$d_{(i,j,\tau+\zeta)} = a_0 + a_1\zeta + a_2\zeta^2 + \cdots + a_\lambda\zeta^{(\lambda)} + \cdots + a_g\zeta^{(g)}. \quad (5.13)$$

Case (i): If the historical demand pattern is not incorporated into the structural state space model, a low-order polynomial trend model with degree $\lambda<g$ can be fitted as

$$\mu_{(i,j,\tau+\zeta)} = a_0 + a_1\zeta + a_2\zeta^2 + \cdots + a_\lambda\zeta^\lambda \quad (5.14)$$

and the residual is $a_{\lambda+1}\zeta^{(\lambda+1)} + \cdots + a_g\zeta^{(g)}$. The magnitude of the approximation error is dominated by its first residual coefficient $\left|a_{\lambda+1}\right|$.

Case (ii): If the estimate of regular demand pattern $\tilde{d}^r_{(i,j,\tau+\zeta)}$ satisfactorily approximates real OD demand $d_{(i,j,\tau+\zeta)}$, one can express $\tilde{d}^r_{(i,j,\tau+\zeta)}$ in the following polynomial form

$$\tilde{d}^r{}_{(i,j,\tau+\zeta)} = \tilde{a}_0 + \tilde{a}_1\zeta + \tilde{a}_2\zeta^2 + \cdots + \tilde{a}_\lambda\zeta^{(\lambda)} + \cdots + \tilde{a}_g\zeta^{(g)}. \quad (5.15)$$

and $\left|a_p - \tilde{a}_p\right| < \left|a_p\right|$ for $0 \le p \le g$ (5.16)

Then, the deviation between the true demand and the historical demand pattern,

$$d_{(i,j,\tau+\zeta)} - \tilde{d}^r_{(i,j,\tau+\zeta)} = (a_0 - \tilde{a}_0) + (a_1 - \tilde{a}_1)\zeta + \cdots + (a_\lambda - \tilde{a}_\lambda)\zeta^{(\lambda)} + \cdots + (a_g - \tilde{a}_g)\zeta^{(g)} \quad (5.17)$$

Using a low-order polynomial trend model to approximate the structural deviations yields the following estimation result

$$\mu_{(i,j,\tau+\zeta)} = (a_0 - \tilde{a}_0) + (a_1 - \tilde{a}_1)\zeta + \cdots + (a_\lambda - \tilde{a}_\lambda)\zeta^{(\lambda)} \quad (5.18)$$

where the dominating truncation error is bounded by coefficient $\left|a_{\lambda+1} - \tilde{a}_{\lambda+1}\right|$.

The assumption in (5.16) indicates that the magnitude of the residual error in the polynomial trend model can be reduced from $\left|a_{\lambda+1} - \tilde{a}_{\lambda+1}\right|$ to $\left|a_{\lambda+1}\right|$ due to additional historical demand information. From the linear regression standpoint, the regular daily pattern can be viewed as a good explanatory regressor that absorbs a considerable amount of variation in the independent variable (i.e. true dynamic demand). Thus, compared to a pure polynomial model, the proposed structural model with the regular pattern component leads to smaller regression residual errors, that is, smaller estimation and prediction variance.

With the help of historical demand information, one can apply a polynomial model with lower order to maintain the same representation accuracy. Since the computational complexity of a Kalman filter is on the order of O(N^3), where $N = N_{od} \times (\lambda+1)$ in our case, the reduction of the model order dramatically decreases the size of the state vector and resulting real-time computational requirements. In summary, incorporating a reliable estimate of the regular demand pattern is always beneficial for improving real-time OD demand prediction quality and computational efficiency. Morrison (1969) and Brookner (1998) provided more comprehensive treatments on value of the nominal trajectory in a polynomial trend filter.

5.3.4 Value of Polynomial Trend Component

In the context of short-term economic forecasting, the first, second and third-order polynomial models can be viewed as the local level model, local linear growth model and local quadratic growth model, respectively. Regarding connections

between the polynomial trend models and other typical time-series ARIMA models, West and Harrison (1997) demonstrated that, if restrictions are imposed on the autocorrelation structure, the limiting case of an $(\lambda+1)^{th}$ polynomial trend model is equivalent to an ARIMA(0, λ, λ) model. According to the generalized state space architecture proposed by Harvey (1989), an auto-regression (AR) term can be also incorporated into the state variable vector in the proposed structural model to represent the autocorrelation structure in the random disturbance.

The following example is intended to show the importance of combining both the polynomial trend model and historical demand information into the prediction of real-time traffic demand flows. In Figure 5-3, solid lines represent the true demand in the morning peak; dotted lines represent the available historical demand data that represent the regular demand patterns; dash lines represent the predicted demand flows given by different state space models.

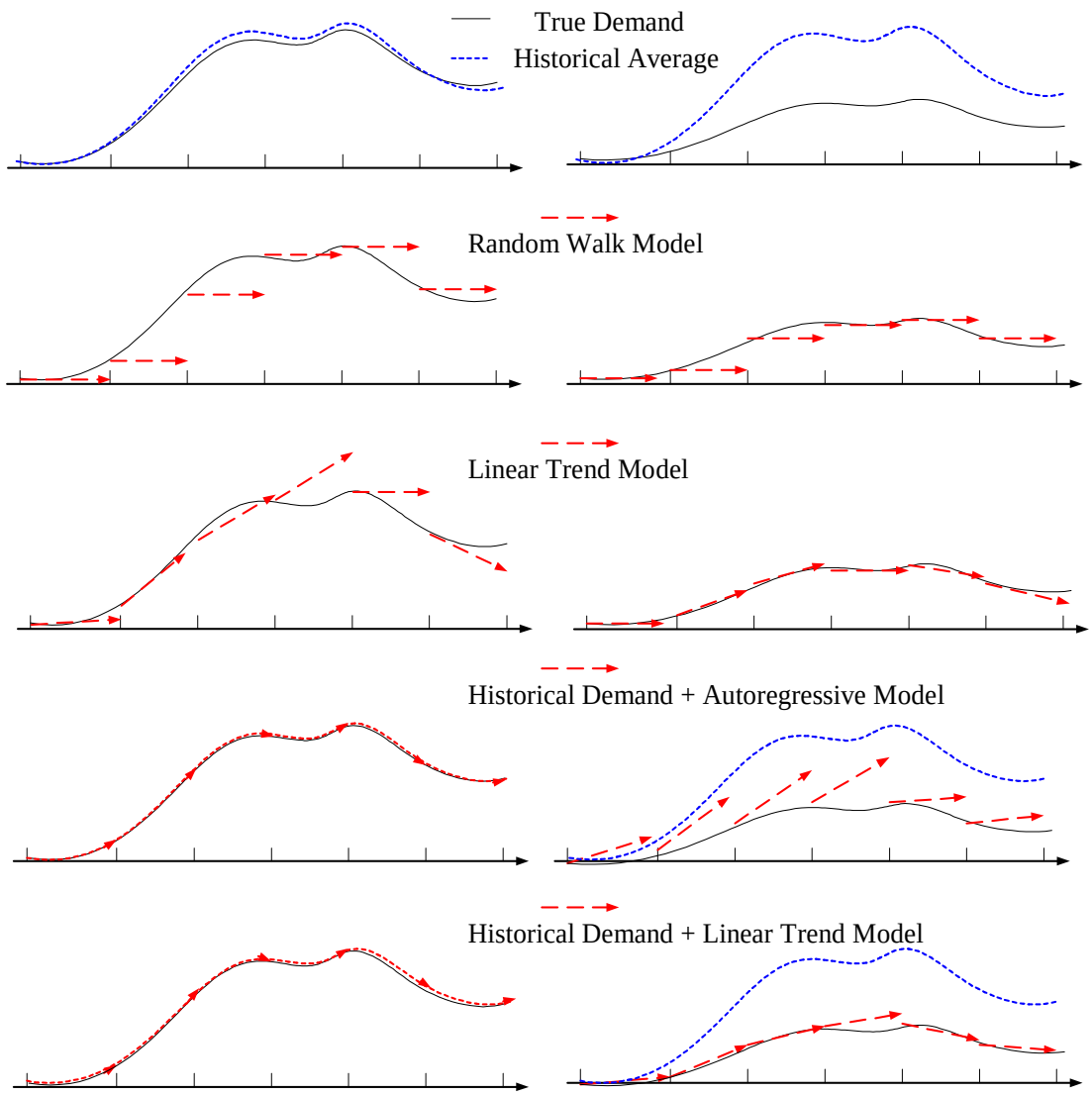

FIGURE 5-3 Illustrative examples of different structural state models.

Two types of demand conditions are shown in the plots on the left hand and on the right hand, respectively

1. The true demand evolves similar to the regular demand pattern.
2. The true demand is significantly lower than the regular demand level, which can be caused by severe weather conditions.

The random walk model in the first row is equivalent to the first order polynomial model (West and Harrison, 1997), and the linear trend model in the second row is the second order polynomial model. Obviously, the linear trend is more suitable to describe the upward or downward pattern in the on-peak and off-peak periods than the random walk model. However, the linear trend model dramatically overestimates the OD demand during the peak time period, because all real-time measurements before the turning point indicate an upward tendency. The error in trend prediction can be explained by the potential uncertainty associated with the high-order polynomial term, as discussed previously.

Under the regular condition, the "historical demand + autoregressive" model perfectly detrends the real-time demand time series and produces accurate prediction results. In the presence of structural changes, however, deviations between the actual demand and the regular demand have non-zero means, leading to significantly biases for the model without the trend component. In contrast, the "historical demand + linear trend" model utilizes the polynomial filter to absorb possible structural difference, producing robust prediction results under both regular and irregular conditions. Comparing the historical demand + linear trend" model with the "linear trend" model, one can also verify that additional historical demand information dramatically decreases prediction variance. In brief, both regular pattern and trend components should by closely integrated in the demand representation so as to ensure the prediction accuracy and robustness. Even if the underlying trends for demand structural deviations are negligible, it is advisable to embed a simple trend component

in the space state representation so that possible future changes in the process structure can be monitored and identified.

5.3.5 Measurement Equation

Similar to off-line OD estimation models, a link proportion matrix is used to connect OD demand flows with link observations:

$$c_{(l,t)} = \sum_{i,j} \sum_{\zeta=-q}^{s-1} \left(LP_{(l,t),(i,j,\tau+\zeta)} \times d_{(i,j,\tau+\zeta)} \right) + u_{(l,t)} \tag{5.19}$$

where $\tau = ks$, $(i,j) = 1, 2, \ldots, Nod$, and $kns \le t < (k+1)ns$.

Specifically, the above equation maps all the lagged and current demand flows involved in the current estimation stage to link measurements obtained during the current stage. Error term $u_{(l,t)}$ can be caused by inconsistencies in assumptions about traffic assignment, traffic control and flow propagation, as well as measurement noise.

To relate link measurements to the state vector in the polynomial filter, *substitute* Equations (5.1) and (5.3), the above equation becomes

$$c_{(l,t)} = \sum_{i,j} \sum_{\zeta=-q}^{s-1} \left[LP_{(l,t),(i,j,\tau+\zeta)} \times (\tilde{d}^{r}_{(i,j,\tau+\zeta)} + \sum_{p=0}^{\lambda} \frac{\zeta^{p}}{g!} \mu^{(p)}{}_{(i,j,\tau)} + \varepsilon_{(i,j,\tau+\zeta)}) \right] + u_{(l,t)}, \tag{5.20}$$

Accordingly, one can define the explicit measurement equation in the Kalman formulation as follows.

$$Y_k = H_k Z_k + v_k \tag{5.21}$$

where measurement vector Y_k consists of elements

$$y_{(l,t)} = c_{(l,t)} - \sum_{i,j} \sum_{\zeta=-q}^{s-1} \left(LP_{(l,t),(i,j,\tau+\zeta)} \times \tilde{d}^{\,r}_{(i,j,\tau+\zeta)} \right), \tag{5.22}$$

measurement noise vector V_k consists of elements

$$v_{(l,t)} = \sum_{i,j} \sum_{\zeta=-q}^{s-1} \left(LP_{(l,t),(i,j,\tau+\zeta)} \times \varepsilon_{(i,j,\tau+\zeta)} \right) + u_{(l,t)} . \tag{5.23}$$

The above measurement error term in the transition equation combines the random fluctuations in the OD demand process, estimation errors related to link proportions.

The dimension of measurement matrix H_k is ($Nod \times ns$, $N_{od} \times (\lambda+1)$), and its $(l,t)^{\text{th}}$, $(i, j, p)^{th}$ element is

$$H_{(l,t),(i,j,p)} = \sum_{\zeta=-q}^{s-1} \left(LP_{(l,t),(i,j,\tau+\zeta)} * \frac{\zeta^{(p)}}{p!} \right) \tag{5.24}$$

By applying a polynomial approximation for OD demands during departure time intervals from $ks - q$ to $ks + s - 1$, the proposed polynomial trend filter neatly incorporates the lagged demands into the estimation procedure for the current stage, leading to an efficient state space representation desirable for large-scale network applications.

5.3.6 Solution Algorithm

Assumption 5.3. System error w_k and measurement error v_k are white noise terms uncorrelated with the initial state Z_0 and with each other, where $w_k \sim N(0, W_k)$ and $v_k \sim N(0, V_k)$.

With system equation (5.9), measurement equation (5.21) and assumption 5.3, the classical Kalman filtering algorithm is ready to be integrated into the following recursive dynamic demand estimation and prediction procedure.

Algorithm 5.2. *Real-time dynamic demand OD estimation and prediction*

Step 1: (Initialization) Construct initial estimates $P_{0,0} = Var(Z_0)$ and $\hat{Z}_{0,0} = E(Z_0)$. Let k=1.

Step 2: (Prediction) Propagate the mean and covariance estimates from k-1 to k.

$$\hat{Z}_{k,k-1} = A_k \hat{Z}_{k-1,k-1} \tag{5.25}$$

$$P_{k,k-1} = A_k P_{k-1,k-1} A_k^T + W_k \tag{5.26}$$

Step 3: (Estimation of state variable) After receiving new link proportions and link observations, calculate the weighting matrix as

$$K_k = P_{k,k-1} H_k^T (H_k P_{k,k-1} H_k^T + V_k)^{-1}, \tag{5.27}$$

and then update the *a posteriori* mean and covariance estimates.

$$\hat{Z}_{k,k} = \hat{Z}_{k,k-1} + K_k (Y_k - H_k \hat{Z}_{k,k-1}) \tag{5.28}$$

$$P_{k,k} = (I - K_k H_k) P_{k,k-1} \tag{5.29}$$

Step 4: (Estimation of real-time demand) Calculate the estimation of real-time demand using new estimates $\hat{\mu}_{(i,j,\tau)}$

$$\hat{d}_{(i,j,\tau)} = E\left(\tilde{d}^r_{(i,j,\tau)} + \mu_{(i,j,\tau)} + \varepsilon_{(i,j,\tau)}\right) = \tilde{d}^r_{(i,j,\tau)} + \hat{\mu}_{(i,j,\tau)} \tag{5.30}$$

where $\tau = ks, ks+1, \ldots, (k+1)s$-1.

Step 5: Advance estimation period forward from k to k+1, and then go back to Step 2.

If assuming independence of measurement errors, one can further apply the scalar updating scheme described by Chui and Chen (1991) in order to avoid complicated matrix inversions in a real-time computation setting.

5.4 Adaptive Day-to-Day Updating of Regular Demand Pattern Information

5.4.1 Notation

As discussed earlier, the initial estimate for the regular demand pattern could be unreliable due to limited observations, and the normal daily pattern could evolve smoothly due to intrinsic day-to-day demand dynamics. Hence, it is necessary to update the *a priori* estimate based on new observation data. A desirable updating formulation should be able to adaptively recognize and capture the systematic day-to-day evolution, and also maintain robustness under disruptions due to special events. In response, an optimal updating formulation based on a Kalman filter framework is proposed.

The notation used in the real-time OD estimation and prediction model is extended to the day-to-day context as follows.

m = index for day.

D_m^r = state variable vector of regular OD demand pattern on day m ,consisting of elements $d_{(i,j,\tau)}^r$.

ξ_m = day-to-day evolution variance on day m.

$\hat{D}_m$ = vector of the real-time demand estimate on day *m*, consisting of estimates $\hat{d}_{(i,j,\tau)}$.

η_m = measurement variance matrix on day *m*.

$\hat{D}^r_{m,m-1}$ = predicted state variable vector D^r_m using observations up to day *m*-1 , consisting of elements $\hat{d}^r_{(i,j,\tau)}$.

$\hat{D}^r_{m,m}$ = estimated state variable vector D^r_m using observations up to day *m*.

$\Sigma_{m,m-1}$ = predicted state covariance matrix for the regular demand pattern on day *m*.

$\Sigma_{m,m}$ = estimated state covariance matrix for the regular demand pattern on day *m*.

K_m= Kalman gain matrix for using real-time demand estimates on day *m*.

C_m = vector of traffic observations on day *m*, consisting of elements $c_{(l,t)}$.

LP_m= link proportion matrix on day *m*, consisting of elements $LP_{(l,t),(i,j\tau)}$.

5.4.2 Transition and Measurement Equations

The transition and measurement equations for the day-to-day demand evolution can be written as

Transition Equation:

$$D^r_{m+1} = D^r_m + \xi_m \tag{5.31}$$

Measurement Equation:

$$\hat{D}_m = D^r_m + \eta_m \ . \tag{5.32}$$

Assumption 5. ξ_m and η_m are white noise terms uncorrelated with the initial state D_0^r and with each other, where ξ_m ~N(0, Q_m) and η_m ~N(0, R_m)

According to transition equation (5.31), the regular demand pattern can evolve smoothly from day to day, where stochastic day-to-day evolution is captured by the evolution random term ξ_m. In the measurement equation, since the true demand state cannot be directly observed, the new real-time demand estimate $\hat{D}_m$ is considered as "measurement" incoming everyday.

5.4.3 Updating Algorithm and Parameter Tuning

Following the standard Kalman filtering algorithm, the updating procedure can be summarized as follows.

Algorithm 3. *Day-to-day updating for regular demand pattern estimate*

Step 1: (Initialization) Construct $\hat{D}_{0,0}^r$ and $\Sigma_{0,0}$ as the initial estimated mean and covariance of the regular demand. Let m=1.

Step 2: (Computation of *a priori* estimate) The *a posteriori* estimate $\hat{D}_{m-1,m-1}^r$ on previous day m-1 is used as the *a priori* estimate for current day m. The corresponding covariance matrix is updated by taking evolution noise into account.

$$\hat{D}_{m,m-1}^r = \hat{D}_{m-1,m-1}^r \tag{5.33}$$

$$\Sigma_{m,m-1} = \Sigma_{m-1,m-1} + Q_m \tag{5.34}$$

Step 3: (Real-time OD estimation and prediction) Run the real-time OD estimation and prediction module in conjunction with real-time DTA simulators, to obtain new estimates $\hat{M}_m$ and $\hat{D}_m$ for day m.

Step 4: (Update of gain matrix) Compute the gain matrix using predicted state covariance matrix and measurement variance matrix.

$$K_m = \Sigma_{m,m-1}(\Sigma_{m,m-1} + R_m)^{-1} \tag{5.35}$$

Step 5: (Update of mean and covariance) Update the estimated mean and covariance matrix for the regular demand state vector.

$$\hat{D}^r_{m,m} = \hat{D}^r_{m,m-1} + K_m(\hat{D}_m - \hat{D}^r_{m,m-1}) = \hat{D}^r_{m,m-1} + K_m(\hat{D}_m - \tilde{D}^r_m) = \hat{D}^r_{m,m-1} + K_m\hat{M}_m \tag{5.36}$$

$$\Sigma_{m,m} = (I - K_m)\Sigma_{m,m-1} \tag{5.37}$$

Step 6: Move to the next day, $m = m+1$, and then go back to Step 1.

After accumulating additional information from the new real-time estimation result, the conditional demand estimate contains less uncertainty than the corresponding *a priori* estimate for the regular demand pattern, i.e. $\Sigma_{m,m} \le \Sigma_{m,m-1}$. In order to make this recursive algorithm operational, the next question is how to specify the values of evolution variance Q_m and measurement variance R_m. By using the multi-day OD estimation method discussed in Chapter 3, the variance of the measurement noise associated with $\hat{D}_m$ can be obtained by evaluating the variance of estimated OD demands across several days. Determining the day-to-day variance is generally more difficult, since one cannot directly observe the day-to-day demand evolution process. Recognizing that the proposed day-to-day evolution process can be described as a random walk plus noise model, existing time series techniques are applied to choose appropriate values of process variance. A common approach is to

first assume a constant signal to noise ratio $\gamma = \frac{Q_m}{R_m}$, indicating the ratio of inherent system variance with respect to observational variance. Based on calibrated R_m, $Q_m = \gamma R_m$, so one can select an appropriate signal to noise ratio so as to minimize average prediction errors in the training data sets. The reader is referred to West and Harrison (1997) for a comprehensive treatment on specification of system and observational variances using a Bayesian estimation framework. In early time period of using this updating procedure, considerable uncertainty in the predicted state covariance $\Sigma_{m,m-1}$ results in a high gain factor, implying that the new real-time estimates receive relatively large weighting. After a certain number of iterations, the gain factor becomes stable as it is gradually reaching a steady state. If constant Q and R are assumed, the following limiting behavior for the Kalman gain matrix can be derived (West and Harrison, 1997).

$$\lim_{m\to\infty} K_m = \frac{\gamma}{2}\left(\sqrt{1+4/\gamma}-1\right) \tag{5.38}$$

A typical value of γ can be 0.05, leading to $\lim_{m\to\infty} K_m = 0.2$, so the most recent real-time estimate receives relatively small weighting eventually. If γ =0.5, corresponding to limiting gain factor of 0.5, then the *a priori* estimate and the new real-time estimate share equal importance in determining priori demand information for the next day.

This Kalman filtering formulation provides a least squares unbiased estimator for the regular demand pattern, with the optimal weights on the *a priori* estimate and new information. It is important to note that this recursive prediction-correction algorithm only requires *a priori* mean and covariance statistics at each iteration

instead of the entire historical data series, resulting in efficient storage implementation for on-line applications. Practically, this updating method can be viewed as a moving average method with adaptive weights, depending on the respective reliability of the *a priori* and real-time information sources.

5.5 Summary

Real-time OD estimation and prediction is an important component in real-time dynamic traffic assignment for ATMS/ATIS network applications. This chapter exploits the potential of using a structural state space model to systematically incorporate regular demand pattern information, structural deviations and random fluctuations.

The contributions of this study are as follows. First, a polynomial trend filter is developed to estimate and predict demand deviations from the *a priori* estimate of the regular demand pattern, so as to capture predictable patterns based on historical information and adaptively response to possible structural deviations in demands. The proposed structural model can assist the analyst to deconstruct the demand process and further reveal the underlying irregularities. Second, based on a Kalman filtering framework, an optimal recursive procedure is proposed for updating the regular demand pattern estimate with new real-time estimates and observations obtained every day. This updating formula explicitly accounts from the uncertainty and variance associated with real-time estimates, and the resulting dynamic learning framework can also assist the analyst to progressively investigate the day-to-day behavior of traffic demand. These models can be naturally integrated into the real-

time DTA system and provide an effective and efficient approach to utilize the real-time traffic data continuously in the operational settings.

6. Recursive Approaches for On-Line Consistency Checking and OD Demand Updating

6.1 Introduction

Previous chapter discussed the real-time OD estimation and prediction model in the context of on-line traffic state estimation and prediction. The proposed structural state model improves forecasting reliability under both regular and irregular conditions, and the day-to-day updating model provides an optimal learning mechanism for capturing demand evolution. In addition to the above important demand estimation and prediction supporting functions, there is still one critical operational requirement to be addressed for any real-time simulation based DTA system, that is, ensuring the consistency between the real-*world* system and the internal representation of the DTA simulator. This chapter is concerned with the formulation and computation issues of an OD demand consistency checking and updating model, which aims to reduce the demand discrepancy between the real-time DTA simulator and the real-world system. Section 6.2 begins with an overview of the recursive prediction-correction procedure in the OD demand estimation and prediction system. A predictive linear quadratic tracking model and a reactive adjustment model are presented in Sections 6.3 and 6.4 to correct demand estimation errors in the DTA simulator. Section 6.5 proposes a spectrum of solution algorithms along with implementation strategies to design a robust and efficient real-time demand adjustment controller.

6.2 Recursive Demand Prediction-Correction Mechanism

This section uses a Kalman filtering framework to illustrate the prediction-correction methodology in real-time DTA and to classify possible error sources in the proposed OD demand estimation and prediction models. First, let k be a subscript for OD demand estimation stages, and let X_k and Y_k be the state variable vector and measurement vector at stage k, respectively. In addition, H_k represents the measurement matrix that relates measurement Y_k and state X_k , and A_k represents the transition matrix that maps the state variables from stage k to k+1. The Kalman filter used in OD demand estimation and prediction relies on a transition equation to describe demand evolution dynamics

$$X_{k+1} = A_k X_k + w_k \tag{6.1}$$

and a measurement equation to link OD demand states to traffic measurements

$$Y_k = H_k X_k + v_k \tag{6.2}$$

where w_k~Normal(0, W_k) and v_k~Normal(0, V_k) are process noise and measurement noise at stage k, respectively.

Based on the demand transition matrix, the dynamic OD demand state at stage k is predicted from stage k-1 by propagating the mean

$$\hat{X}_{k,k-1} = A_k \hat{X}_{k-1,k-1} \tag{6.3}$$

and covariance estimates

$$P_{k,k-1} = A_k P_{k-1,k-1} A_K^T + W_k \,, \tag{6.4}$$

where $\hat{X}_{k,k-1}$ = prediction of X_k using observations up to stage k-1,

$\hat{X}_{k,k}$ = estimation of X_k using observations up to stage k,

$P_{k,k-1}$ = predicted state variance covariance matrix of X_k at stage k-1,i.e. $Var\left(X_k - \hat{X}_{k,k-1}\right)$,

$P_{k,k}$ = estimated state variance covariance matrix of X_k at stage k, i.e.$Var\left(X_k - \hat{X}_{k,k}\right)$.

After receiving new traffic observations obtained at estimation stage k, the Kalman filter adds new information of measurement Y_k into the *a posteriori* estimates of OD trip demand. Specifically, the mean of OD demand is updated by adding a correction term in proportion to deviations between measurement Y_k and predicted system output $H_k \hat{X}_{k,k-1}$, i.e.,

$$\hat{X}_{k,k} = \hat{X}_{k,k-1} + K_k(Y_k - H_k \hat{X}_{k,k-1}), \tag{6.5}$$

where the weighting matrix is

$$K_k = P_{k,k-1} H_k^T (H_k P_{k,k-1} H_k^T + V_k)^{-1}. \tag{6.6}$$

The variance of the OD demand estimate is updated according to

$$P_{k,k} = P_{k,k-1} - K_k H_k P_{k,k-1}. \tag{6.7}$$

Equation (6.7) clearly indicates, from the *a priori* demand estimate to the *a posteriori* demand estimate, the uncertainty in the OD demand estimator is reduced by $K_k H_k P_{k,k-1}$.

When the real-time DTA simulator starts at stage k, it can only use available OD prediction $\hat{X}_{k,k-1}$ as the demand input, indicating that there exists demand prediction errors $\hat{X}_{k,k} - \hat{X}_{k,k-1}$. In order to correct the errors in the *a priori* estimate, a demand updating formula equivalent to Equation (6.5) should be applied to the real-

time DTA simulator, after measurement Y_k is collected from the traffic surveillance system. Without this correction step, the demand prediction errors would propagate to the subsequent network state prediction through the simulation-assignment process. Since the real-time OD demand estimator relies on the DTA simulator to provide measurement matrix H_k (i.e. link proportions) in updating equations (6.5) and (6.7), an inconsistent internal representation of the DTA simulator could in turn make the OD estimator and predictor gradually diverge from the real-world demand process.

One brute force strategy of maintaining system consistency is to backtrack the simulator to the start time of stage *k*, and to re-simulate the network traffic conditions using the *a posteriori* estimate $\hat{X}_{k,k}$. If the state variable vector X_k also covers the demand flows departing from previous stages, the DTA simulator should be further backtracked to an earlier stage so as to truly reflect the demand update $\hat{X}_{k,k} - \hat{X}_{k,k-1}$ for lagged OD demand. Obviously, this computationally intensive strategy is not suitable for a real-time DTA implementation, so the following analysis focuses on the an OD demand adjustment method that does not require state resetting.

6.3 Predictive OD Demand Adjustment Model

As shown in Figure 6-1, OD demand consistency checking and updating in this study is modeled as a feedback controller. For the DTA traffic simulator, i.e. the plant to be controlled, the control input is the adjusted demand from the consistency checking and updating module; and the system output is the simulated network flow pattern. Taking the real-world measurements as the control reference, the controller seeks to (1) reduce deviations between the simulated states and the real-world

observations and (2) keep the adjustment magnitude as small as possible. Similar to the commonly used minimum energy criterion in the optimal control, the second objective is intended to avoid unsteady responses of the DTA simulator due to dramatic demand adjustments.

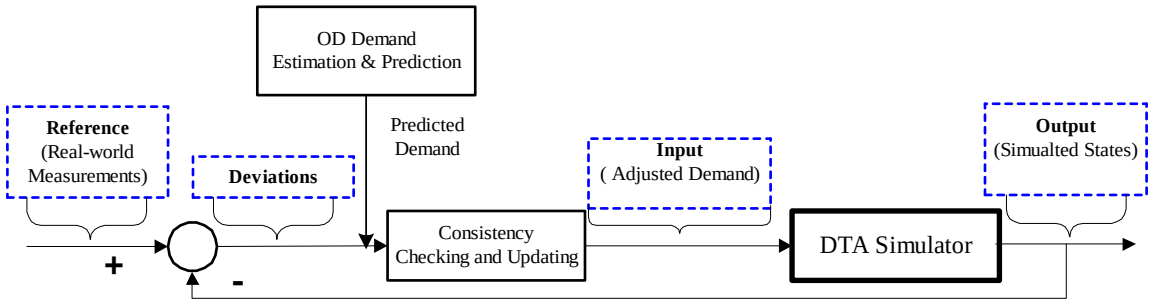

FIGURE 6-1 Feedback control model for OD demand adjustment.

The controller can be formulated using a 1-step look-ahead linear quadratic control model.

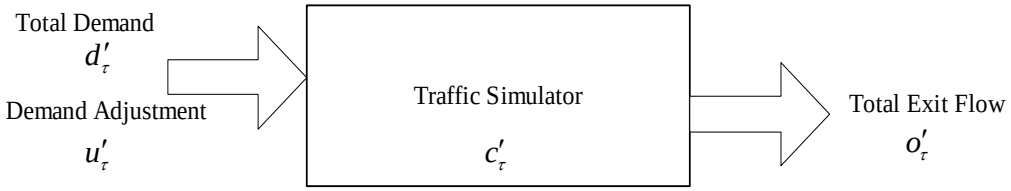

FIGURE 6-2 Conceptual state space representation of traffic simulator.

The state space representation of the DTA simulator is simplified using a liquid tank model as shown in Figure 6-2. Let τ be subscript for departure time intervals. Let c_τ and c'_τ represent total numbers of vehicles in the real-world system and in the traffic simulator, respectively, at departure time interval τ. Furthermore, let d'_τ and u'_τ denote total predicted demand volume and total demand adjustment,

respectively, at departure time interval τ. In the following linear quadratic tracking model, the total number of vehicles in the simulation system is considered as the state variable.

$$J_\tau = (c_{\tau+1} - c'_{\tau+1})^2 + r(u'_\tau)^2 \quad (6.8)$$

where r = weight for the demand adjustment magnitude.

The transition equation is give by

$$c'_{\tau+1} = c'_\tau + d'_\tau + u'_\tau - o'_\tau \quad (6.9)$$

where o'_τ = total number of vehicles exiting from the traffic simulator at departure time interval τ.

Performance index J_τ simultaneously minimizes the state deviation at the next departure time interval and adjustment magnitude. In the above simplified single-input single-output system, the system transition equation shows that the total number of vehicles remaining in the simulator at departure time interval τ+1 changes by the addition of the demand input (i.e. predicted demand d'_τ plus demand adjustment u'_τ) minus the total exit flow o'_τ at departure time interval τ. This state space representation explicitly considers the intermediate effects of the adjustment implemented at departure time interval τ on the future state at departure time interval τ+1. As a result, the above predictive control strategy can compute manipulated variable adjustments to optimize the future performance of the plant (i.e. DTA simulator). However, this predictive control procedure has one critical shortcoming that prevents its application in real-time settings. That is, the future state of the real-world traffic system $c_{\tau+1}$ has not been observed at departure time interval τ, so the

adjustment u'_τ cannot be determined as a function of $c_{\tau+1}$. An alternative strategy is to use a predicted value of $c_{\tau+1}$ to generate the reference point for this controller, but the resulting prediction errors might dramatically degrade the actual adjustment performance.

6.4 Reactive OD Demand Adjustment Model

To circumvent the above-mentioned difficulty in the predictive model, a reactive approach is adopted in this study to optimize OD demand adjustments for *prevailing* state deviations. Let e_τ denote the deviation between the total number of vehicles in real-world system and that in the simulator at departure time interval τ. The reactive adjustment model assumes that

$$e_{\tau+1} = e_\tau - u'_\tau, \tag{6.10}$$

meaning that the state deviation at next time interval $\tau+1$ is changed by the addition of the demand adjustment at departure time interval τ. Accordingly, a new performance index is

$$\underset{u'_\tau}{Min}(e_\tau - u'_\tau)^2 + r(u'_\tau)^2 \tag{6.11}$$

with a closed form solution as

$$u_\tau = \frac{1}{1+r} e_\tau. \tag{6.12}$$

Without considering the impact of predicted demand d'_τ and existing flow o'_τ, the system transition equation (6.10) focuses on system state changes due to demand adjustment u'_τ so calculated adjustments are only suitable for a short time interval. Essentially, the reactive control strategy shown in Equation (6.12) follows the closed

loop feedback law: if the current number of vehicles in the simulator is lower (higher) than the observed measures, corresponding to the positive (negative) deviation term e_τ, then positive (negative) adjustment u'_τ is applied to the demand input.

As the above univariate liquid tank model is only an extremely simplified representation of the complex DTA traffic simulation system, the following further proposes a detailed demand prediction feedback adjustment model to deal with a realistic traffic network, which consists of multiple origins and destinations as well as a set of nodes connected by a set of directed links. If both origin and destination of each vehicle are observed by point-to-point sensors, then the unique state of OD demand flows is determined and the corresponding errors in the demand representation can be explicitly identified. However, for most applications, only limited point detectors are available on a subset of links, resulting in inability to reveal the true origin destination states. Alternatively, link density is selected as the reference measure in the proposed model.

Using link proportions to describe relationships between OD demand adjustments and changes in link density, the proposed optimization problem seeks to minimize (1) deviations between the simulated link density and the real-world link density and (2) demand adjustment magnitude.

$$\underset{u'}{Min} \quad \sum_{l,t}\left[e_{(l,t)} - \sum_{i,j}(\hat{p}_{(l,t)(i,j)}u'_{(i,j)})\right]^2 + r_{(i,j)}(u'_{(i,j)})^2 \tag{6.13}$$

Subject to

$$d'_{(i,j)} + u'_{(i,j)} = 0. \tag{6.14}$$

where l = subscript for link with traffic measurements, l=1,…,m,

t = subscript for observation time interval in an OD adjustment period, t=1, 2, …, T,

(i, j) = subscript for origin-destination pair, (i, j) =1,…, n_{od},

$c_{(l,t)}$ = number of vehicles measured on link l, during observation interval t,

$c'_{(l,t)}$ = number of vehicles simulated on link l, during observation interval t,

$e_{(l,t)}$ = deviation between $c_{(l,t)}$ and $c'_{(l,t)}$,

$\hat{p}_{(l,t)(i,j)}$ = link proportion in the DTA simulator, i.e. the proportion of vehicular demand flows from origin i to destination j, contributing to the number of vehicles on link l during observation interval t,

$u'_{(i,j)}$ = demand adjustment from origin i to destination j to be applied in the DTA simulator,

$d_{(i,j)}$ = demand volume of OD pair (i, j) during the current adjustment period,

$r_{(i,j)}$ = positive weight for adjustment magnitude of OD pair (i, j).

The above optimization problem is a linear quadratic model with inequality constraints, and the positive link proportion matrix and positive diagonal weighting matrix $[r_{i,j}]$ can guarantee the positive definiteness of the optimization problem and the existence of a unique solution. In other words, the combined weighting objective function successfully overcomes a possible rank deficient problem posed by limited traffic measurements during an adjustment time interval. If ignoring the nonnegativity constraint, the unconstrained problem has a closed form solution:

$$u' = \left(P^T P + R\right)^{-1} P^T e \qquad (6.15)$$

where R = the weighting matrix on demand adjustment, i.e. $n_{od} \times n_{od}$ diagonal matrix consisting of elements $r_{(i,j)}$,

P = link proportion matrix consisting of elements $\hat{p}_{(l,t)(i,j)}$,

e = deviation vector of link density,

u' = demand adjustment vector.

The regulation term in the objective function limits the magnitude of adjustment and consequently reduces the chance of violating the nonnegativity constraint. The constrained problem can be solved by the Lagrangian method. Alternatively, one can first solve the unconstrained problem and ensure the nonnegativity constraints by resetting $u'_{(i,j)} = - d'_{(i,j)}$ for any demand pair such that $d'_{(i,j)} + u'_{(i,j)} <0$. In essence, the proposed OD demand updating function (6.15) and the OD demand estimation function (6.5) in the Kalman filter share a similar feedback function form, that is, the adjustment magnitude u' and $\hat{X}_{k,k} - \hat{X}_{k,k-1}$ are proportional to gain factors $\left(P^T P + R\right)^{-1} P^T$ and error term, respectively. On the other hand, the proposed OD demand adjustment model can be viewed as a simple least squares optimization program, where the Kalman filter is a least squares estimator that fully considers the second order noise statistics.

6.5 Efficient Algorithms and Implementation Issues

Integrated in the real-time DTA system, the OD demand adjustment module operates at every updating time interval as follows.

(1) Receive real-world observations from the surveillance system,

(2) Measure deviations between the real-world observations and the internal states of the DTA simulator,

(3) Obtain link proportions from the DTA simulator,

(4) Solve the proposed optimization problem and calculate appropriate demand adjustments,

(5) Feed the demand adjustment into the DTA simulator.

The subsequent discussion is devoted to the development of tractable and computationally efficient solution algorithms. The closed form solution (6.15) requires inverting a (n_{od}×n_{od}) matrix, and the complexity of a direct matrix inversion is O(n_{od}^3). For a short updating period, the number of observations (m×T) is typically less than the number of OD pairs to be updated. For example, at a 5-minute updating time interval, the ratio of the number of observations over the number of OD demand pairs is only 14.6% in the Irvine test bed network. In this case, a recursive updating scheme without involving direct matrix inversions can reduce the computational time dramatically. The recursive least squares algorithm is first presented here, which has been widely used in the field of Kalman filtering (Chui and Chen, 1991). Its each iteration requires $2(n_{od})^2+3(n_{od})$ multiplications, leading to the time complexity of the algorithm is $O(n_{od})^2$×m×T, In addition, storing the variance matrix Φ(n) requires $O(n_{od})^2$ memory space.

Notation

n= observation index at each adjustment period, n =1, 2, …, m×T

$e(n)$ = deviation in terms of the number of vehicles based on observation n

$u'(n)$ = vector of demand adjustment taking into account observations 1, 2, …, n

$p(n)$ = vector of link proportions related to observation n

$g(n)$ = gain vector related to observation n

$\Phi(n)$ = error correlation matrix considering observations 1, 2, …, n

$\alpha(n)$ = step size scalar related to observation n

Note that, $u'(n)$, $p(n)$ and $g(n)$ are column vectors with n_{od} elements.

Algorithm 1: Recursive least squares algorithm

Initialization

n=0; $u'(n)$=0

$$\Phi(n)= diag[\frac{1}{r(1)}, \frac{1}{r(2)}, \cdots, \frac{1}{r(n_{od})}]$$

Main loop

For each measurement n = 1 to m×T,

(1) Compute the gain vector

$$g(n) = \frac{\Phi(n-1)p(n)}{1 + p^T(n)\Phi(n-1)p(n)} \tag{6.16}$$

(2) Update the estimate of the adjustment vector

$$u'(n) = u'(n-1) - g(n)[p^T(n)u'(n-1) - e(n)] \tag{6.17}$$

(3) Update the error correlation matrix

$$\Phi(n) = \Phi(n-1) - g(n)p^T(n)\Phi(n-1) \tag{6.18}$$

End Loop

Although the recursive least algorithm is able to reduce the complexity from $O(n_{od})^3$ to $O(n_{od})^2$, calculating and maintaining a huge variance covariance matrix

Φ(n) is both time and memory demanding, especially for a real-time feedback controller with hundreds of state variables. A simple sub-optimal algorithm with stable performance is preferred in this context. Simplifying time-varying matrix Φ(n) to a constant scalar $\phi(\mathrm{n}) = \frac{1}{\mathrm{r}}$ yields

$$\mathrm{g}(n) = \frac{\phi(n-1)\mathrm{p}(n)}{1+\phi(n-1)\mathrm{p}^T(n)\mathrm{p}(n)} = \frac{\mathrm{p}(n)}{\frac{1}{\phi(n-1)}+\mathrm{p}^T(n)\mathrm{p}(n)} = \frac{\mathrm{p}(n)}{r+\mathrm{p}^T(n)\mathrm{p}(n)}. \qquad (6.19)$$

This leads to the following normalized incremental gradient algorithm, which is extensively used in the field of adaptive filtering and neural networks. Each iteration of this algorithm only requires $3\times n_{od}$ multiplications and n_{od} memory space. Note that, a positive step size term should be added for this gradient algorithm, as the above gain factor is an approximation for the optimal search direction.

Algorithm 2: Normalized incremental gradient algorithm

Initialization

n=0; $u'(n)$=0

Main Loop

For each measurement n = 1 to m×T,

(1) Compute the gain vector

$$\mathrm{g}(n) = \frac{\mathrm{p}(n)}{r+\mathrm{p}^T(n)\mathrm{p}(n)} \qquad (6.20)$$

(2) Update the estimate of the adjustment vector

$$u'(n) = u'(n-1) - \alpha(\mathrm{n})\mathrm{g}(n)[\mathrm{p}^T(n)u'(n-1) - \mathrm{e}(\mathrm{n})] \qquad (6.21)$$

End Loop

In the above algorithm, the demand adjustment for each OD pair is proportional to $p^T(n)p(n)$. If ignoring this product term, then one can obtain a standard incremental gradient algorithm, where $g(n) = \frac{1}{r} p(n)$. This simplified algorithm only requires $2 \times n_{od} \times m \times T$ multiplications and n_{od} memory space. Compared to the optimal recursive least squares algorithm, both sub-optimal algorithms are substantially faster with less memory requirements.

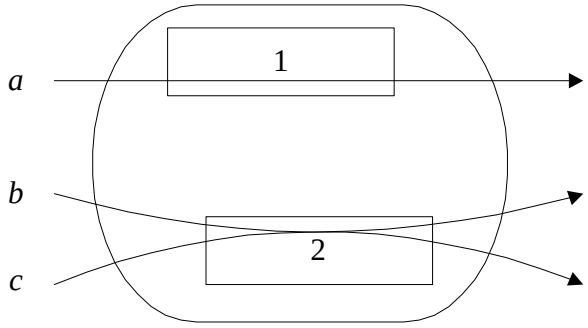

FIGURE 6-3 Idealized network for illustrating the advantage of the normalization

Principally, the normalized gain factor is able to recognize the impact of overlapping OD pairs in the OD demand adjustment process. Its advantage can be illustrated using a simplified network shown in Figure 6-3, where vehicular flows along OD pair *a* pass through link 1 and vehicular flows along OD pairs *b* and *c* pass through link 2. Further assume the link proportions from OD pair *a* to link 1, from OD pair *b* to link 2, and from OD pair *c* to link 2 are all 100%. If the same deviations are observed on links 1 and 2, i.e. e(1)=e(2), the standard incremental gradient algorithm would suggest that the OD demand adjustments for three OD pairs have the same magnitude, leading to possible over-adjustments for traffic states on link 2.

Since product terms $p^T(1)p(1)=1$ and $p^T(2)p(2)= 2$, according to the normalized incremental algorithm, the individual demand correction for overlapping OD pairs *b* and *c* would be only half of the correction for independent OD pair *a*.

As shown by Bertsekas (1999), the above three algorithms are variants of the incremental gradient method to solve the sequential linear quadratic problem. These algorithms process the data blocks in sequence, and they mainly differ in terms of the calculation related to the gain matrix (factor). In addition, the recursive least squares algorithm only needs a single pass through the entire data to reach the optimum, where two sub-optimal algorithms might require multiple passes to find the minimum, although a single pass typically produces a very dramatic decrease in the value of the objective function. Because the link proportion matrix in the adjustment formulation is still a simplified representation of the input-output relation for the DTA simulator, it is difficult to make the simulator completely reach the desired state even using an "optimal" adjustment based on the proposed mathematical model. In fact, computational efficiency is a more important consideration. A desirable real-time demand regulator should require substantially small computation efforts while maintaining adjustment performance. In addition, for a stable feedback controller in a complex and dynamic real-time environment, the weighting matrix (factor) *R* must be properly tuned to obtain a desired response to a given disturbance. A small *R* indicates a large gain, quick response and small errors when a disturbance occurs. On the contrary, the choice of a large *R* will result in great errors under disturbances, but tends to be stable when the process structure changes. Experiments are also needed to select an appropriate demand updating frequency, which is jointly determined by the

execution time for retrieving link proportions from the DTA simulator and computational efficiency of the demand adjustment algorithm.

6.6 Summary

To maintain an internally consistent representation with actual traffic conditions, this section presents an OD demand consistency checking and updating model for on-line dynamic traffic assignment operation. Both predictive and reactive approaches are proposed to minimize (1) the deviations between simulated states and real-world observations and (2) OD demand adjustment magnitude. The two objectives are combined into a weighted linear quadratic function to construct a guaranteed over-determined optimization problem. Alternative recursive solution algorithms are presented to design an efficient feedback controller that regulates the demand input for the real-time dynamic traffic assignment simulator.

7. Experimental Analysis of Dynamic OD Demand Estimation and Prediction Methods

7.1 Introduction

After presenting the OD demand estimation and prediction mathematical models for on-line applications, this chapter aims to evaluate the performance of the proposed models under different degrees of information availability and for alternative formulation strategies. Since the dynamic OD demand estimation and prediction process in this study relies on a simulation-based dynamic assignment model to describe the spatial and temporal interactions of travelers, the resulting complexity precludes the analytical derivation of the properties and the performance of the demand estimators in a general network. Alternatively, numerical experiments are conducted to achieve the following objectives: (1) illustrate the effectiveness of the proposed methods using available real-world measurements in a real network, (2) quantify the relative benefits of additional information and model enhancement for OD demand estimation and prediction, and (3) perform sensitivity analysis of the consistency checking algorithm under different parameter settings. The results are intended to provide insight on effective real-time demand estimation and prediction and traffic data collection schemes.

The chapter is organized as follows. Section 7.2 evaluates the performance of on-line OD demand estimation and prediction components under various model structures, and Section 7.3 tests alternative demand consistency checking and

updating algorithms and implementation strategies. This chapter is concluded by a summary of the analysis results in Section 7.4.

7.2 Real-Time OD Demand Estimation and Prediction

This section is intended to demonstrate the effectiveness of the dynamic OD estimation and prediction models for on-line DTA applications, and to test the performance of alternative model structures.

7.2.1 Network Configuration and System Settings

The numerical experiments in this section are conducted based on the detailed Irvine test bed traffic network, as shown in Figure 7-1. This network includes 61 OD demand zones, 326 nodes and 626 links, where traffic counts are measured at 30-second intervals on 19 freeway links and at 5-minute interval on 28 arterial links.

Table 7-1 defined the data sampling frequency, execution frequency and cycle length for different components in the real-time DTA system. All the experiments are performed on a PC with 2 GHz CPU and 4 GB memory, and all the algorithms are implemented in Visual Fortran and Visual C++ on the Windows platform.

Table 7-1 System Execution Parameters.

Parameter	Value
Observation interval for OD demand estimation	0.5 min
Observation interval for consistency checking and updating	0.5 min
RT-DYNA roll period	0.5 min
P-DYNA roll period	5 min
P-DYNA prediction horizon	20 min

ODE estimate time interval	5 min
ODP execution cycle	10 min
ODP prediction horizon	45 min
Long term consistency checking cycle	5 min
Short term consistency checking cycle	0.5 min

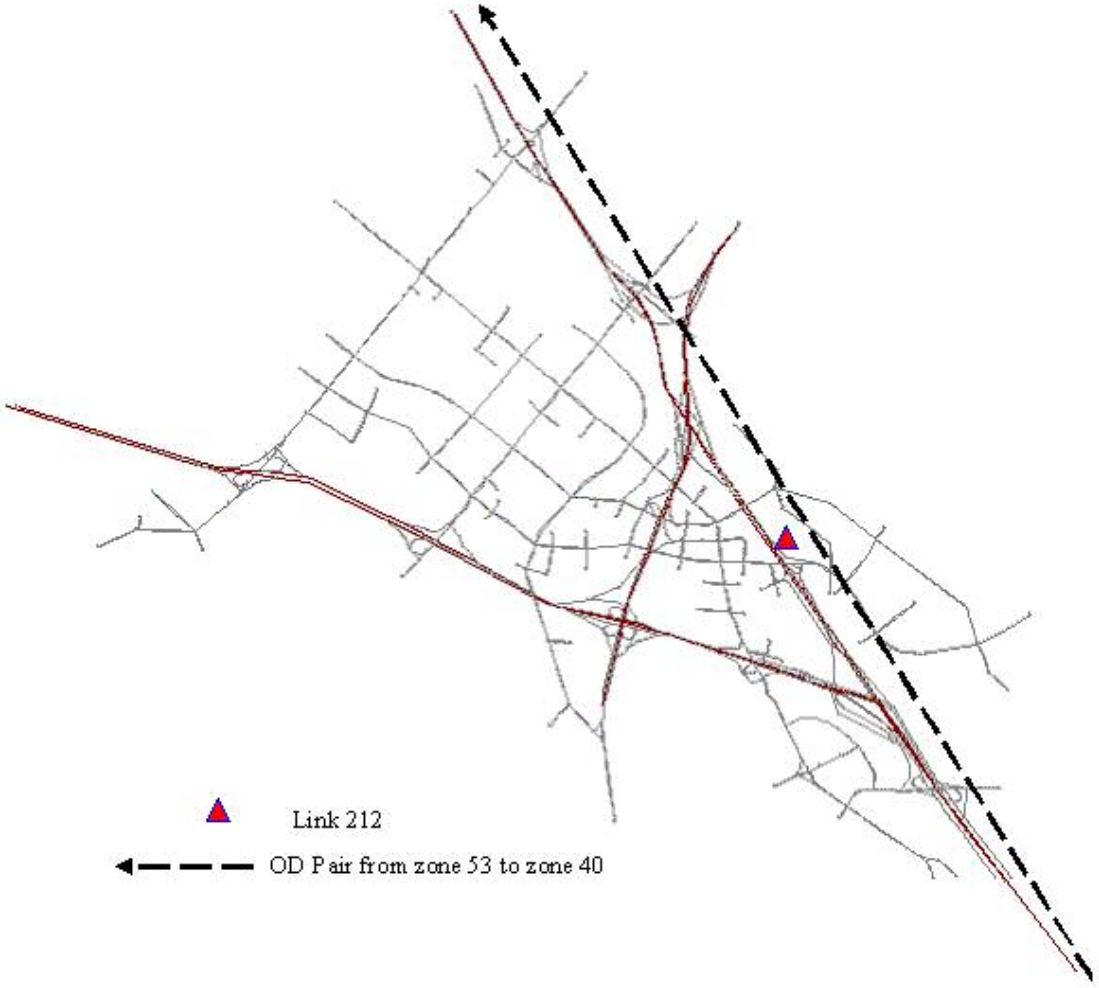

FIGURE 7-1 Detailed Irvine network.

7.2.2 *A priori* Estimation vs. Real-time Estimation

The *a priori* estimate of the regular demand pattern is constructed by the off-line OD estimation method using the first day data. Real-world observations on the second day are used to calibrate the system and measurement variances in the real-time OD estimation and prediction model. The third day data are used to validate the proposed real-time OD estimation and prediction algorithm. The time of interest in the following experiments is the morning peak period (4:00 AM – 10:00 AM).

The *a priori* estimation results are generated by loading the *a priori* OD estimate (from the first day) into the traffic network, while on-line estimation refers to real-time traffic assignment results using real-time OD estimates (on the third day). Comparing simulated results and real-world observations, Table 7-2 summarizes the root mean squared errors of density, volume and speed for observed links in the network. In particular, link density and speed measures are processed at 1-minute time intervals, while link volume is processed based on 15-minute time intervals to obtain reliable samples. The consistent error reduction in these three major traffic measures clearly demonstrates that the on-line estimator is able to utilize the real-time information to improve the final quality of network state estimation and prediction.

Table 7-2 Average RMSE in on-line estimation vs. *A Priori* estimation.

	A priori Estimation	*Online Estimation*	*Percentage Improvement*
Density	11.6	10.5	9.5%
Volume	288.8	208.5	27.8%
Speed	16.7	14.1	25.6%

Taking link 1 in the network as a specific example, Figures 7-3 and 7-3 further illustrate the difference between the *a priori* estimate and real-time estimate in terms of link density. The off-line simulation result generated from the *a priori* demand estimate with the first day observations does not capture the morning traffic peak between 6 AM and 7 AM on the third day, while the on-line estimator recognizes the day-to-day demand changes from the real-time measurements.

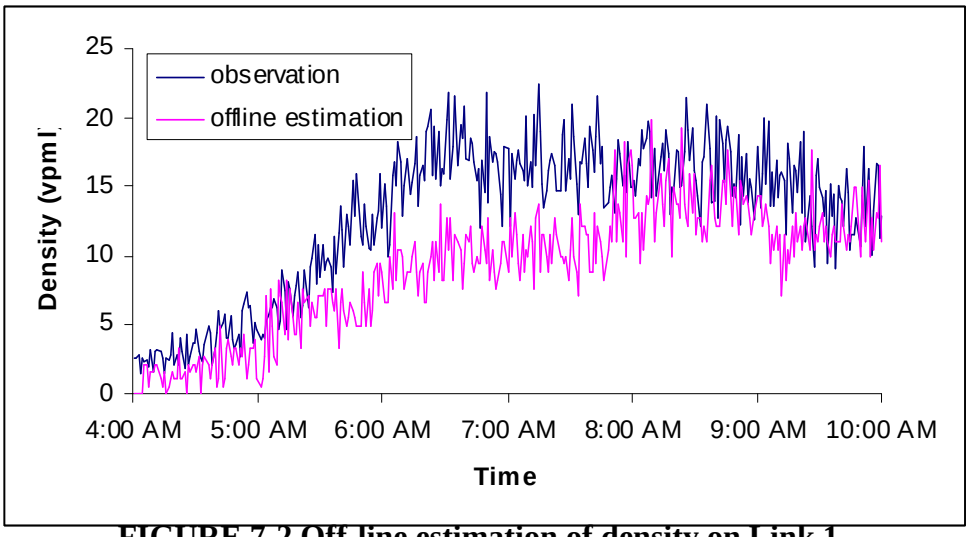

FIGURE 7-2 Off-line estimation of density on Link 1

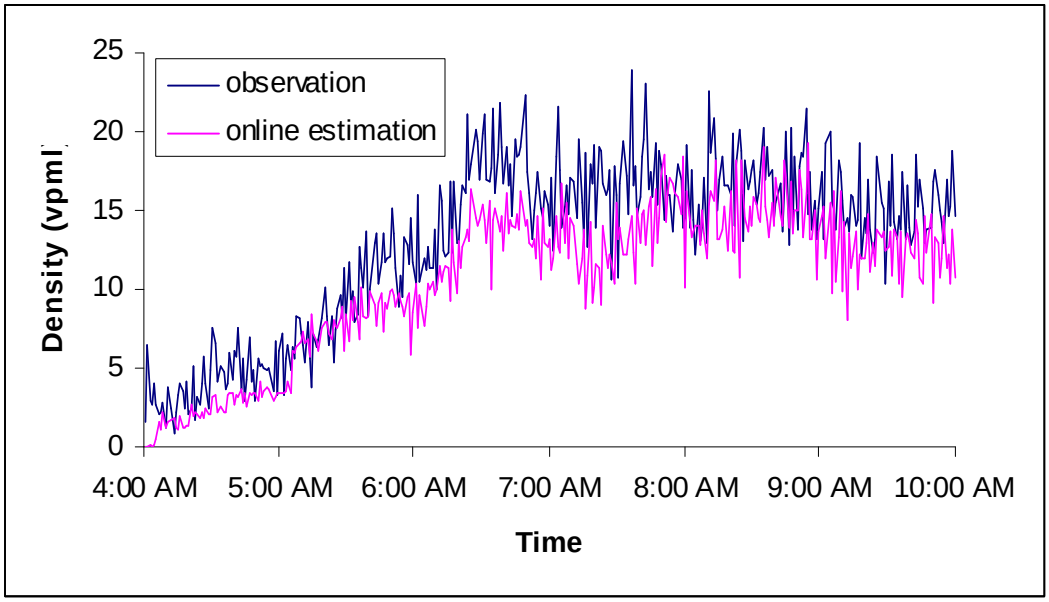

FIGURE 7-3 Online estimation of density on Link 1.

7.2.3 Polynomial Model

First, a second-order polynomial trend model (i.e. local linear trend model) is applied to estimate the demand deviations from the *a priori* estimate of the regular demand pattern. In Figure 7-4, the *a priori* regular pattern estimate and the corresponding demand deviations are displayed for the OD pair from zone 53 to zone 40, which carries the highest trips among all the OD pairs in the study network. On average, the *a priori* demand data underestimates the real-time demand on the third day for this selected OD pair, but the prior estimate still represents similar time-varying dynamic patterns as the real-time demand. As expected, the demand deviations exhibit a much slower changing pattern than the corresponding real-time demand flows over the same time. As discussed previously, the day-to-day dynamics is one of the major causes of the estimated structural demand deviations. In additional, the estimation noise in the *a priori* demand data, which only utilizes one-day observations, may have contributed to the deviations shown in this case.

Considering the smooth trend for demand deviations in Figure 7-4, it is desirable to further reduce the second-order polynomial model to the first-order polynomial model. To compare the estimation performance of alternative models, the root mean squared error in density is defined as

$$RMSE_t = \sqrt{\frac{\sum_i (c_{i,t} - \hat{c}_{i,t})^2}{n_{obs}}}$$

Where $c_{(i,t)}$ = density measured on link *i*, during observation interval *t*,

$\hat{c}_{(i,t)}$ = simulated density from the real-time DTA estimator on link *i*, during observation interval *t*,

n_{obs}= number of observations.

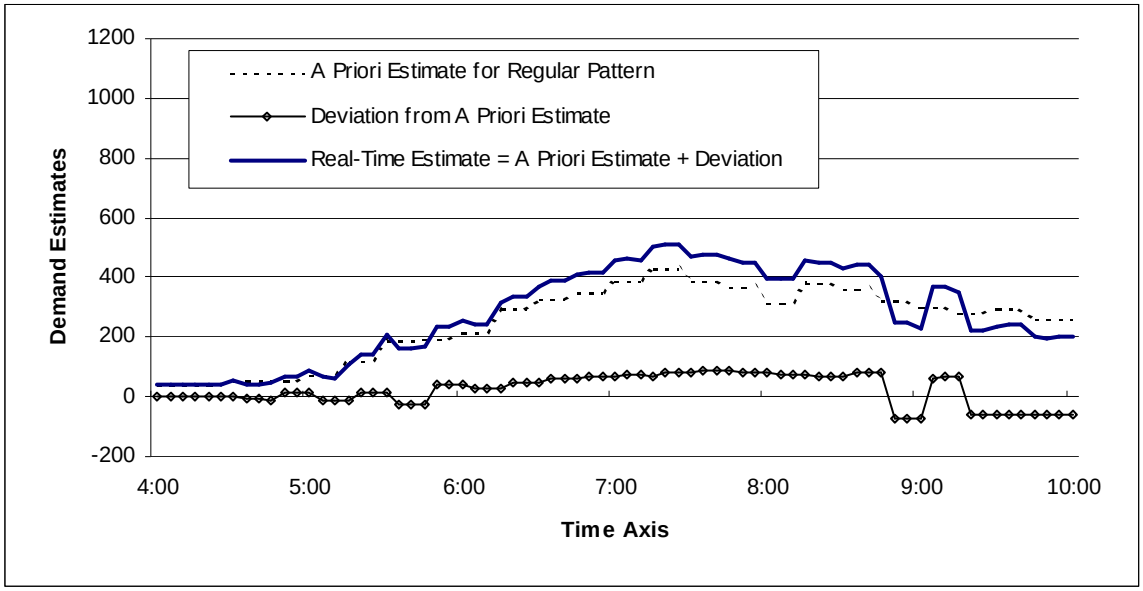

FIGURE 7-4 Dynamic demand estimates for OD pair from zone 53 to zone 40.

The RMSE at every 5 minutes are plotted in Figure 7-5, for the first-order and second-order polynomial models, respectively. The average RMSE of the first-order model during the study horizon (10.2608) is marginally greater than the average RMSE of the second-order model (10.8588) by 1.8%. These two models produce smaller estimation errors in the early morning (from 4:00 AM to 6:00 AM), compared to the peak hour period (from 7:00 AM to 9:00 AM). Such large estimation errors around 8 AM can be explained by the increasing demand variability and significant dynamics in the underlying traffic flow propagation process during the peak hour period. Based on experiment results from the third day data, the first-order polynomial model seems to be more attractive than the first-order model, since it

offers acceptable accuracy with considerable computation efficiency. On the other hand, if real-time response constraints can be satisfied, keeping a higher order polynomial model is always preferable, because it is more capable of capturing nonlinearity in the demand structure changes.

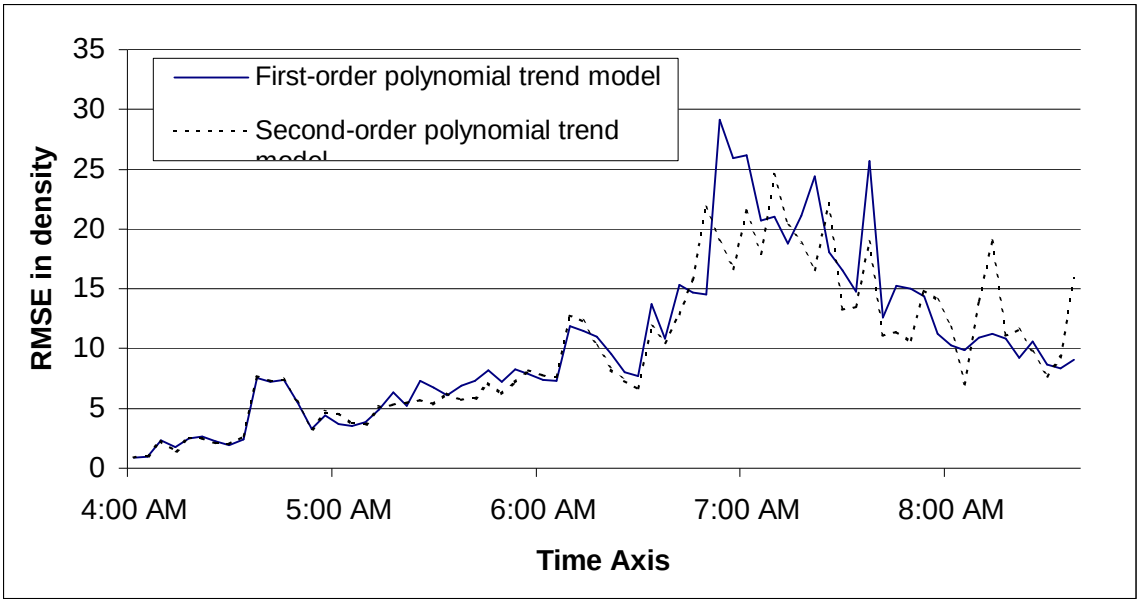

FIGURE 7-5 RMSE in density for the first-order and second polynomial models.

Figure 7-6 plots simulated density, predicted density and the observed density on link 212, using a 20-minute prediction horizon. Link 212 is a freeway link going northbound, and its location is marked in Figure 7-1. The simulated density and the predicted density are generated from the DTA network state estimation module and the DTA state prediction module, respectively. As revealed by the graph, the DTA network state estimator is able to capture the time-varying trends of real-world traffic, while the DTA network state predictor can forecast dynamic flow propagation with

acceptable quality. The above results further validate the effectiveness of the proposed real-time OD estimation and prediction framework.

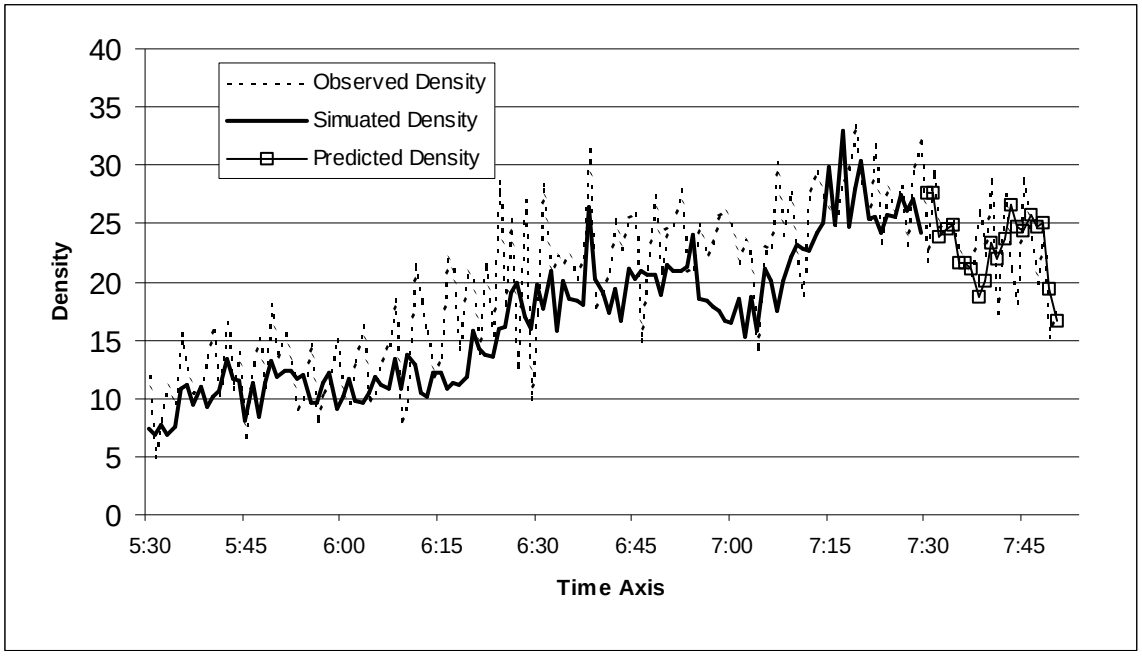

FIGURE 7-6 Observed density vs. simulated density and predicted density on link 212

7.3 Online OD Demand Consistency Checking and Updating

This section is intended to evaluate the performance of different OD demand adjustment algorithms proposed in Chapter 6. In addition, it also conducts sensitivity analysis for the key system parameters, such as the step size for sub-optimal algorithms and updating frequency.

7.3.1 Description of the Test Network and Data Set

The experiments in this section are conducted based on the Irvine test bed network shown in Figure 7-1. The root mean square error in terms of number of vehicles on links with observations is used as the consistency measure.

$$RMSE = \sqrt{\frac{\sum_{t}\sum_{l}(c'_{l,t} - c_{l,t})^2}{n_{obs}}}$$

where $c_{(l,t)}$ = number of vehicles measured on link l, during observation interval t,

$c'_{(l,t)}$ = number of vehicles simulated on link l, during observation interval t,

nod_s= number of observations.

To specifically examine the influence of demand consistency checking on the system performance, the short-term (flow propagation) consistency checking module is not activated in the following experiments.

7.3.2 Experimental Design and Results

1. **Computational performance**

The first task is to compare the computational performance of the three consistency solution algorithms. Table 7-3 lists average execution time and estimation errors for these three algorithms at a 1-min updating interval. Obviously, the efficient sub-optimal approach is suitable for real-time operation, and the exact solution algorithm does not satisfy the real-time response constraint for the medium-scale network considered in the study. However, the optimal adjustment solution result can still serve as a useful benchmark for comparing the performance of sub-optimal algorithms. The percentage improvements shown in Table 1 are the relative

error reduction compared to the do-nothing case (i.e. without OD demand adjustment). Overall, both sub-optimal algorithms can dramatically reduce the estimation errors, and the normalized incremental algorithm outperforms the standard algorithm, because the former can capture the overlapping effect in OD demand adjustments.

Table 7-3 Computational performances of consistency solution algorithms.

	Avg. computational time (sec)	Avg. RMSE	Error reduction
Recursive least square algorithm	192.830	18.17	44.8%
Normalized incremental gradient	0.829	23.99	28.4%
Standard incremental gradient	1.131	27.12	19.1%

Figure 7-7 details the time-varying performance for different consistency checking algorithms. From 4:00 AM to 6:30 AM, the benefit of all three adjustment algorithms is insignificant, as the overall OD demand level remains steady and OD prediction errors are relatively small. After 6:30 AM, the estimation errors in the do-nothing case increase sharply, which can be attributed to high OD prediction errors during the peak period. Between 6:30 AM and 7:30 AM, all three algorithms are able to reduce the system inconsistency, and the optimal solution procedure clearly generates a lower bound of error reduction relative to the other two sub-optimal regulators. During the off-peak period (7:30 AM to 8:30 AM), the deviations between the real-time simulator and real-world system cannot be significantly reduced by any OD demand consistency checking algorithm. By carefully comparing the simulation results with the real-world observations on a link-by-link basis, the author finds that the remaining large state deviations are more likely due to incorrect route choice

prediction or incorrect traffic flow modeling. As only the OD demand consistency checking module is activated in this study, the system might attribute all the state inconsistency to the OD demand prediction errors, leading to slightly worse results compared to the do-nothing case. This observation underscores the need for a joint consistency checking system, which should correctly recognize and correct system inconsistency caused by different error sources. After 9:00 AM, the adjustment algorithms again produce significant error reductions.

Figure 7-7 details the time-varying performance for the three consistency checking algorithms. From 4:00 AM to 6:30 AM, the benefit of all three consistency checking algorithms is insignificant, as the overall OD demand level remains steady and OD prediction errors are relatively small (The demand pattern of a major OD pair is shown in Figure 7-4). After 6:30 AM, the estimation errors in the do-nothing case increase sharply, which is due to high estimation errors associated with the OD demand predictor during the on-peak period. Between 6:45 AM and 7:30 AM, all three consistency checking algorithms are able to reduce the system inconsistency, and the optimal solution procedure clearly generates the lower bound of error reduction relative to other two sub-optimal regulators. During the off-peak period (7:30 AM to 8:30 AM), the deviations between the real-time simulator and real-world system cannot be significantly reduced by any OD demand consistency checking algorithm. Such large remaining deviations in terms of number of vehicles can be attributed to the other error sources, e.g., incorrect route choice prediction or incorrect traffic flow modeling. In this case, if attributing all the system inconsistency to the OD demand prediction errors, that is, only the OD demand consistency checking

module is activated, then such incorrect modeling could even lead to worse results compared to the do-noting case. This observation underscores the need for a joint consistency checking system, which should correctly recognize and correct system inconsistency caused by different error sources. After 9:00 AM, the recursive least squares algorithm and the normalized incremental algorithm again consistently perform well in reducing the traffic state deviations.

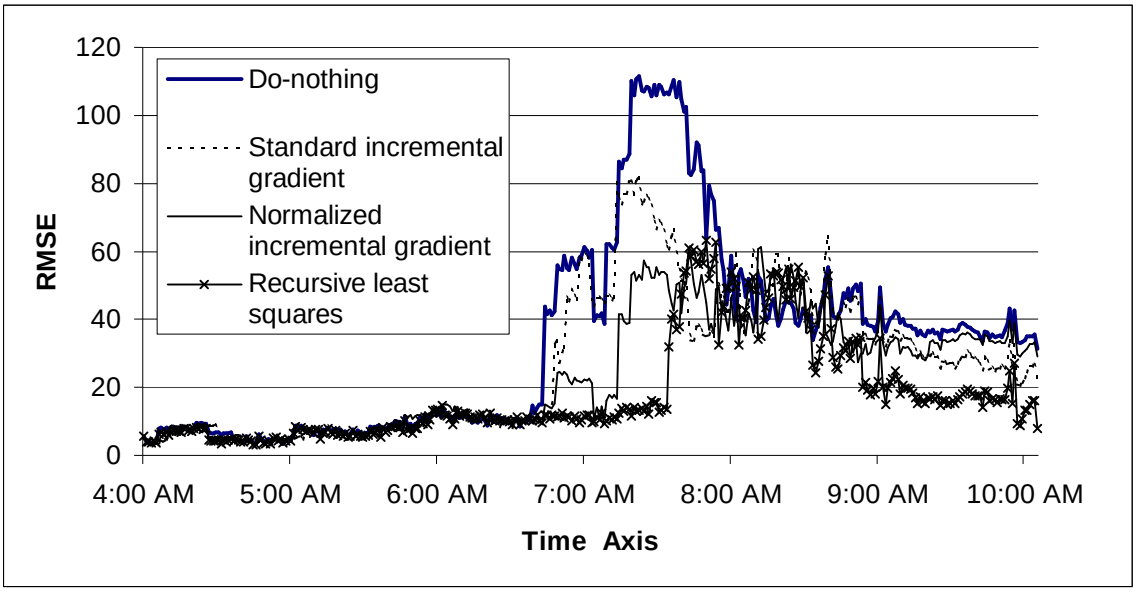

FIGURE 7-7 Performance of three consistency checking algorithms

1. **Step size in sub-optimal algorithms**

The weighting factor of the objective in the OD consistency checking formulation can be calibrated using the multi-objective programming techniques discussed in Chapter 3. The following experiments are designed to tune the step size in the two sub-optimal algorithms. The minimum estimation error for the normalized incremental gradient algorithm is obtained for a step size of 0.5. The standard

incremental gradient algorithm gains the best performance when the step size is around 0.05. The dramatic difference between these two step size parameters is due to the existence of the term $p^T(n)p(n)$ in the normalization algorithm. As expected, when the step size parameter in both sub-optimal algorithms decreases, the experiment results show that the OD demand correction magnitude tends asymptotically toward zero, and the performance improvement due to demand consistency checking and updating becomes gradually negligible.

2. Updating time interval

With different updating time intervals, Figure 7-8 depicts the performance of the normalized incremental gradient algorithm in terms of network state estimation and prediction errors. Under all three updating intervals, the sub-optimal algorithm produces dramatic error reductions for the real-time traffic state estimation. Moreover, the short (i.e. 1-min and 2.5-min) updating interval can control the near-term prediction errors within a certain range, compared to a sharp prediction error increase in the do-nothing case. This indicates that frequent updating is preferred in order to rapidly respond to possible OD demand prediction disturbances. In comparison, a longer updating interval corresponds to less frequent updates, implicitly allowing more time for OD demand prediction errors propagating in the DTA simulator. The figure clearly illustrates that, at a 5-min updating interval, the influence of error propagation becomes very significant when the prediction horizon extends from 10 minutes to 20 minutes. In this case, OD demand consistency

checking only provides marginal quality improvement for medium-term network state prediction.

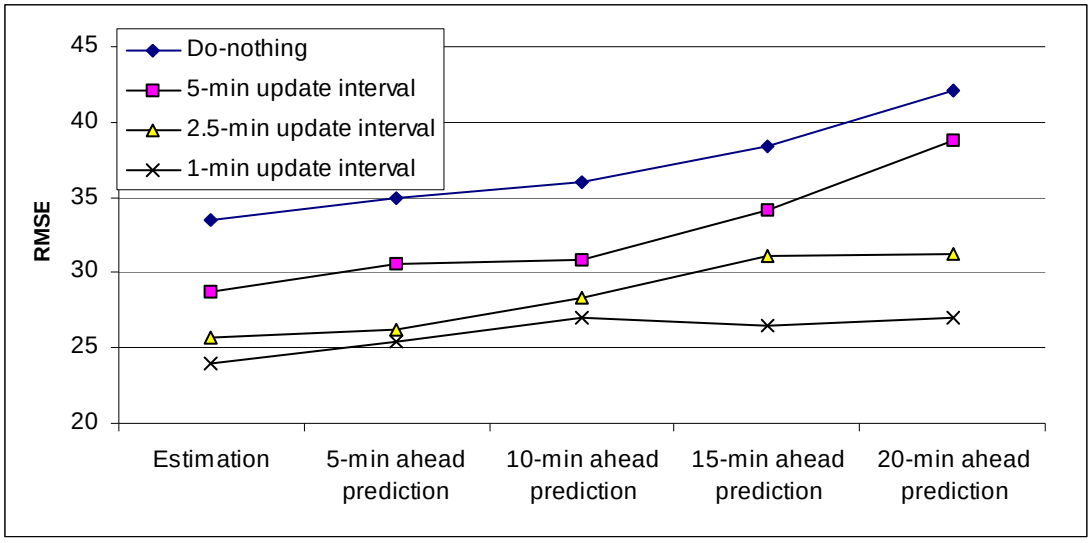

FIGURE 7-8 Estimation and prediction errors under different updating time intervals

7.4 Summary

This section uses the field data from the Irvine network to evaluate real-time OD estimation and prediction and real-time demand consistency checking algorithms. In summary, the experiments provide the following important findings for on-line DTA operations:

1. Compared to the *a priori* estimator, the on-line demand estimator is able to utilize the real-time information to significantly improve the final quality of network state estimation and prediction. The first-order polynomial model can offer acceptable OD estimation and prediction accuracy with considerable computation efficiency.

2. The exact solution algorithm does not satisfy the real-time response constraint for the given medium-scale network in the study. In contrast, the normalized incremental algorithm and standard algorithm are capable to generate on-line OD correction solutions for the real-time DTA simulator, and make significant reductions in both estimation and prediction errors, especially with a short updating interval.

8. Conclusions and Future Extensions

This chapter begins with a brief summary of the proposed formulations and findings. Section 8.2 presents the author's perspective on the contributions of this research to the state of the art of dynamic OD demand estimation and prediction. Section 8.3 discusses further extensions and directions for future research in this area.

8.1 Summary

This research deals with a series of critical issues in the estimation and prediction of dynamic OD trip flows using various information sources for off-line and on-line DTA operations, namely, OD estimation using link counts and AVI counts for planning applications, real-time OD estimation and prediction, as well as real-time demand consistency checking and updating.

8.1.1 Off-line OD Demand Estimation

The inability of providing accurate OD demand estimates becomes a critical bottleneck in the deployment of advanced dynamic traffic assignment methodologies and other promising traffic information and management scenarios. Mary studies have been devoted to the dynamic OD demand estimation problem, in which time varying traffic link counts in a single day are the most widely used input. Since limited traffic observations from single data source typically lead to non-unique solutions and large estimation errors, current planning and operational applications place a great need for an effective modeling approach that can extract more information from other available data sources.

Based on an iterative bi-level estimation framework, this dissertation aims to enhance the quality of OD demand estimates by combining historical static demand information and time-varying archived traffic measurements into a flexible multi-objective optimization framework that minimizes the overall sum of squared errors. The one-day demand estimation formulation is extended to utilize multi-day link counts for inferring day-to-day demand variations. The field data from the Irvine test bed network are used to illustrate the proposed methodology and demonstrate the effectiveness of new estimation models.

To circumvent the difficulties of obtaining market penetration rates of transponder tags in a demand population, this research proposes a novel OD demand estimation formulation to effectively exploit OD demand distribution information provided by emerging Automatic Vehicle Identification (AVI) sensor data, and presents robust formulations to accommodate possible deviations from idealized conditions in the demand estimation process. Using a synthetic data set, this study evaluates the performance of new estimation models and provides the following key findings. (1) Sufficient sample size, which is jointly determined by market penetration rates and observation time interval, is necessary to obtain accuracy demand estimates from partially observed vehicle identification data. (2) In the presence of identification errors, a parsimonious structure accounting for imprecise information can provide more robust and accurate estimates than a complex joint estimation model for OD demand flow and identification rates.

8.1.2 Real-time OD Demand Estimation and Prediction

A structural real-time OD demand estimation and prediction model and a polynomial trend filter are developed to systematically model regular demand pattern information, structural deviations and random fluctuations so as to provide reliable prediction and capture the structural changes in time-varying demand, which can be caused by severe weather conditions and special events. Based on a Kalman filtering framework, an optimal adaptive updating procedure is further presented to use the real-time demand estimates to obtain *a priori* estimates of the mean and variance of regular demand patterns.

To maintain a representation of the network states consistent with that of the real-world traffic measurements in a real-time DTA system, this research proposes a dynamic OD demand optimal adjustment model and efficient sub-optimal feedback controllers to regulate the demand input for the real-time DTA simulator while reducing the adjustment magnitude. Without correcting OD demand estimation errors in the DTA simulator, the inconsistency in OD flows would accumulate in the DTA simulator and further propagate into the internal representation of path and link flows, making the network state prediction become highly unreliable. The proposed real-time OD estimation & prediction and consistency checking models are implemented and tested using the field data from the Irvine test bed network. In addition, this research conducts sensitivity analysis for various model structures and parameter settings, such as updating frequency and the magnitude of gain factor.

8.2 Research Contributions

This section presents specific contributions of this research to the theoretical and algorithmic developments of dynamic OD estimation and prediction.

To date, the potential benefits of utilizing new types of traffic measures to enhance travel demand modeling capability have not been adequately exploited in the traditional approaches. Many existing OD estimation and prediction models lack the capability to systematically capture the dynamic nature of OD trip desires, particularly the day-to-day evolution and possible structural changes encountered in real-world traffic systems. In addition, existing OD demand consistency models either lack efficient real-time solution implementations for realistic networks, or fail to systematically account for spatial and temporal interaction of demand errors.

Based on the simulation-based solution methodology for dynamic traffic assignment (Mahmassani, 1998), this dissertation significantly enriches the capabilities of the basic bi-level dynamic OD estimation model and iterative solution procedure proposed by Tavana (2001), and the real-time OD estimation and prediction framework proposed by Kang (1999). In this research, the simulation-based DTA program is used to describe the network flow pattern and provide measurement matrices for the estimation process. A wide range of optimization formulations and efficient algorithms are constructed to estimate and predict real-world OD demand flows, and control the OD demand loading in the DTA simulator. This dissertation provides the following key contributions.

(i) The proposed formulation can effectively incorporate historical static OD demand information while avoiding potential modeling biases. The multiple-objective

framework provides a tractable and intuitive approach to incorporate multiple information sources, and it can considerably alleviate the common under-specification problem encountered in traditional models. In addition, the multi-day estimation model offers valuable opportunities to utilize traffic flow time series to capture the day-to-day demand variations.

(ii) This research provides an effective mechanism to extract OD distribution pattern information from AVI data. The formulation based on point-to-point split samples circumvents the principle difficulties in estimating dynamic market penetration rates for different OD pairs. Moreover, the robust formulation utilizes the imprecise knowledge from the analyst and obviates the need for estimating dynamic and location-dependent identification rates.

(ii) The polynomial model in the real-time OD estimation and prediction model provides key capability to recognize the non-stationary characteristic in dynamic demand time series, and produces a compact state space representation for modeling lagged OD demand without complex state argumentation. The structural model represents a significant advantage over the existing forecasting models where regular pattern, structural trend and random variation are viewed in isolation of one another. This systematic integration reduces the order of the polynomial trend model, and decreases the demand prediction variability and computational complexity in both time and space dimensions.

(iii) The day-to-day updating formulation provides an integrated framework to account for the inherent uncertainty nature of OD estimates, including stochasticity and correlation in estimated values. The bi-objective consistency checking model

with the efficient normalized least mean squares algorithm can account for the spatial interaction of OD demand flows in the network, providing a tractable solution procedure for general networks.

Overall, this research successfully satisfies various requirements in different problem contexts with diverse information sources. More importantly, this dissertation strengthens the inherent connections between off-line OD estimation, on-line estimation & prediction and real-time consistency, rather than approaching these problem stages as a set of isolated problems. These models and algorithms are systematically integrated into off-line and on-line DTA systems, and are rigorously tested using the field data and synthetic data based on the real Irvine network.

8.3 Future Research

With enhanced demand estimation and prediction formulations, this dissertation illustrates considerable potential for generalizing the modeling framework into the field of traffic state estimation and prediction. These innovative methods still require further investigation into numerous issues, especially in the following dimensions.

Multi-process day-to-day demand updating

This study uses a single process model to represent the multi-day demand dynamics. To better understand the day-to-day OD demand evolutions in a real traffic system, there is a need to uncover complex multi-process characteristics by using more elaborate system representation models. More verification tests using real-world data are also necessary to investigate the effectiveness of day-to-day demand

updating models, especially for DTA planning applications in large-scale metropolitan networks.

Off-line and on-line estimation models using real-world AVI data

In this research, possible benefits of AVI data are investigated through experimental control using the synthetic data, and the AVI data are only utilized in the off-line OD estimation process. The real-world data can provide further insight on the actual performance of alternative estimation models. The on-line DTA applications also call for further integration of point-to-point sensor measurements in the real-time estimation and prediction of OD demand.

Dynamic estimation and prediction of route choice model parameters

The proposed OD demand estimation and prediction models suggest a promising modeling approach for estimating and predicting the other two key components in dynamic traffic states, namely, route choice behavior and traffic flow propagation. On the other hand, reducing the route choice and traffic flow propagation errors in the DTA system can significantly improve the performance of OD estimation, which relies on the DTA simulator to provide input-output mapping matrices. Collectively, these extensions will offer a comprehensive and integrated framework for estimating and predicting traffic network states.

Bibliography

Ahmed, M. and Cook, A. (1979) Analysis of Freeway Traffic Time Series Data by Using Box-Jenkins Techniques. *Transportation Research Board,* 722, pp.1-9.

Antoniou, C., Ben-Akiva, M. and Koutsopoulos, H. (2004) Incorporating Automated Vehicle Identification Data into Origin-Destination Estimation, Transportation Research Board CD-ROM Paper Preprints (TRB Paper No. 04-3341), National Research Council, Washington, D.C.

Asakura, Y., Hato, E. and Kashiwadani, M. (2000) OD Matrices Estimation Model Using AVI Data and its Application to the Han-Shin Expressway Network. *Transportation,* 27(4), pp.419-438.

Ashok, K. (1996) Estimation and Prediction of Time-Dependent Origin-Destination Flows, Ph.D. Dissertation, MIT.

Ashok, K. and Ben-Akiva, M. (1993) Dynamic Origin-Destination Matrix Estimation and Prediction for Real-Time Traffic Management Systems, *Proceedings the 12th International Symposium on Transportation,* C. Daganzo(ed.), Elsevier Science Publishing Company Inc.

Ashok, K. and Ben-Akiva., M. (2000) Alternative Approaches for Real-Time Estimation and Prediction of Time-Dependent Origin-Destination Flows. *Transportation Science,* 34(1), pp. 21-36.

Beckmann, M.J., McGuire, C.B. and Winston, C.B. (1956). *Studies in the Economics of Transportation.* Yale University Press, Connecticut.

Bell, M.G.H. (1983) The Estimation of an Origin-Destination Matrix from Traffic Counts. *Transportation Science,* 17(2), pp. 198-217.

Bertsekas, D.P. (1999) *Non-linear Programming,* Athena Scientific.

Bikowitz, E. and Ross, S. (1982) Evaluation and Improvement of Inductive Loop Detectors. *Transportation Research Record,* 841.

Brookner, E. (1998) Tracking and Kalman Filtering Made Easy, John Wiley and Sons, Inc, New York.

Cascetta, E., Inaudi, D. and Marquis, G. (1993) Dynamic Estimators of Origin-Destination Matrices Using Traffic Counts. *Transportation Science,* 27, pp. 363-373.

Cascetta, E. and Nguyen, S. (1988) A Unified Framework for Estimating or Updating Origin/Destination Matrices from Traffic Counts. *Transportation Research*, 22B, pp. 437-455.

Chang, G. L. and Tao, X. (1996) Estimation of Dynamic O-D Distributions for Urban Networks. In *Proceedings of 13th International Symposium on Transportation and Traffic Theory*, Lesort, J. B. ed. Tarrytown, NY.

Chang, G.L. and Tao, X. (1999) An Integrated Model for Estimating Time-Varying Network Origin-Destination Distributions. *Transportation Research* 33A (5), pp. 381-399.

Chang, G.L. and Wu, J. (1994) Recursive Estimation of Time-Varying Origin-Destination Flows from Traffic Counts in Freeway Corridors. *Transportation Research*, 28B (2), pp.141-160.

Chui, C. and Chen, G. (1991) Kalman Filtering with Real-Time Applications, Springer-Verlag.

Coifman, B. (1998) Vehicle Reidentification and Travel Time Measurement in Real-Time on Freeways Using the Existing Loop Detector Infrastructure. *Transportation Research Record*, 1643, pp. 181-191.

Coifman, B., Beymer, D., Mc Lauchlan, P. and Malik, J. (1998) A Real-Time Computer Vision System for Vehicle Tracking and Traffic Surveillance, *Transportation Research*, 6C, pp. 271-288.

Cremer, M. and Keller, H. (1981) Dynamic Identification of Flows from Traffic Counts at Complex Intersections. *Proceedings the 8th International Symposium on Transportation and Traffic Theory*, Toronto University.

Cremer M. and Keller, H. (1984) A Systems Dynamics Approach to the Estimation of Entry and Exit O-D Flows. *Proceedings of the 9th International Symposium on Transportation and Traffic Theory.* Volmuller, I. and Hamerslag, R. eds., VNU Science Press, Utrecht, The Netherlands.

Cremer M. and Keller, H. (1987) A New Class of Dynamic Methods for Identification of Origin-Destination Flows. *Transportation Research*, 21B, pp. 117-132.

Cremer, M. and Papageorgiou, P. (1981) Parameter Identification for a Traffic Flow Model. *Automatica*, 17, pp. 837-843.

Davis, G. and Nihan, N. (1991). Nonparametric Regression and Short-Term Freeway Traffic Forecasting. *Journal of Transportation Engineering*, 1991, pp. 78-188.

Dixon, M. P. and Rilett, L. R. (2002). Real-Time OD Estimation Using Automatic Vehicle Identification and Traffic Count Data. *Journal of Computer-Aided Civil and Infrastructure Engineering*, 17(1), pp. 7–21.

Dixon, M.P. (2000). Incorporation of Automatic Vehicle Identification Data into Synthetic OD Estimation Process, Ph.D. Dissertation, Texas A&M University.

Doan D.L., Ziliaskopoulos, A. and Mahmassani, H.S. (1999). An On-line Monitoring System for Real-Time Traffic Management Applications. *Transportation Research Record*, 1678, pp. 142-149.

Florian M. and Y. Chen (1994) A Coordinate Descent Method for the Bilevel O-D Matrix Adjustment Problem. Preprints of the 7th IFAC/IFORS Symposium on Transportation Systems: Theory and Application of Advanced Technology. Tianjin, China, pp. 1029-1034.

FHWA (1973) Urban Traffic Control System and Bus Priority System Traffic Adaptive Network Signal Timing Program: Software Description. Federal Highway Administration, US. Dept. of Transportation, Washington, D.C.

Fisk, C. S. (1989) Trip matrix Estimation from Link traffic Counts: The Congested Network Case. *Transportation Research*, 23B, pp.331-336.

Gazis, D. C. and Szeto, M. W. (1972). Application of Kalman Filtering to the Surveillance and Control of Traffic Systems. *Transportation Science*, 6(4), pp. 419-439.

Haas, C., Mahmassani, H. S., Khoury, J., Haynes, M., Rioux, T. and Logman, H., (2001) Evaluation of Automatic Vehicle Identification for San Antonio's Transguide for Incident Detection and Advanced Traveler Information Systems, Center for Transportation Research, University of Texas at Austin, Research Report, No. 4957-1.

Harvey, A. C. (1989) Forecasting Structural Time Series Models and the Kalman Filter. Cambridge University Press.

Kang, Y. (1999) Estimation and Prediction of Dynamic Origin-Destination (O-D) Demand and System Consistency Control for Real-Time Dynamic Traffic Assignment Operation. Ph.D. Dissertation, The University of Texas at Austin.

Koppelman, F.S. and Pas, E.I.. (1984) Estimation of Disaggregate Regression Models of Person Trip Generation with Multiday Data. *Proceedings of the 9th International Symposium on Transportation and Traffic Theory.* Volmuller, I. and Hamerslag, R. eds., VNU Science Press, Utrecht, The Netherlands. pp. 513-529.

Kwon, J., Coifman, B. and Bickel, P., (2000). Day-to-Day Travel Time Trends and Travel Time Prediction from Loop Detector Data. *Transportation Research Record*, 1717, pp. 120-129.

Leblanc L.J. and Farhangian K. (1982) Selection of a Trip Table which Reproduces Observed Link Flows. *Transportation Research*, 16B, pp. 83-88.

Lomax, T. J. and Schrank, D. L. (2002) The 2002 Urban Mobility Report, Texas Transportation Institute, Texas A&M University, College Station, TX.

Mahfoud, R. H. (2002) Distributed Software Design and Implementation of a Real-Time Traffic Estimation and Prediction System, M.S. Thesis, The University of Texas at Austin.

Mahmassani, H.S. et al. (2000) *DYNASMART-P, Volume II, User's Guide.* Technical Report STO67-85-PII, Center for Transportation Research, The University of Texas at Austin.

Mahmassani, H. S. (2001) Dynamic Network Traffic Assignment and Simulation Methodology for Advanced System Management Applications. *Networks and Spatial Economics*, 1:2, pp. 267-292.

Mahmassani, H. S., Hawas, Y., Abdelghani, K., Abdelfatah, A., Chiu, Y-C. and Kang, Y. (1998) *DYNASMART-X; Volume II: Analytical and Algorithmic Aspects.* Technical Report ST067-85-Volume II, Center for Transportation Research, The University of Texas at Austin.

Mahmassani, H.S., Hu, T., Peeta, S. and Ziliaskopoulos, A. (1994) Development and Testing of Dynamic Traffic Assignment and Simulation Procedure for ATIS/ATMS applications. Technical Report DTFH61-90-R-00074-FG, Center for Transportation Research, The University of Texas at Austin.

Maher, M.J. (1985) The Analysis of Partial Registration-Plate Data. *Traffic Engineering and Control*, 26, pp. 495-497.

Makowski, G.G. and Sinha, K.C. (1976) A Statistical Procedure to Analyze Partial License-Plate Surveys. *Transportation Research*, 10, pp. 131-132.

Merchant, D. and Nemhauser, G. (1978). A Model and an Algorithm for the Dynamic Traffic Assignment Problem. *Transportation Science*, 12, pp. 183-199.

Merchant, D. and Nemhauser, G. (1978). Optimality Conditions for a Dynamic Traffic Assignment Model. *Transportation Science*, 12, pp. 200-207.

Michalopoulos, P. G., Blake, W. and Benke, R. (1990) Testing and Field Implementation of the Minnesota Video Detection System (AUTOSCOPE). *Transportation Research Record*, 1287, pp. 176-185.

Morrison, N. (1969). Introduction to Sequential Smoothing and Prediction. McGraw-Hill, New York.

Nguyen S. (1977) Estimating an OD matrix from Network Data: A Network Equilibrium Approach, *Publication No. 87, Centre de Recherche sur les Tranports*, Universite de Montreal, Montreal, Quebec.

Nicholson, H. and Swann, C. (1974) The Prediction of Traffic Flow Volumes Based on Spectral Analysis. *Transportation Research*, 8, pp. 533-538.

Peeta, S. and Bulusu, S. (1999) A Generalized Singular Value Decomposition Approach for Consistent Online Dynamic Traffic Assignment, *Transportation Research Record*, 1667, pp. 77-87.

Peeta, S. and Mahmassani, H.S. (1995) Multiple User Classes Real-Time Traffic Assignment for On-Line Operations: a Rolling Horizon Solution Framework, *Transportation Research*, 3C(2), pp 83-98.

Peeta, S. and Ziliaskopoulos, A. (2001) Foundations of Dynamic Traffic Assignment: the Past, the Present and the Future. *Networks and Spatial Economics*, 1:2, pp. 233-266.

Okutani, I. and Stephanedes, Y. (1984) Dynamic Prediction of Traffic Volume Through Kalman Filtering Theory. *Transportation Research*, 18B(1), pp. 1-11.

Patriksson, M. (1994) The Traffic Assignment Problem: Models and Methods. Koninklijke Wohrmann, Zutphen, The Netherlands.

Sheffi, Y. (1985). Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods, Prentice-Hall, NJ.

Shuldiner, P.W., D'gostino, S.A. and Woodson, J.B. (1996). Determining Detailed Origin-Destination and Travel Time Patterns Using Video and Machine Vision License Plate Matching, *Transportation Research Record*, 1551, pp. 8-17.

Spiess H. (1987). A Maximum Likelihood Model for Estimating Origin-Destination Matrices. *Transportation Research*, 21B, pp. 395-412.

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

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

Tavana H., Mahmassani, H.S. and Haas, C. (1999) Effectiveness of Wireless Phones in Incident Detection: A Probabilistic Analysis. *Transportation Research Record*, 1683, pp. 31-37.

Turner, S.M., Eisele, W.L., Benz, R.J. and Holdener, D.J. (1998) Travel Time Data Collection Handbook. Report FHWA-PL-98-035, Federal Highway Administration, Office of Highway Information Management, Washington, D.C.

Van Aerde, M., Hellinga, B., Yu, L. and Rakha, H. (1993) Vehicle Probes as Real-Time ATMS Sources of Dynamic O-D and Travel Time Data. Large Urban Systems. Proceedings of the ATMS Conference, St. Petersburg, FL, pp. 207-230.

Van der Zijpp, N.J. (1997) Dynamic OD-Matrix Estimation from Traffic Counts and Automated Vehicle Identification Data. *Transportation Research Record,* 1607, pp. 87-94.

Van Zuylen H.J. and Willumsen L.G. (1980) The Most Likely Trip Matrix Estimated from Traffic Counts. *Transportation Research,* 14B, pp 281-293.

Wardrop, J.G. (1952). Some Theoretical Aspects of Road Traffic Research, Proceedings, Institution of Civil Engineers, II(1), pp. 325-378.

Watling, D.P. and Maher, M.J. (1992) A Statistical Procedure for Estimating a Mean Origin-Destination Matrix from a Partial Registration Plate Survey. *Transportation Research*, 26B, pp.289-314.

West, M. and Harrison, P.J. (1997) Bayesian Forecasting and Dynamic Models (2nd Edition). Springer-Verlag.

Whittaker, J., Garside, S. and Lindveld, K. (1997). Tracking and Predicting a Network Traffic Process. *International Journal of Forecasting*, 13, pp. 51-61

Yang, H., Sasaki, T., Iida, Y. and Asakura, Y. (1992) Estimation of Origin-Destination Matrices from Link Traffic Counts on Congested Networks. *Transportation Research,* 26B, pp. 417-434.